Applied Stochastic Processes and Control for Jump-Diffusions: Modeling, Analysis and Computation

Floyd B. Hanson
University of Illinois
Chicago, Illinois, USA

January 7, 2007

To five generations of women in my life,
Margaret Geiger, Violet Bliss, Ethel Hutchins, Lisa Hanson
and Chiara Hanson Whitehurst

Contents

List of Figures

List of Tables

Preface

Everything should be as simple as it is,
but not simpler.
—Albert Einstein (1879-1955).

A mathematical theory is not to be considered complete
until you have made it so clear that you can explain it
to the first man whom you meet on the street.
—David Hilbert (1862-1943).

Always take a pragmatic view in applied mathematics:
the proof of the pudding is in the eating.
—N. H. Bingham and Rüdiger Kiesel (2004) [33].

Overview of This Book

The aim of this book is to be a self-contained, practical, entry level text on stochastic processes and control for jump-diffusions in continuous time, technically Markov processes in continuous time.

The book is intended for graduate students as well as a research monograph for researchers in applied mathematics, computational science and engineering. Also, the book may be useful for practicianers of financial engineering who need fast and efficient answers to stochastic financial problems. Hence, the exposition is based upon integrated basic principles of applied mathematics, applied probability and computational science. The target audience includes mathematical modelers and students in many areas of science and engineering seeking to construct models for scientific applications subject to uncertain environments. The prime interest is in modeling and problem solving. The utility of the exposition, based upon systematic derivations along with essential proofs in the spirit of classical applied mathematics, is more important to setting up a stochastic model of an application than abstract theory. However, a lengthy last chapter is intended to bridge the gap between the applied world and the abstract world in order to enable applied students and readers to understand the more abstract literature.

More rigorous theorem formulation and proving is not of immediate importance compared to modeling and solving an applied problem, although many proofs are given here. Many research problems deal with new applications and often these new applications require models beyond those in the existing literature. So, it is important to have a reasonably understandable derivation for a nearby model that can be perturbed to obtain a proper new model. The level of rigor here is embodied in correct and systematic derivations, with many proofs and results not available elsewhere, under reasonable conditions, not necessarily the tightest possible conditions. In fact, much of this book and the theory of Markov processes in continuous time is based upon modifying the formulations for continuous function in calculus to extend them to the discontinuous and non-smooth functions of stochastic calculus.

Origin of the Book

The book is based upon the author's courses *Math 574 Applied Optimal Control*, *Math 590 Special Topics: Applied Stochastic Control*, *MCS 507 Mathematical, Statistical and Scientific Software for Industry* and partly on *MCS 571 Numerical Methods for Partial Differential Equations*. In addition, the results from research papers on computational stochastic dynamic programming are included. Courses in asymptotic analysis and numerical analysis play a role as well. However, as with lectures, every attempt is made to keep the book self-contained through an integrated approach, without depending heavily on prerequisites, especially with a diverse readership and interdisciplinary topics.

This book integrates many of the research and exposition advances made in computational stochastic dynamic programming and stochastic modeling. They exhibit the broader impact of the applications and the computationally oriented approach. The stochastic applications are wide-ranging, including the optimal economics of biological populations in uncertain and disastrous environments, biomedical applications in cancer modeling and optimal treatment, and financial engineering with applications in option pricing and optimal portfolios.

How This Book is Organized and How to Use It

- A *prependix* rather than an appendix, Chapter B of preliminaries is intended as a reference for topics in probabilIty, matrix algebra, analysis and other topics that are too numerous to expect with a wide-ranging interdisciplinary book such as this one. Over-specification of prerequisites tend to filter out too many students who could benefit from this material. This prependix is intended to bring all readers up to the same level by self-study, where necessary, of the basic concepts and notations of probability and analysis needed for jump-diffusion processes and their deviations from continuity. It is not meant to be taught or read in sequence, but to include relevant results when needed and to make the presentation as self-contained as possible.

- Simple jump-diffusion Chapters 1, 2, 3 and 4 cover the basics for simple jump-diffusions, i.e., stochastic diffusion (Wiener of Brownian motion) and simple

Poisson driven processes, including stochastic integration and stochastic calculus for transformations of stochastic differential equations (SDEs). The speed and depth of coverage for the student or reader will depend on their level of knowledge, particularly with respect to prior knowledge of probability and diffusion processes which are more well known. The presentation is more elementary than that of later chapters to reduce the likelihood that readers will get lost at the basic level.

- Advanced and special topics are found in Chapters 5 to 12 and can be selected according to the instructor's or reader's interests. There are more chapters than can be covered in any one course.

  - Chapter 5 covers more advanced and general topics for SDEs. These include jumps driven by compound Poisson or Poisson random measure processes that allow randomly distributed jump-amplitudes, state-dependent jump-diffusions and multidimensional jump-diffusions.
  - Chapters A and 6 can form a control theory component of a course with either deterministic or stochastic optimal control chapters or both. Chapter A gives a summary of deterministic optimal control results to provide a background for comparison to the stochastic optimal control results, but could be skipped if a deterministic control course is a prerequisite or if only stochastic optimal control are of interest.

    In Chapter 6 stochastic optimal control problems are introduced and the equation of stochastic dynamic programming is systematically derived from the basic principles of applied mathematics.
  - Chapter 7 concerns partial differential equation (PDE) methods for solving stochastic problems using the forward and backward Kolmogorov equations, Dynkin's integral formulas (also Feynmann-Kac's as Dynkin's with an integrating factor) that help provide PDE solutions without directly solving the PDE, boundary conditions and stopping time problems. Knowledge of partial derivatives from advanced courses in calculus is all that should be needed, a course in PDEs will be of little help, since a course is not essential and only these integral formulas are used in this chapter. PDE methods are an applied alternate to the abstract method of using martingales to solve stochastic problems, such as those in finance (see Chapter 12 for martingale and other abstract approaches.)
  - Chapters 8 and 9 form a computational component of a course with either computational stochastic dynamic programming or computational simulations or both. Chapter 8 has treatments using either modified finite difference methods for optimal control problem or the Markov chain approximation methods. Computational methods are important for stochastic optimal control problems because there are so few exact analytical solutions.

    Chapter 9 contains treatments for direct simulations of SDEs and general simulations by the Monte Carlo method.

- Chapter 10 on financial applications and Chapter 11 on biomedical applications provide substantial examples of application of the theory and techniques treated in this book. Chapter 10 explains Merton's mathematical justification and generalization of the classical Black-Scholes option pricing problem in sufficient detail for those familiar with the diffusion processes properties in Chapters 1-4 and is a good motivating application for Chapter 5. Also treated are option pricing models for jump-diffusions, optimal portfolio and consumption models, and an important events model that modifies the jump-diffusion model with a quasi-deterministic jump model for scheduled announcements and random responses.

  Chapter 11 includes applications to stochastic optimal control or bioeconomic models, diffusion approximation models of tumor growth and a deterministic optimal control model of PDE-driven drug delivery model for the brain.

- Chapter 12 is an applied description of abstract probability methods, including probability measure, probability space, martingales and change in probability measure using either Radyn-Nikodým and Girsanov theorems. The last section is a generalization of jump-diffusions called Lévy processes that permit the jump-rate to be infinite. This chapter is meant to be a bridge between the applied view of stochastic processes and the abstract view to ease the transition to reading some of the more abstract literature on stochastic processes. However, depending on the instructor or reader, parts of this chapter can be woven into the coverage of the earlier chapters. For instance, a colleague said that Girsanov's measure change transformation was needed in his financial applications course and there are a pure diffusion version and a jump-diffusion version of the Girsanov theorem in this chapter.

Distinct Features of This Book

The book is based upon a number of distinct features:

- Both analytical and computational methods are emphasized based on the utility, with respect to the computational complexity, of the problems. Exercises and examples in the elementary chapters include both computational and analytic ones. Students need to have good analytic and computational skills to do well, since diverse skills are needed for many jobs.

- The treatment of jump and diffusion processes is balanced as well, rather than a stronger or nearly exclusive emphasis on diffusion processes. This is a unique feature of this book. This treatment of jump-diffusions is important for training graduate students to do research on stochastic processes, since the analysis of diffusion processes is so well-developed, there are many opportunities for open problems on jump-diffusions.

- It clearly shows the strong role that discontinuous as well as non-smooth properties of stochastic processes play compared to the random properties by emphasizing a concrete jump calculus, without much reliance on measure-theoretic constructs.

- Basic principles of probability theory in the spirit of classical applied mathematics are used to set up the practical foundations through clear and systematic derivations, making the book accessible as a research monograph to many who work with applications.

- It shows how analytical-canonical control problem models, such as the linear-quadratic, jump-diffusion (LQJD) problem and financial risk-adverse power utilities, can be used to reduce computational dimensional complexity of approximate solutions along with other computational techniques.

- Insightful and useful material are used so that the book can be readily used to model realistic applications and even modify the derivations when new applications do not quite fit the old stochastic model.

- Clear explanations for the entry level student are used. In particular, clear and consistent notation is used, such that the notation is clearly identified with the quantity it symbolizes, rather than arbitrarily selected. Sometimes this has meant some compromise on some standard notation, for instance, P is used for the Poisson process to be consistent with the W used for the Wiener process. This means that P could not be used for probability, so Prob is used in place of P (or Pr) and is clearer to a diverse audience. Similarly, probability distributions are denoted by by Φ and densities by ϕ since P is used for Poisson and F is used for transformation functions throughout the book.

Target Audience

Colleagues and students have requested a more accessible, practical treatment of these topics. They are interested in learning about stochastic calculus and optimal stochastic control in continuous time, but reluctant to invest time to learn it from more advanced treatments relying heavily on abstract concepts. Hence, this book should be of interest to an interdisciplinary audience of applied mathematicians, applied probabilists, engineers (including control engineers dealing with deterministic problems and financial engineers needing fast as well as useful methods for modeling rapidly changing market developments), statisticians and other scientists. After this primary audience, a secondary audience would be mathematicians, engineers and scientists, using this book as a research monograph, seeking more intuition to more fully understand stochastic processes and how the more advanced analytical approaches fit in with important applications like financial market modeling.

Prerequisites

For optimal use of this book, it would be helpful to have had prior introduction to applied probability theory including continuous random variables, mathematical analysis at least at the level of advanced calculus. Ordinary differential equations, partial differential equations and basic computational methods would be helpful but the book does not rely on prior knowledge of these topics by using basic calculus style motivations. In other words, the more or less usual preparation for students of applied mathematics, science and engineering should be sufficient. However, the author has strived to make this book as self-contained as practical, not strongly relying on prior knowledge and explaining or reviewing the prerequisite knowledge at the point it is needed to justify a step in the systematic derivation of some mathematical result.

MATLAB Computation

As part of the theme of balancing computation and analysis, MATLAB™, the matrix laboratory computation system is used for almost all computational examples and figure illustrations. Simple MATLAB codes are described in class and the code for all text figures are given in Appendix C. MATLAB greatly facilitates the development of code and is ideally suited to stochastic processes and control problems. Also, MATLAB now comes with the Maple™ kernel built into the MATLAB student package for including elementary symbolic computations with numeric computations. Beyond the initial elementary assignments, the students are required to submit their assignments with professionally done illustrations for which they can find examples in Appendix C. Many students surveyed at the end of the class actually list MATLAB with the other topics that they were happy to learn. MATAB is also helpful later for producing professional research papers and theses.

Acknowledgments

The author is grateful to a number of co-workers and students who helped as reviewers or contributed to this applied stochastic book through research contributions, as well as other authors and agencies giving grant support for computational stochastic dynamic programming:

My research assistants and graduate students, Siddhartha Pratim Chakrabarty, Zongwu Zhu, Guoqing Yan and Jinchun Ye have helped review drafts of this book with the the keen eyes of applied mathematics and computer science graduate students to make sure that it would be useful and understandable to other graduate students. My wife Ethel did a major job at the final proof-read.

Over the years many have helped develop pieces of the underlying applied theory or model applications: Abdul Majid Wazwaz, Dennis Ryan, Kumarss Naimipour, Siu-Leung Chung, Huihuang (Howard) Xu, Dennis J. Jarvis, Christopher J. Pratico, Michael S. Vetter, Raghib abu-Saris, and Daniel L. Kern.

This work has been influenced, consciously and subconsciously, from books and related works by many authors such as

Applebaum [12], Arnold [13], Bingham and Kiesel [33], Blis [40], Çinlar [55], Clark [56], Cont and Tankov [59], Feller [83, 84], Fleming and Rishel [85], Gihman and Skorohod [94, 95], Goel and Richter-Dyn [98], Glasserman [96], Hammersley and Handscomb [104], D. Higham [139, 140], Hull [147], Itô [149], Jäckel [150] Jazwinski [154], Karlin and Taylor [161, 162, 265] Kirk [163], Kloeden and Platen [165], Kushner [173, 175], Kushner and Dupuis [179], Ludwig [187], Merton [203], Mikosch [209], Øksendal [222], Øksendal and Sulem, [223], Parzen [224], Protter [232], Runggaldier [239], Schuss [244], Snyder and Miller [252], Tuckwell [270], Steele [256], Wonham [285], and others. Although this influence may not be directly apparent here, some have shown how to make the presentation much simpler, while others have supplied the motivation to simplify the presentation, making it more accessible to a more general audience and other applications.

This material is based upon work supported by the National Science Foundation under Grants No. 02-07081, 99-73231, 96-26692, 93-01107, 91-02343 and 88-0699 in the Computational Mathematics Program entitled: *Advanced Computational Stochastic Dynamic Programming for Continuous Time Problems* at the University of Illinois at Chicago. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. In addition, NSF supplied a Research Experience for Undergraduates support for Mike Vetter to develop a portable object-oriented version of our multi-dimensional computational control visualization system.

Argonne National Laboratory Advanced Computing Research Facility (ANL/ACRF) supplied parallel processing training through summer and sabbatical support that enabled the development of large scale computational stochastic applications from 1985-1988.

Many of our national supercomputing centers have provided supercomputing time on the the currently most powerful supercomputers for continuing research for solving large scale stochastic control problems in *Advanced Computational Stochastic Dynamic Programming* and also for computational science education. In addition to Argonne National Laboratory, these were National Center for Supercomputing Applications (NCSA), Los Alamos National Laboratory's Advanced Computing Laboratory (LANL/ACL), Cornell Theory Center (CTC/CNSF), Pittsburgh Supercomputing Center (PSC) and the San Diego Supercomputing Center (SDSC/NPACI) during 1987-2003.

At the University of Illinois Chicago, the Laboratory of Advanced Computing (UIC/LAC) and associate centers have supplied us with cluster computing and the Electronic Visualization Laboratory (UIC/EVL) supplied a most capable master's graduate student, Chris Pratico, and facilities for the developing a

multi-dimensional computational control visualization system using a real-time socket feed from our Los Alamos National Laboratory account.

Chapter 1

Stochastic Jump and Diffusion Processes: Introduction

Life is good for only two things, discovering mathematics and teaching mathematics.
—Siméon Denis Poisson (1781-1840).

I do not regret my attempts, for it is only by trying problems that exceed his powers that the mathematician can ever learn to use these powers to their full extent.
—Norbert Wiener (1894-1964) in *Ex-Prodigy*.

The generation of random numbers is too important to be left to chance.
—Robert Coveyou at `http://www.xs4all.nl/~jcdverha/scijokes/1_5.html#subindex`.

1.1 Poisson and Wiener Processes Basics

This chapter introduces Wiener processes $W(t)$ and simple Poisson jump processes $P(t)$ in differential and integral forms. The Wiener and Poisson processes form the tools of a toolbox to create jump-diffusion process models. Wiener processes also called diffusion or loosely Brownian motion.

The processes $W(t)$ and $P(t)$ are **continuous-time stochastic processes** which basically means they are continuous time-dependent random variables[1]. They are also a special form of stochastic processes called a Markov process that is without memory of all but the prior state and can be simply defined [55], repeating the essential definition given in the previous chapter, as

[1] In this book, the words **stochastic** and **random** have the same meaning, involving probability or chance.

Definition 1.1. *The stochastic process $X(t)$ is a* ***Markov process*** *provided the conditional probability satisfies*

$$\text{Prob}[X(t+\Delta t) = x \mid X(s), 0 \le s \le t] = \text{Prob}[X(t+\Delta t) = x | X(t)] \,,$$

for any $t \ge 0$, any $\Delta t \ge 0$ and x is in the state space, $\mathcal{D}_x$.

The stochastic processes serve as useful concepts for modeling random changes in time with stochastic differential equations, similar to the use of ordinary differential equations to model deterministic (non-stochastic) problems. These standard processes have basic infinitesimal moments

$$\text{E}[dW(t)] = 0 \;\text{ and }\; \text{Var}[dW(t)] = dt \tag{1.1}$$

for the **differential Wiener process** with initial condition $W(0^+) = 0$ **with probability one (w.p.o.)**, while

$$5\text{E}[dP(t)] = \lambda dt = \text{Var}[dP(t)] \tag{1.2}$$

for the **differential of the simple Poisson counting process** with rate $\lambda > 0$ and initial condition $P(0^+) = 0$ **with probability one**. The Wiener process is a mathematical idealization of **Brownian motion**, but often the term Brownian motion is used instead of the term Wiener process.

Remark 1.2. *If the W and P processes started at a different initial time other than zero, say at $t = t_0$, then the initial conditions would be changed to $W(t_0^+) = 0^+$ and $P(t_0^+) = 0^+$, respectively. There is not much special about the zero initial conditions, just convenience and standardization.*

The simplest and very useful view of these differential stochastic processes is to consider them defined as increments, i.e.,

$$dW(t) \equiv W(t+dt) - W(t) \tag{1.3}$$

and

$$dP(t) \equiv P(t+dt) - P(t) \,, \tag{1.4}$$

for infinitesimal increments in time dt. The property that

$$\text{Var}[dW(t)] = \text{E}[(dW(t))^2] = dt \tag{1.5}$$

is motivation for the non-differentiability of the $W(t)$ process since the limit of

$$\sqrt{\text{Var}[dW(t)]}/dt = \sqrt{\text{E}[(dW(t))^2]}/dt = \frac{1}{\sqrt{dt}} \to +\infty \tag{1.6}$$

as $dt \to 0^+$, i.e., the variance of the ratio of differentials $\text{Var}[dW(t)/dt] \to +\infty$ as $dt \to 0^+$. Hence, the differentiability of $W(t)$ is inconsistent with the failure of the variance of the quotient $dW(t)/dt$ in the limit $dt \to 0^+$. Equation (1.6) says that the root mean square (RMS) derivative becomes unbounded as $dt \to 0^+$. This is not a rigorous proof that $W(t)$ is a non-smooth process, although $W(t)$ is a continuous process from (1.1). (For a proof that $W(t)$ is non-differentiable see the theorem below.)

1.2 Wiener Process Basic Properties

The assumptions for the Wiener process, including that of being normally distributed, are the properties:

Properties 1.3. *The* ***standard Wiener process*** $\boldsymbol{W(t)}$

- $W(t)$ *is a* ***continuous process****, since*

$$W(t^+) = W(t) = W(t^-), \quad t > 0\,.$$

- $W(t)$ *has* ***independent increments****, since the Wiener increments*

$$\Delta W(t_i) = W(t_i + \Delta t_i) - W(t_i)$$

are mutually independent for all t_i *on non-overlapping time intervals. The* ***non-overlapping time intervals*** *are defined such that* $t_i \geq 0$, $t_{i+1} = t_i + \Delta t_i$ *and any* $\Delta t_i > 0$ *for* $0 = 1 : n$, *so that*

$$t_i < t_{i+1} \text{ for } i = 0\!:\!n.$$

Noting that $W(t_i) = W(0) + \sum_{j=0}^{i-1} \Delta W(t_j)$, *so depends on all preceding increments, recalling that* $W(0) = 0$ *with probability one at* $t_0 = 0$ *i.e.,*

$$\text{Prob}[\Delta W(t_i) \leq w_i, \Delta W(t_j) \leq w_j] = \text{Prob}[\Delta W(t_i) \leq w_i] \cdot \text{Prob}[\Delta W(t_j) \leq w_j]\,,$$

if $j \neq i$, *such that there is no overlap in the time intervals* $[t_i, t_{i+1})$ *and* $[t_j, t_{j+1})$. *Note that* $\Delta W(t_i)$, *as a forward increment is independent (see Definition B.35 for independent random variables) of* $W(t_i)$ *and that* $\Delta W(t_i) \equiv W(t_i + \Delta t_i) - W(t_i)$ *is associated with the time interval* $[t_j, t_j + \Delta t_j)$, *open on the right to be compatible with right continuity of the Poisson process.*

- $W(t)$ *is a* ***stationary process****, since the distribution of the increment* $\Delta W(t) = W(t + \Delta t) - W(t)$, *with* $\Delta t > 0$, *is independent of* t.

- $W(t)$ *is a* ***Markov process****, since*

$$\text{Prob}[W(t + \Delta t) = w \mid W(s), s \leq t] = \text{Prob}[W(t + \Delta t) = w | W(t)]\,,$$

for any $t \geq 0$, *any* $\Delta t \geq 0$. *(It is helpful to note that* $W(t)$ *is synonymous with the increment* $(W(t) - W(0))$.*)*

- $W(t)$ *is normally distributed with mean* $\mu = 0$ *and variance* $\sigma^2 = t$, $t > 0$, *i.e., the density of* $W(t)$ *is*

$$\phi_{W(t)}(w) = \phi_n(w; 0, t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{w^2}{2t}\right)\,, \tag{1.7}$$

when $-\infty < w < +\infty$ *and* $t > 0$. *(The actual distribution function for* $W(t)$, $\Phi_{W(t)}(w)$, *has been given already in (B.22).)*

- $W(0) = 0$ ***with probability one***, *since* $\phi_{W(0^+)}(w) = \delta(w)$ *from (1.7), i.e., in the limit as* $t \to 0^+$ *(see the Exercise 22 on p. B75 in Section B.16).*

Thus, the increments $\Delta[W(t + i\Delta t)] \equiv W(t + (i + 1)\Delta t) - W(t + i\Delta t)$ for $i = 0, 1, \ldots$ are stationary, independent and identically distributed (IID) as a normal distribution given time step Δt and $t \geq 0$, i.e.,

$$\phi_{\Delta W(t)}(w) = \phi_n(w; 0, \Delta t) = \frac{1}{\sqrt{2\pi\Delta t}} \exp\left(-\frac{w^2}{2\Delta t}\right) , \tag{1.8}$$

when $-\infty < w < +\infty$ and $\Delta t > 0$. So the basic moments of the Wiener increments are

$$\mathrm{E}[\Delta W(t)] = 0, \quad \mathrm{Var}[\Delta W(t)] = \Delta t. \tag{1.9}$$

Similarly, by the stationarity property of the $dW(t) = W(t + dt) - W(t)$ differential process when $dt > 0$ has the same probability distribution as the process $W(dt)$ when $t > 0$ and that the distribution from (1.7) is normal with mean $\mu = 0$ and variance $\sigma^2 = dt$,

$$\phi_{dW(t)}(w) = \phi_n(w; 0, dt) = \frac{1}{\sqrt{2\pi dt}} \exp\left(-\frac{w^2}{2dt}\right) , \tag{1.10}$$

when $-\infty < w < +\infty$ and $dt > 0$.

Theorem 1.4. ***Covariance of*** $\mathbf{W(t)}$***:*** *If* $W(t)$ *is a Wiener process, then*

$$\mathrm{Cov}[W(t), W(s)] = \min[t, s] . \tag{1.11}$$

Proof. This theorem is a very elementary application of the independent increment and mean zero properties of Wiener or diffusion processes, also demonstrating how application of independent increments rely on the zero mean property. The zero mean property implies that $\mathrm{E}[W(t)] = 0 = \mathrm{E}[W(s)]$. First consider the case $s < t$ and write $W(t) = W(s) + (W(t) - W(s))$, i.e., as independent increments (see Definition B.35 for expectations of products independent random variables) and noting that the first increment is $W(s) - W(0) = W(s)$ on $[0, s)$ since $W(0) = 0$ and the second increment is on $[s, t)$, then

$$\begin{aligned}
\mathrm{Cov}[W(t), W(s)] &= \mathrm{E}[W(t)W(s)] = \mathrm{E}[W^2(s) + W(s)(W(t) - W(s))] \\
&= \mathrm{E}[W^2(s)] + \mathrm{E}[W(s)(W(t) - W(s))] \\
&= \mathrm{Var}[W(s)] + \mathrm{E}[W(s)]\mathrm{E}[(W(t) - W(s))] \\
&= s + 0 \cdot 0 = s ,
\end{aligned}$$

using the linearity of the expectation operator (B.9), the definition of the variance (B.10) together with the separability of expectations (B.80) for independent

increments $W(s)$ and $(W(t) - W(s))$, and finally that $W(s)$ denotes the independent increment $W(s) - W(0)$ with variance s (B.22, 1.7). In the case $t < s$, then $\mathrm{Cov}[W(t), W(s)] = t$ by symmetry using the splitting $W(s) = W(t)+(W(s)-W(t))$, and combining both cases produces the conclusion $\mathrm{Cov}[W(t), W(s)] = \min[s, t]$, where the function $\min[s, t]$ denotes the minimum of s and t. □

When computing diffusion sample paths, i.e., the trajectory of $W(t)$ in time t, it is necessary to break up the time domain, say $[0, T]$ into small increments $\Delta T = T/N$ where N is the number of random samples that will be used, so that each corresponding Wiener increment $\Delta W(t_i)$ will be independent. Since $W(0) = 0$ **with probability one**, let $t_i = i \cdot \Delta T$ for $i = 0 : N$, then

$$W(t_{i+1}) = \sum_{j=0}^{i} \Delta W(t_j) \, .$$

Using MATLAB™ , for instance, an integer state , say 0, is selected with the MATLAB command

```
randn('state',0);
```

where `'state'` is a literal script argument specified that this call is to set the random state of the function `randn`. A row N-vector set of diffusion increments can be computed wholesale by the formula,

```
DWv = sqrt(DT)*randn(1,N);
```

where `randn(N,1)` is the $N \times 1$ standard zero-mean, unit-variance normal random generator of MATLAB. The factor `sqrt(DT)` is the Wiener scaling for the square root of the variance (1.9). Then the simulated trajectory can be computed by

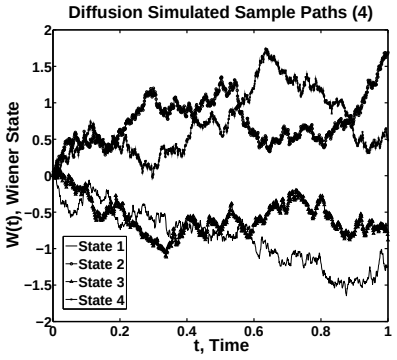
```
tv = 0:DT:T;  % time vector tv(1:N+1).
for  i = 1:N
     Wv(i+1) = sum((DWv(1:i)));
end
```

assuming $Wv(1) = 0.0$ in the MATLAB shifted subscript base at one, rather than at zero. Finally, the diffusion sample path can be plotted with

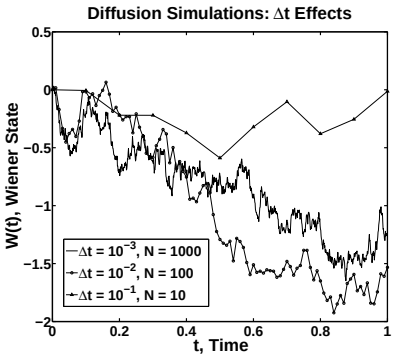
```
plot(tv,Wv,'k-');
```

and results for four sample paths are displayed in Fig. 1.1(a) using $N = 1000$, $T = 1.0$ and $k = 1:4$ `randn` states. The MATLAB program used to generate this part of the figure is given in Program C.7 given in Appendix C.

In Fig. 1.1(b), the variation of the fine structure of the sample path is displayed, with time step size using subsets of the same random sample state. The sample paths in this case differ markedly since the sample subsets are quite different in quantity, being $N = 1000, 100$ and 10 random sample points for $\Delta t = 10^{-3}, 10^{-2}$ and 10^{-1}, respectively, so the different cumulative set of random points leads to quite different random trajectories.



(a) Diffusion sample paths using four random states.



(b) Diffusion sample paths using three different time steps.

Figure 1.1. *In Figure 1.1(a), paths were simulated using MATLAB [210] with* $N = 1000$ *sample points, four* `randn` *states and maximum time* $T = 1.0$*. In Figure 1.1(b), paths were simulated using subsets of the same random state of* `randn` *used for the finer grid* 0.001*.*

1.3 More Wiener Process Moments

The expectations for the integer powers of the Wiener increment follow from the mean using the Wiener increment normal density (1.8). Only the even integer powers, $m = 2k$, need some calculation since the means will be zero for the odd integer powers due to the even-ness of the density on $(-\infty, +\infty)$, i.e., $\mathrm{E}[(\Delta W(t))^{2k+1}] = 0$ when $k = 0, 1, 2, \ldots,$

$$\begin{aligned}
\mathrm{E}\left[(\Delta W(t))^m\right] = \mathrm{E}\left[(\Delta W(t))^{2k}\right] &= \int_{-\infty}^{+\infty} \phi_n(w; 0, \Delta t) w^{2k} dw \\
&= \frac{2}{\sqrt{2\pi\Delta t}} \int_0^{+\infty} \exp\left(-\frac{w^2}{2\Delta t}\right) w^{2k} dw \\
&= \frac{(2\Delta t)^k}{\sqrt{\pi}} \int_0^{+\infty} \exp(-u)\, u^{k-1/2} du \\
&= \frac{(2\Delta t)^k \Gamma(k + 1/2)}{\Gamma(1/2)}\,, \qquad (1.12)
\end{aligned}$$

for $k = 0, 1, 2, \ldots,$ where Γ is the gamma function [2] defined by

$$\Gamma(x) \equiv \int_0^{\infty} e^{-u} u^{x-1} du\,, \quad x > 0\,, \qquad (1.13)$$

with initial condition $\Gamma(1) \equiv 1$ and special value $\Gamma(1/2) = \sqrt{\pi}$. The gamma function is like a generalized factorial function, due to the recursive form $\Gamma(x + 1) = x\Gamma(x)$ so that $\Gamma(x + 1) = x!$. The final formula (1.12) satisfies the recursion

$$g_{2k+2}(\Delta t) \equiv \mathrm{E}[(\Delta W)^{2k+2}(t)] = (k + 1/2)(2\Delta t) g_{2k}(\Delta t).$$

Further, note that the final formula (1.12) holds for any integer m when the $\Delta W(t)$ is replaced by the absolute value, i.e.,

$$\mathrm{E}[|\Delta W(t)|^m] = (2\Delta t)^{m/2}\Gamma((m+1)/2)/\Gamma(1/2). \tag{1.14}$$

The final formula (1.12) satisfies the recursion

$$g_{m+2}(\Delta t) \equiv \mathrm{E}[(\Delta W)^{m+2}(t)] = (m+1)\Delta t g_m(\Delta t)\,,$$

for $m = 0:\infty$, starting from $g_0(\Delta t) = 1$ or $g_1(\Delta t) = \sqrt{2\Delta t/\pi}$. The results for the first few powers are summarized in Table 1.1:

Table 1.1. *Some expected moments (powers) of absolute value of the Wiener increments.*

m	$\mathrm{E}[\lvert\Delta W(t)\rvert^m]$
0	1
1	$\sqrt{2\Delta t/\pi}$
2	Δt
3	$2\Delta t\sqrt{2\Delta t/\pi}$
4	$3(\Delta t)^2$
5	$8(\Delta t)^2\sqrt{2\Delta t/\pi}$
6	$15(\Delta t)^3$
⋮	⋮
2k	$(2k-1)!!(\Delta t)^k$
2k+1	$k!(2\Delta t)^k\sqrt{2\Delta t/\pi}$

In Table 1.1, the function $(2k-1)!!$ is defined below.

Definition 1.5. ***Double Factorial Function:***

$$(2k-1)!! = (2k-1)\cdot(2k-3)\cdots 1\,, \tag{1.15}$$

denotes the ***double factorial function****, given here for odd arguments. For example,* $1!! = 1$*,* $3!! = 3$ *and* $5!! = 15$*.*

For even arguments the double factorial function is proportional to the standard factorial function,

$$(2k)!! = 2^k k!\,.$$

Example 1.6. *These results can be applied to other expected moments, for example,*

$$\begin{aligned}\mathrm{Var}[(\Delta W)^2(t)] &= \mathrm{E}[((\Delta W)^2(t) - \Delta t)^2] \\ &= \mathrm{E}[(\Delta W)^4(t)] - 2\Delta t\mathrm{E}[(\Delta W)^2(t)] + (\Delta t)^2\mathrm{E}[1] \\ &= 2(\Delta t)^2\,,\end{aligned} \tag{1.16}$$

upon expanding the square and using the linear property of the expectation.

The moment calculation in (1.12) can be implemented directly by symbolic computation, for example by Maple.

Example 1.7. ***Wiener Moments by Maple:***

\# ***Maple Functions for Wiener Moments and Wiener Deviation Moments:***

```
> restart : interface(showassumed = 0) : assume(s2 > 0) :
> fnormal := (x, m, s2)− > exp(−(x − m) ∗ (x − m)/(2 ∗ s2))/sqrt(2 ∗ Pi ∗ s2);
```

$$fnormal := (x, m, s2) \rightarrow \frac{e^{\left(1/2\frac{(x-m)^2}{s2}\right)}}{\sqrt{2\pi s2}}$$

```
> momentdw := (n, m, s2)− > simplify(int(x^n ∗ fnormal(x, m, s2),
> x = −infinity..infinity));
```

$$momentdw := (n, m, s2)) \rightarrow simplify\left(\int_{-\infty}^{\infty} x^n fnormal(x, m, s2)dx\right)$$

```
> momentdevdw := (n, m, s2)− > simplify(int((x − m)^n fnormal(x, m, s2),
> x = −infinity..infinity));
```

$$momentdevdw := (n, m, s2) \rightarrow simplify\left(\int_{-\infty}^{\infty} (x-m)^n fnormal(x, m, s2)dx\right)$$

\# *Sample illustrations for moment functions:*

```
> assume(dt > 0) : assume(sigma > 0) :
> mom6dw := collect(momentdw(6, mu ∗ dt, sigma^2 ∗ dt), dt);
```

$$mom6dw := \mu^6 dt^6 + 15\mu^4\sigma^2 dt^5 + 45\mu^2\sigma^4 + 15\sigma^6 dt^3$$

```
> mom6devdw := momentdevdw(6, mu ∗ dt, sigma^2 ∗ dt);
```

$$mom6dw := 15dt^3$$

```
> mom5absdevdw := momentabsdevdw(5, mu ∗ dt, sigma^2 ∗ dt);
```

$$mom6dw := \frac{8dt^{(5/2)}\sqrt{2}}{\sqrt{\pi}}$$

Remarks 1.8.

- *The results can also be applied to expected moments of Wiener differential process, $dW(t) = W(t+dt) - W(t)$, by replacing single appearances Δt's by dt, i.e., $\Delta t \to dt$ is assumed, and neglecting terms of $\mathrm{O}^2(\Delta t)$ as $\Delta t \to 0^+$ since they are treated as negligible compared to terms of $\mathrm{ord}(\Delta t)$ as $\Delta t \to 0^+$.*

- *Sometimes to keep the steps in a derivation simple, the infinitesimal dt will be treated as being an infinitesimally small object such that as $dt \to 0^+$, $(dt)^2 \ll 1$ or $(dt)^2 < \mathrm{ord}(dt)$ and similarly for higher powers of dt. However, when there are no order Δt terms in the answer, then, as in (1.16), the proper leading order (by definition nonzero) would be of interest. Expected moments of $W(t)$ also follow by replacing Δt by t, except the higher powers of t would not be negligible compared to the first power, unless t is small.*

1.4 Wiener Process Non-Differentiability

Theorem 1.9. ***Non-differentiability of*** $\mathbf{W(t)}$***:***
For any fixed $x > 0$ and $t > 0$,

$$\mathrm{Prob}\left[\lim_{\Delta t \to 0^+}\left[\left|\frac{\Delta W(t)}{\Delta t}\right| > x\right]\right] = 1. \tag{1.17}$$

Proof. Let $x > 0$ be fixed, $t > 0$, $0 < \Delta t \ll 1$, then interchanging limit with probability operations since time is deterministic and using the normal distribution of the increment $\Delta W(t) = W(t+\Delta t) - W(t)$ in (1.8),

$$\begin{aligned}
\mathrm{Prob}\left[\lim_{\Delta t \to 0^+}\left[\left|\frac{\Delta W(t)}{\Delta t}\right| > x\right]\right] &= \lim_{\Delta t \to 0^+}\left[\mathrm{Prob}\left[\left|\frac{\Delta W(t)}{\Delta t}\right| > x\right]\right] \\
&= \lim_{\Delta t \to 0^+}\left[\mathrm{Prob}\left[|\Delta W(t)| > x\Delta t\right]\right] \\
&= \lim_{\Delta t \to 0^+}\left[\frac{2}{\sqrt{2\pi\Delta t}}\int_{x\Delta t}^{\infty}\exp\left(-\frac{w^2}{2\Delta t}\right)dw\right] \\
&= \lim_{\Delta t \to 0^+}\left[\frac{2}{\sqrt{2\pi}}\int_{x\sqrt{\Delta t}}^{\infty}\exp\left(-\frac{v^2}{2}\right)dv\right] \\
&= \frac{2}{\sqrt{2\pi}}\int_0^{\infty}\exp\left(-\frac{v^2}{2}\right)dv = 1\,,
\end{aligned}$$

for any $x > 0$ and $t > 0$ fixed. Note that the error is

$$\frac{2}{\sqrt{2\pi}}\int_0^{x\sqrt{\Delta t}}\exp\left(-\frac{v^2}{2}\right)dv \le \frac{2}{\sqrt{2\pi}}\int_0^{x\sqrt{\Delta t}} 1\,dv = \frac{2}{\sqrt{2\pi}}x\sqrt{\Delta t} \ll 1\,,$$

since $\exp(-v^2/2) \le 1$. Further note that we can take x as large as we please, as long as it is fixed, so that $\Delta W(t)/\Delta t$ must be unbounded as $\Delta t \to 0^+$ **with probability one** for each t. Hence, the Wiener process $W(t)$ is non-differentiable or non-smooth

with probability on for each t. (See also Mikosch [209, Sect. A3, p. 188], for a similar proof using less direct methods; see Steele [256, Sect. 5.2, p. 63] for more precise conditions.) □

1.5 Wiener Process Expectations Conditioned on Past

Example 1.10. ***Illustration of Independent Increments and Markov Properities for Wiener Process:***

- $\mathrm{E}[W(t)|W(r), 0 \le r \le s] = W(\min[s,t])$.
  Note that the conditioning set $\{W(r), 0 \le r \le s\}$ denotes the past when $t > s \ge 0$, viewing $W(t)$ as the sum of two independent increments $(W(s) - W(0)) + (W(t) - W(s))$, noting that $W(0) = 0$. However, when $0 \le t \le s$, then the increment $W(t) \equiv (W(t) - W(0))$ is a constant relative to the conditioning set, so the result depends on the relation between t and s using the rule $\mathrm{E}[f(X)|X] = f(X)$ given in Chapter B on Page B28. Hence,

$$\begin{aligned}
\mathrm{E}[W(t)|W(r), 0 \le r \le s] &= \left\{ \begin{array}{l} W(t),\ 0 \le t \le s \\ \mathrm{E}[W(s) + (W(t) - W(s))|W(r), 0 \le r \le s], \\ \qquad 0 \le s < t \end{array} \right\} \\
&= \left\{ \begin{array}{l} W(t),\ 0 \le t \le s \\ \mathrm{E}[W(s)|W(r), 0 \le r \le s] + \mathrm{E}[(W(t) - W(s))], \\ \qquad\qquad\qquad\qquad\qquad 0 \le s < t \end{array} \right\} \\
&= \left\{ \begin{array}{l} W(t),\ 0 \le t \le s \\ W(s) + 0,\ 0 \le s < t \end{array} \right\} = \left\{ \begin{array}{l} W(t),\ 0 \le t \le s \\ W(s),\ 0 \le s < t \end{array} \right\} \\
&= W(\min[s,t])\,,
\end{aligned}$$

  where the independent increment property was used along with the zero mean property of the increment, $\mathrm{E}[\Delta W(t)] = 0$ and the completely conditioned rule that $\mathrm{E}[f(X)|X] = f(X)$. The function $\min[s,t]$ denotes the minimum of s and t. The linear property of the conditional expectation was also used.

  When $0 \le s < t$ then the formula,

$$\mathrm{E}[W(t)|W(r), 0 \le r \le s] = W(s)\,, \tag{1.18}$$

  signifies that the average information conditioned on the past data, $\{W(r), r \in [0,s]\}$, is given by the most recent past data $W(s)$, which may imply a significant reduction in uncertainty for the present data, $W(t)$.

  The form of the expectation result (1.18) is the principal characteristic form for a ***martingale*** *$X(t)$,*

$$\mathrm{E}[X(t)|X(r), 0 \le r \le s] = X(s)\,, \tag{1.19}$$

  where $X(t) = f(W(t))$ for instance. The martingale is an abstract model of a fair game (see the beginning preliminary chapter of Mikosch [209] for a clear description of martingales, but in an abstract presentation; martingales will be described at the end of this book in Chapter 12 with full qualifications).

- $\mathrm{E}[W^2(t)|W(r), 0 \le r \le s] = W^2(\min[s,t]) + (t-s)H(t-s)$,
  where $H(X)$ is the Heaviside step function (B.158). This result is derived similarly to the prior result for the conditional mean, but much more algebra is required, although many of the small details of the prior derivation are omitted.

$$\begin{aligned}
\mathrm{E}[W^2(t)|W(r), 0 \le r \le s] &= \left\{ \begin{array}{l} W^2(t),\ 0 \le t \le s \\ \mathrm{E}[(W(s) + (W(t) - W(s)))^2 | W(r), 0 \le r \le s], \\ \qquad\qquad\qquad\qquad 0 \le s < t \end{array} \right\} \\
&= \left\{ \begin{array}{l} W^2(t),\ 0 \le t \le s \\ W^2(s) + 2W(s)\mathrm{E}[(W(t) - W(s))] + \mathrm{E}[(W(t) - W(s))^2], \\ \qquad\qquad\qquad\qquad 0 \le s < t \end{array} \right\} \\
&= \left\{ \begin{array}{l} W^2(t),\ 0 \le t \le s \\ W^2(s) + 2W(s) \cdot 0 + (t-s),\ 0 \le s < t \end{array} \right\} \\
&= \left\{ \begin{array}{l} W^2(t),\ 0 \le t \le s \\ W^2(s) + (t-s),\ 0 \le s < t \end{array} \right\} \\
&= W^2(\min[s,t]) + (t-s)H(t-s).
\end{aligned}$$

  Here, the increment variance $\mathrm{Var}[\Delta W(t)] = \Delta t$ has been used.

  The general technique for powers $W^m(t)$ when $s < t$ with conditioning on $W(s)$ is to use the decomposition into independent increments $W(t) = W(s) + (W(t) - W(s))$ and then expand the power of m by the binomial expansion (B.152)

$$(W(s) + (W(t) - W(s)))^m = \sum_{k=0}^{m} \left(\begin{array}{c} m \\ k \end{array} \right) W^k(s)(W(t) - W(s))^{m-k},$$

  and then use independence of the increments and conditioning to calculate for each term,

$$\begin{aligned}
&\mathrm{E}\left[\left(\begin{array}{c} m \\ k \end{array} \right) W^k(s)(W(t) - W(s))^{m-k} \middle| W(r), 0 \le r \le s \right] \\
&\qquad = \left(\begin{array}{c} m \\ k \end{array} \right) W^k(s)\mathrm{E}\left[(W(t) - W(s))^{m-k} \right],
\end{aligned}$$

  relying on Table 1.1 for the remaining expectation.

The term normal distribution is more often used in mathematics and statistics, while the term Gaussian distribution may be used more often in other sciences and engineering.

1.6 Poisson Process Basic Properties

Since the Poisson process suffers from positive jumps of integer magnitude the Poisson process is also discontinuous, which makes the differentiability problems of the Poisson process of secondary importance. For this reason, the Poisson process is also called a **counting process** or **point process**. Thus, the analytical problems are even more severe than for the Wiener process, since the singularities of the Poisson process arise at the zeroth order with the value of $P(t)$ jumping, while those of $W(t)$ arise at the first order derivative. However, the jumps of the Poisson

process have a modeling benefit over the Wiener process in that the Poisson process is useful for applications with disasters or crashes and those with bonanzas or rallies.

In summary, **Poisson process $P(t)$ is a discontinuous process** and satisfies the following properties:

Properties 1.11. ***Simple Poisson Process $P(t)$:***

- *$P(t)$ has **unit jumps**, since if the value of $P(t)$ jumps at time $T_k > 0$, then*

$$P(T_k^+) = P(T_k^-) + 1\,,$$

where $P(T_k^+)$ denotes the limit from the right and $P(T_k^-)$ the limit from the left, so $P(t)$ is discontinuous, increasing and has instantaneous jumps.

- *$P(t)$ is **right-continuous**, since*

$$P(t^+) = P(t) \geq P(t^-), \quad t > 0\,. \tag{1.20}$$

- *$P(t)$ has **independent increments**, since the Poisson increments*

$$\Delta P(t_i) \equiv P(t_i + \Delta t_i) - P(t_i)$$

are mutually independent for all t_i on non-overlapping time intervals defined such that $t_i \geq 0$, $t_{i+1} = t_i + \Delta t_i$ and any $\Delta t_i > 0$ for $0 = 1 : n$ so that

$$t_i < t_{i+1} \text{ for } i = 0 : n\,,$$

noting that $P(t_i) = P(0) + \sum_{j=0}^{i-1} \Delta P(t_j)$, depending on all preceding increments, recalling that $P(0) = 0$ with probability one at $t_0 = 0$, i.e.,

$$\text{Prob}[\Delta P(t_i) \leq p_i, \Delta P(t_j) \leq p_j] = \text{Prob}[\Delta P(t_i) \leq p_i] \cdot \text{Prob}[\Delta P(t_j) \leq p_j]\,,$$

if $j \neq i$, such that there is no overlap in the time intervals $(t_i, t_{i+1}]$ and $(t_j, t_{j+1}]$. Note that $\Delta P(t_i)$, as a forward increment is independent (see Definition B.35 for expectations of products independent random variables), of $P(t_i)$ and recall that $\Delta P(t_i) \equiv P(t_i + \Delta t_i) - P(t_i)$ is associated with the time interval $[t_j, t_j + \Delta t_j)$, open on the right since the process $P(t_i)$ is right continuous.

- *$P(t)$ is a **stationary process**, since the distribution of the increment $\Delta P(t) = P(t + \Delta t) - P(t)$ is independent of t.*

- *$P(t)$ is a **Markov process**, since*

$$\text{Prob}[P(t + \Delta t) = k \,|P(s), s \leq t] = \text{Prob}[P(t + \Delta t) = k \,|P(t)]\,,$$

for any $t \geq 0$, any $\Delta t > 0$. (It is helpful to note that $P(t)$ is synonymous with the increment $(P(t) - P(0))$.)

- $P(t)$ *is* ***Poisson distributed*** *with mean* $\mu = \lambda t$ *and variance* $\sigma^2 = \lambda t$, $t > 0$, *i.e.,*

$$\Phi_{P(t)}(k; \lambda t) = \text{Prob}[P(t) = k] \equiv p_k(\lambda t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} , \tag{1.21}$$

  for integer values $k = 0, 1, 2, \ldots$, *with constant* $\lambda > 0$ *and* $t \geq 0$.

- $P(0^+) = 0^+$ ***with probability one****, since from (1.21),* $p_k(0^+) = \delta_{k,0}$, *i.e., in the limit as* $t \to 0^+$.

See also Çinlar [55] or Snyder and Miller [252] for a more essential list of assumptions.

Thus, for $P(t)$, the increments $\Delta[P(t + i\Delta t)] \equiv P(t + (i + 1)\Delta t) - P(t + i\Delta t)$ for $i = 0, 1, \ldots$ are independent and identically distributed (IID) given time step $\Delta t > 0$ and $t \geq 0$.

By the stationarity property of the Poisson process increment $\Delta P(t) = P(t + \Delta t) - P(t)$ has the same discrete distribution as $P(\Delta t)$ in (1.21), so has the parameter $\lambda \Delta t$ instead of the λt in (B.50), i.e.,

$$\Phi_{\Delta P(t)}(k; \lambda \Delta t) = \text{Prob}[\Delta P(t) = k] = p_k(\lambda \Delta t) = e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^k}{k!} , \tag{1.22}$$

for $k = 0, 1, 2, \ldots$, $t \geq 0$ and $\Delta t \geq 0$.

Similarly, by the stationarity property of the differential, $dP(t) = P(t + dt) - P(t)$, for Poisson process has the same discrete distribution as $P(dt)$ in (1.21), except that $dP(t)$ has the parameter λdt instead of the λt in (B.50) for $P(t)$. Thus $dP(t)$ has the distribution,

$$\Phi_{dP(t)}(k; \lambda dt) = \text{Prob}[dP(t) = k] = p_k(\lambda dt) = e^{-\lambda dt} \frac{(\lambda dt)^k}{k!} , \tag{1.23}$$

for $k = 0, 1, 2, \ldots$, $t \geq 0$ and $dt \geq 0$. The distribution (1.23) might be considered as a limiting version of the more basic and proper incremental version in (1.22).

The simulation of the simple Poisson process $P(t)$ is usually based upon simulating the time between jumps, the inter-arrival time $T_{k+1} - T_k$, since the inter-arrival time can be shown to be exponentially distributed as sketched in Chapter B.

Lemma 1.12. ***Exponential Distribution of Time Between Jumps:***
Let $P(t)$ *be a simple Poisson process with fixed jump frequency* $\lambda > 0$ *and let* T_j *denote the* j*th jump time, then the distribution of the* ***inter-jump time*** $\Delta T_j \equiv T_{j+1} - T_j$ *for* $j = 0, 1, 2, \ldots$, *defining* $T_0 \equiv 0$, *conditioned on* T_j, *is*

$$\Phi_{\Delta T_j}(\Delta t) = \text{Prob}[\Delta T_j \leq \Delta t \,|\, T_j] = 1 - e^{-\lambda \Delta t} . \tag{1.24}$$

Proof. The basic idea of this proof is that the probability of the time between jumps $\Delta T_j = T_{j+1} - T_j$ less than Δt, conditioned on the prior jump time T_j, will

be the same as the probability that there be at least one jump in the time interval, which is the same as one minus the probability that there are no jumps in the time interval, i.e.,

$$\begin{aligned}\text{Prob}[\Delta T_j \le \Delta t \mid T_j] &= 1 - \text{Prob}[\Delta T_j > \Delta t \mid T_j] \\ &= 1 - \text{Prob}[\Delta P(T_j) = 0 \mid T_j] \,.\end{aligned}$$

However, by the stationary property of the simple Poisson process $P(t)$ the probability of the difference does not depend on the common time T_j, but on the difference in time ΔT_j,

$$\begin{aligned}\text{Prob}[\Delta T_j \le \Delta t \mid T_j] &= 1 - \text{Prob}[P(\Delta t) - P(0) = 0] \\ &= 1 - \text{Prob}[P(\Delta t) = 0] = 1 - p_0(\lambda \Delta t) \\ &= 1 - e^{-\lambda \Delta t} = \Phi_e(\Delta t; 1/\lambda) \,,\end{aligned}$$

where the fact that $P(0) = 0$ with probability one has been used, Poisson distribution $p_k(\lambda \Delta t)$ is given in (1.22) and the exponential distribution $\Phi_e(t; \mu)$ is given in (B.40). □

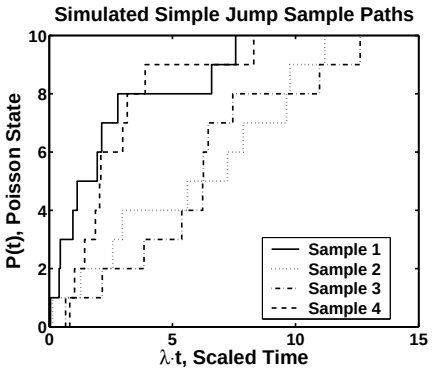
Using MATLAB with the efficient and fundamental distribution transformation from uniform to exponential distribution (B.42) , a uniformly distributed pseudo-random number generator can be used. These numbers can be generated wholesale, in vector form, for plotting or other applications, using a given K samples and the Poisson parameter value `lambda`, by the following code fragment,

```
Uv = rand(1,K); T(1) = 0; kv(1) = 0;
for k = 1:K, kv(k+1) = k;
    T(k+1) = T(k) - log(Uv(k))/lambda;
end
plot(kv,T,'k-');
```

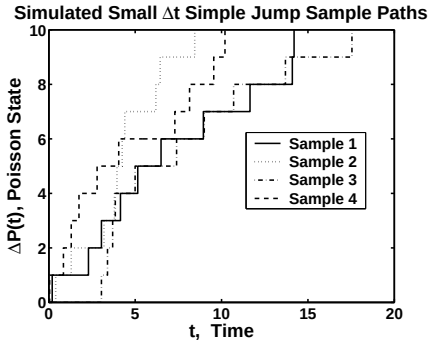
where `log` is the MATLAB natural logarithm notation. See the comments about (B.44) explaining why the proper term $\log(\texttt{Uv}(k))$ is used here rather than the less efficient term $\log(1 - \texttt{Uv}(k))$.

Since the natural time variable for Poisson is scaled as $\lambda * t$, four sample paths for $P(t)$ are illustrated in Fig. 1.2(a) versus the dimensionless time $\lambda * t$. The variation with the jump rate λ can be deduced since higher frequencies ($\lambda > 1$) compress the time axis and lower frequencies ($\lambda < 1$) expand the time axis. Note that the exponentially distributed inter-jump or inter-arrival times must used for simulating $P(t)$ since the Poisson distribution is not useful in simulating the jump times directly. The MATLAB source code for the left figure is given in Program C.9 in Appendix C.

In Fig. 1.2(b) are the corresponding sample paths for the Poisson process increment $\Delta P(t)$ when the time increments between jumps are sufficiently small so that the zero-one jump law, discussed more extensively in Theorem 1.19 in Section 1.7,



(a) Four Poisson jump $P(t)$ sample paths.



(b) Incremental Poisson jump $\Delta P(t)$ sample paths using different time steps.

Figure 1.2. *In Figure 1.2(a), Simulated sample paths for the simple Poisson Process $P(t)$ versus the dimension-less time λt using four different MATLAB [210] random states for four different sample paths and the exponential distribution of the time between jumps. In Figure 1.2(b) is a similar illustration for the simple Poisson process increment simulations versus t with $\lambda = 1.0$ and $\Delta t = 0.05$, based upon the zero-one jump law implemented with a uniform distribution paths were simulated using subsets of the same random state of* `rand` *used for the finer grid* 0.001.

applies and the time between jumps is uniformly distributed with asymptotic probability $\lambda\Delta t$ for the next jump and $(1 - \lambda\Delta t)$ for zero jumps, since

$$\text{Prob}[T_{k+1} - T_k \leq \Delta t \,|\, T_k] = 1 - e^{-\lambda\Delta t} \sim \lambda\Delta t \,,$$

provided $\lambda\Delta t \ll 1$, i.e., small, taking $\Delta t = 0.05$ and $\lambda = 1.0$. The small time increment process can be numerically simulated by a standard uniform number generator like MATLAB's `rand` and the **method of acceptance-rejection** [230, 96] such that the open interval $(0, 1)$ is partitioned into a centered interval of length $\lambda\Delta t$ and the complement of $(0, 1)$. When a uniformly generated point lands in the centered interval then a jump is counted, while there is no jump if it lands in the complement. The centered interval, $[(1 - \lambda\Delta t)/2, (1 + \lambda\Delta t)/2]$ is used to avoid the bias of open interval property of pseudo-random number generators where the neighborhood of the end points is excluded by a very small amount that is the order of the **machine epsilon** (the smallest positive floating number that is significant when added to one, theoretically, 2^{-53} in IEEE standard double precision). A sufficiently large sample should approximately satisfy the zero-one jump law probabilities, since the rejection method is equivalent to the unit step function applied $U(X_u; (1 - \lambda\Delta t)/2, (1 + \lambda\Delta t)/2)$ to the uniform variate X_u approximately

generated by `rand` and the expectation is

$$\begin{aligned}\mathrm{E}\left[U(X_u;(1-\lambda\Delta t)/2,(1+\lambda\Delta t)/2)\right] &= \int_0^1 U(u;(1-\lambda\Delta t)/2,(1+\lambda\Delta t)/2)du \\ &= \int_{(1-\lambda\Delta t)/2}^{(1+\lambda\Delta t)/2} du = \lambda\Delta t\,.\end{aligned}$$

The MATLAB source code for the right figure is given in Program C.10 of Appendix C.

Theorem 1.13. ***Covariance of*** $\mathbf{P(t)}$***:*** *If $P(t)$ is a Poisson process, then*

$$\mathrm{Cov}[P(t),P(s)] = \lambda\min[t,s]\,. \tag{1.25}$$

Proof. This theorem is a very elementary application of the independent increment property of Poisson or Markov jump processes, also demonstrating how application of independent increments rely on the zero mean properties. For the Poisson process, unlike the standardized diffusion process, the zero mean property comes from using the Poisson deviation or centered Poisson processes $\widehat{P}(t) = P(t) - \lambda t$ where $\mathrm{E}[P(t)] = \lambda t$, such that $\mathrm{E}[\widehat{P}(t)] = 0 = \mathrm{E}[\widehat{P}(s)]$. First consider the case $s < t$ and write

$$\widehat{P}(t) = \widehat{P}(s) + (\widehat{P}(t) - \widehat{P}(s)),$$

i.e., as independent increments noting the time increment $\Delta t = t - s$, the first increment is $\widehat{P}(s) - \widehat{P}(0) = \widehat{P}(s)$ since $\widehat{P}(0) = 0$ and that subtracting the mean terms λt and λs preserves the independent increment property since functions of independent random variables are independent (B.80). Then

$$\begin{aligned}\mathrm{Cov}[P(t),P(s)] &= \mathrm{E}[\widehat{P}(t)\widehat{P}(s)] = \mathrm{E}[\widehat{P}^2(s) + \widehat{P}(s)(\widehat{P}(t) - \widehat{P}(s))] \\ &= \mathrm{E}[\widehat{P}^2(s)] + \mathrm{E}[\widehat{P}(s)(\widehat{P}(t) - \widehat{P}(s))] \\ &= \mathrm{Var}[P(s)] + \mathrm{E}[\widehat{P}(s)]\mathrm{E}[\widehat{P}(t) - \widehat{P}(s)] \\ &= \lambda s + 0\cdot 0 = \lambda s\,,\end{aligned}$$

using the linearity of the expectation operator (B.9), the definition of the variance (B.10) together with the independence of the expectations (B.80) for independent increments $\widehat{P}(s)$ and $(\widehat{P}(t) - \widehat{P}(s))$, and finally that $\widehat{P}(s)$, with $P(s)$, has variance λs (1.21). In the case $t < s$, then $\mathrm{Cov}[P(t),P(s)] = \lambda t$ by symmetry, and both cases together produce the conclusion $\mathrm{Cov}[P(t),P(s)] = \lambda\min[s,t]$. □

1.7 Poisson Process Moments

The expectations for the integer powers of the Poisson increment follow from the mean over the Poisson distribution (1.22) and summed by differentiation of the exponential series (B.53):

Lemma 1.14. ***Poisson Sums by Differentiation:***

$$\mathrm{E}[(\Delta P)^m(t)] = e^{-\lambda\Delta t}\sum_{k=0}^{\infty}\frac{(\lambda\Delta t)^k k^m}{k!} \tag{1.26a}$$

$$= \left[e^{-u}\left(u\frac{d}{du}\right)^m e^u\right]\Big|_{u=\lambda\Delta t}, \tag{1.26b}$$

for $m = 0, 1, 2, \ldots$.

The result (1.26b) can be shown by induction from the definition (1.26). Either the direct summation form (1.26) or the differentiation form (1.26b) can be implemented by symbolic computation, for example the summation definition form can be coded in Maple as

Example 1.15. ***Poisson Moment Summations by Maple:***

\# ***Maple Functions for Poisson Moments and Poisson Deviation Moments:***

```
> fpoisson := (k, u)- > exp(−u) ∗ u^k/k!;
```

$$fpoisson := (k,u) \rightarrow \frac{e^{(-u)}u^k}{k!}$$

```
> momentdp := (n, u)- > simplify(sum('k^n fpoisson(k, u)','k' = 0..infinity));
```

$$momentdp := (n,u) \rightarrow simplify\left(\sum_{'k'=0}^{\infty} 'k^n fpoisson(k,u)'\right)$$

```
> momentdevdp := (n, u)- > simplify(sum('(k − u)^n fpoisson(k, u)',
> 'k' = 0..infinity));
```

$$momentdevdp := (n,u) \rightarrow simplify\left(\sum_{'k'=0}^{\infty} '(k-u)^n fpoisson(k,u)'\right)$$

\# *Sample illustrations for 5th moment of both moment functions:*

```
> mom5dp := momentdp(5, lambda ∗ dt);
```

$$mom5dp := \lambda dt(1 + 15\lambda dt + 25\lambda^2 dt^2 + 10\lambda^3 dt^3 + \lambda^4 dt^4)$$

```
> mom5devdp := momentdevdp(5, lambda ∗ dt);
```

$$mom5dp := 10\lambda^2 dt^2 + \lambda dt$$

Table 1.2. *Some expected moments (powers) of Poisson increments and their deviations.*

m	$\mathrm{E}[(\Delta P)^m(t)]$	$\mathrm{E}[(\Delta P(t)-\lambda\Delta t)^m]$
0	1	—
1	$\lambda\Delta t$	0
2	$\lambda\Delta t(1+\lambda\Delta t)$	$\lambda\Delta t$
3	$\lambda\Delta t(1+3\lambda\Delta t+(\lambda\Delta t)^2)$	$\lambda\Delta t$
4	$\lambda\Delta t(1+7\lambda\Delta t+6(\lambda\Delta t)^2+(\lambda\Delta t)^3)$	$\lambda\Delta t(1+3\lambda\Delta t)$
5	$\lambda\Delta t(1+15\lambda\Delta t+25(\lambda\Delta t)^2+10(\lambda\Delta t)^3+(\lambda\Delta t)^4)$	$\lambda\Delta t(1+10\lambda\Delta t)$

The results for the first few powers are summarized in Table 1.2: The second column of this table can be quickly calculated by recursion, since if $u=\lambda\Delta t$ and $g_m(u)=\mathrm{E}[(\Delta P)^m(t)]$, then it can be shown that $g_{m+1}(u)=u\cdot(g_m(u)+g'_m(u))$. See Exercise 6 on Page 27 for the asymptotic form of $\mathrm{E}[(\Delta P)^m(t)]$. The expectation of a general function, $\mathrm{E}[f(\Delta P(t))]$, in terms of an infinite series of the finite differences of $f(0)$, which terminates if $f(\Delta P(t))$ is an integer power of $\Delta P(t)$, is the topic of Exercise 7 on Page 28.

These tabulated results can be applied to other expected moments, for example,

$$\begin{aligned}\mathrm{Var}[\Delta P(t)] &= \mathrm{E}[(\Delta P(t)-\lambda\Delta t)^2]\\ &= \mathrm{E}[(\Delta P)^2(t)]-2\lambda\Delta t\mathrm{E}[\Delta P(t)]+(\lambda\Delta t)^2\mathrm{E}[1]=\lambda\Delta t\,,\end{aligned}$$

upon expanding the square and using the linear property of the expectation. See the third column of Table 1.2. The results can also be applied to expected moments of Poisson differential process as an increment process, $dP(t)=P(t+dt)-P(t)$, by replacing Δt by dt and neglecting terms of $\mathrm{O}^2(dt)$ since they are treated as negligible compared to term of $\mathrm{ord}(dt)$, dt being infinitesimally small. Expected moments of $P(t)$ also follow by replacing Δt by t, except the higher powers of t would not be negligible compared to the first power, unless t is small.

1.8 Poisson Poisson Zero-One Jump Law

Theorem 1.16. ***Zero-One Jump Law Order of Magnitude of Error for*** $\Delta P(t)$***:***
As $\Delta t\to 0^+$ *and* $\lambda>0$ *as well as bounded, then*

$$\mathrm{Prob}[\Delta P(t)=0]=1-\lambda\Delta t+\mathrm{O}^2(\lambda\Delta t)\,, \tag{1.27}$$

$$\mathrm{Prob}[\Delta P(t)=1]=\lambda\Delta t+\mathrm{O}^2(\lambda\Delta t)\,, \tag{1.28}$$

$$\mathrm{Prob}[\Delta P(t)>1]=\mathrm{O}^2(\lambda\Delta t)\,, \tag{1.29}$$

$$\mathrm{Prob}[(\Delta P)^m(t)=\Delta P(t)]=1-\frac{1}{2}(\lambda\Delta t)^2+\mathrm{O}^3(\lambda\Delta t),\quad m\geq 2\,. \tag{1.30}$$

Proof. Taking the Poisson increment distribution (1.22) and expanding it asymptotically using primarily the exponential series expansion (B.53) for $\Delta t \ll 1$ yields,

$$\begin{aligned}
\text{Prob}[\Delta P(t) = 0] &= e^{-\lambda\Delta t} = 1 - \lambda\Delta t + \frac{1}{2}(\lambda\Delta t)^2 + \text{O}^3(\lambda\Delta t)\,, \\
\text{Prob}[\Delta P(t) = 1] &= e^{-\lambda\Delta t}\lambda\Delta t = \lambda\Delta t - (\lambda\Delta t)^2 + \text{O}^3(\lambda\Delta t)\,, \\
\text{Prob}[\Delta P(t) > 1] &= 1 - \text{Prob}[\Delta P(t) = 0] - \text{Prob}[\Delta P(t) = 1] \\
&= \frac{1}{2}(\lambda\Delta t)^2 + \text{O}^3(\lambda\Delta t)\,.
\end{aligned}$$

Since $\text{O}^2(\lambda\Delta t) + \text{O}^3(\lambda\Delta t) = \text{O}^2(\lambda\Delta t)$, the first three equations are proved. The last equation (1.30) follows from the fact that $x^m = x$ is only true for $m \geq 2$ if $x = 0$ or $x = 1$, so

$$\begin{aligned}
\text{Prob}[(\Delta P)^m(t) = \Delta P(t)] &= \text{Prob}[\Delta P(t) = 0] + \text{Prob}[\Delta P(t) = 1] \\
&= 1 - \text{Prob}[\Delta P(t) > 1] = 1 - \frac{1}{2}(\lambda\Delta t)^2 + \text{O}^3(\lambda\Delta t)\,.
\end{aligned}$$

The significance of this result is that if $\lambda\Delta t$ is sufficiently small and terms of order $(\lambda\Delta t)^2$ can be neglected, then only jumps of zero or one are very likely, i.e., very probable. □

Remarks 1.17.

- *In some other texts, the three small Poisson increment properties, Eqs. (1.27, 1.28, 1.29), are used as an elementary definition of the simple Poisson process. Here, we have started at a higher level of definition to facilitate the use of the Poisson process in applications.*

- *Combining the asymptotic probability relations (1.28) for $\Delta P(t) = 1$ and (1.29) for $\Delta P(t) > 1$ leads to*

  $$\text{Prob}[\Delta P(t) > 1] \ll \text{Prob}[\Delta P(t) = 1]$$

  *when $\lambda\Delta t \ll 1$. This asymptotic relationship characterizes the **orderliness** property of Poisson process (see Snyder and Miller [252]).*

With this result, the corresponding results for differential Poisson processes follow. First, a definition to specify that the square of a differential as been neglected.

Definition 1.18. ***Equality to Precision-***dt***:***
Let $f(dt; x)$ and $g(x)$ be bounded functions for $dt \geq 0$ and parameter x. Write

$$f(dt; x) \stackrel{\text{dt}}{=} g(x)dt \tag{1.31}$$

if

$$f(dt;x) = g(x)dt + \mathrm{o}(dt)$$

as $dt \to 0^+$ *and fixed* x.

Theorem 1.19. ***Zero-One Jump Law for*** $dP(t)$***:***
Let $dt > 0$ *and let* λ *be positive and bounded, then*

$$\mathrm{Prob}[dP(t) = 0] \stackrel{\mathrm{dt}}{=} 1 - \lambda dt\,, \tag{1.32}$$

$$\mathrm{Prob}[dP(t) = 1] \stackrel{\mathrm{dt}}{=} \lambda dt\,, \tag{1.33}$$

$$\mathrm{Prob}[dP(t) > 1] \stackrel{\mathrm{dt}}{=} 0\,, \tag{1.34}$$

$$\mathrm{Prob}[(dP)^2(t) = dP(t)] \stackrel{\mathrm{dt}}{=} 1\,, \tag{1.35}$$

$$\mathrm{Prob}[(dP)^m(t) = dP(t)] \stackrel{\mathrm{dt}}{=} 1\,, m > 0\,. \tag{1.36}$$

Proof. The proof follows easily from the increment approximate Theorem 1.16 upon neglecting all terms $\mathrm{O}^2(\lambda\Delta t)$. The last equation in precision-dt (1.36) holds for the same reason that the prior equation (1.35) holds as long as $m > 0$. Note that $(dP)^m(t) = dP(t)$ is obviously valid for $dP(t) = 0$, but if $dP(t) \neq 0$ then division by $dP(t)$ is permissible so $(dP)^{m-1}(t) = 1$ and we must have $dP(t) = 1$, one being the only real root in this real problem. It is reasonable to assume that $\lambda dt \leq 1$ to avoid nominally violating probability bounds $\mathrm{Prob}[dP(t) = 0] \geq 0$ and $\mathrm{Prob}[dP(t) = 1] \leq 1$, even though the error is hidden in the order symbols. The rules (1.32-1.36) will come in very handy for simplifying powers of $dP(t)$ in the Poisson jump calculus later in this text. □

This **zero-one jump law** immediately leads to the following corollary for Poisson differential distribution and expectations:

Corollary 1.20. ***Zero-One Distribution and Expectation for*** $dP(t)$***:***

$$\Phi_{dP(t)}(k) = p_k(\lambda dt) \stackrel{\mathrm{dt}}{=} (1 - \lambda dt)\delta_{k,0} + \lambda dt\delta_{k,1}\,, \tag{1.37}$$

is a generalized representation of the differential Poisson distribution and

$$\mathrm{E}[f(dP(t))] \stackrel{\mathrm{dt}}{=} (1 - \lambda dt)f(0) + \lambda dt f(1)\,, \tag{1.38}$$

is the expectation, provided $f(p)$ *is a bounded and continuous function.*

The Poisson zero-one jump law is a special case of a **Bernoulli distribution**, concerning Bernoulli trials which have only two outcomes, here with failure probability $p = 1 - \lambda dt$ for zero jump or success probability $1 - p = \lambda dt$ for one jump, provided λdt is small compared to unity.

1.9 Temporal, Non-Stationary Poisson Process

Properties 1.21. ***Temporal Poisson process:***

- *For the* ***temporal or non-stationary Poisson process*** $P(t)$ *the jump rate is time dependent,* $\lambda = \lambda(t)$*, so that* $P(t)$ *is no longer simple or stationary, but non-stationary. First consider the differential process* $dP(t)$ *replacing the simple Poisson jump-count* λdt *by the time-dependent one,*

$$d\Lambda(t) \equiv \lambda(t)dt\,. \tag{1.39}$$

Letting $\Lambda(0) = 0$ *initially, then*

$$\Lambda(t) = \int_0^t \lambda(s)ds\,, \tag{1.40}$$

with increment

$$\Delta\Lambda(t) \equiv \Lambda(t+\Delta t) - \Lambda(t) = \int_t^{t+\Delta t} \lambda(s)ds\,. \tag{1.41}$$

Thus, $\Delta\Lambda(t) \sim \lambda(t)\Delta t$ *only when* $\Delta t \ll 1$*, i.e., is small.*

- *The* ***temporal Poisson distribution for the differential Poisson process*** $dP(t)$ ***remains unchanged*** *from the fixed jump rate Poisson, except for* $\lambda = \lambda(t)$ *and*

$$\begin{aligned}\Phi_{dP(t)}(k;\lambda(t)dt) &= \text{Prob}[dP(t) = k] \\ &= p_k(\lambda(t)dt) = e^{-\lambda(t)dt}\frac{(\lambda(t)dt)^k}{k!}\,,\end{aligned} \tag{1.42}$$

for $k = 0, 1, 2, \ldots$*, with* $t \geq 0$ *and temporal parameter* $\lambda(t) > 0$*.*

However, the ***Poisson distribution property*** *(1.21) of the Poisson process needs to be changed for the temporal increment process* $\Delta P(t)$ *(1.22) using the modified parameter* $\Delta\Lambda(t)$*,*

$$\begin{aligned}\Phi_{\Delta P(t)}(k;\Delta\Lambda(t)) &= \text{Prob}[\Delta P(t) = k] \\ &= p_k(\Delta\Lambda(t)) = e^{-\Delta\Lambda(t)}\frac{(\Delta\Lambda(t))^k}{k!}\,,\end{aligned} \tag{1.43}$$

for $k = 0, 1, 2, \ldots$*, with* $t \geq 0$*,* $\Delta t \geq 0$ *and temporal parameter* $\Delta\Lambda(t)$*. Thus, the temporal Poisson process is also a* ***time-inhomogeneous*** *process. The Poisson increment distribution is fundamental for the temporal Poisson process. Note that* $\Lambda(t)$ *will be nondecreasing if* $\lambda(t) > 0$ *and continuous.*

Finally, since the full temporal Poisson process $P(t)$ *is the increment* $P(t) - P(0) = P(t)$*, then it has the distribution*

$$\begin{aligned}\Phi_{P(t)}(k;\Lambda(t)) &= \text{Prob}[P(t) = k] \\ &= p_k(\Lambda(t)) = e^{-\Lambda(t)}\frac{(\Lambda(t))^k}{k!}\,,\end{aligned} \tag{1.44}$$

inherited from (1.43).

- *The **non-stationary behavior** follows from the fact that the distribution of the increment (1.43) depends on t through the parameter $\Delta\Lambda(t)$ or more simply from the **Poisson increment expectation** given in (B.51) or Table 1.2 with $\Delta\Lambda(t)$ replacing the parameter $\lambda\Delta t$,*

$$\mathrm{E}[\Delta P(t)] = \Delta\Lambda(t)\,, \tag{1.45}$$

since it will be, in general, a function of time t. Thus,

$$\mathrm{E}[P(t) - P(t_0)] = \Lambda(t) - \Lambda(t_0) = \int_{t_0}^{t} \lambda(s)ds\,.$$

*The **Poisson increment variance** must be the same as its expectation (B.51B.51–B.52),*

$$\mathrm{Var}[\Delta P(t)] = \Delta\Lambda(t). \tag{1.46}$$

However, treating the increment as an integral leads to another form

$$\begin{aligned}
\mathrm{Var}[\Delta P(t)] &= \mathrm{Var}\left[\int_t^{t+\Delta t} dP(s)\right] \\
&= \mathrm{E}\left[\left(\int_t^{t+\Delta t} dP(s) - \Delta\Lambda(t)\right)^2\right] \\
&= \mathrm{E}\left[\left(\int_t^{t+\Delta t} (dP(s) - \lambda(s)ds)\right)^2\right] \\
&= \mathrm{E}\left[\int_t^{t+\Delta t} (dP(s_1) - \lambda(s_1)ds_1)\int_t^{t+\Delta t} (dP(s_2) - \lambda(s_2)ds_2)\right] \\
&= \int_t^{t+\Delta t}\int_t^{t+\Delta t} \mathrm{E}\left[(dP(s_1) - \lambda(s_1)ds_1)(dP(s_2) - \lambda(s_2)ds_2)\right] \\
&= \int_t^{t+\Delta t}\int_t^{t+\Delta t} \mathrm{Cov}[dP(s_1), dP(s_2)]\,.
\end{aligned} \tag{1.47}$$

Since

$$\mathrm{Var}[\Delta P(t)] = \Delta\Lambda(t) = \int_t^{t+\Delta t} \lambda(s)ds,$$

noting that $dP(s_1)$ and $dP(s_2)$ are independent increments as differentials as long as $s_2 \neq s_1$. Hence, $\mathrm{Cov}[dP(s_1), dP(s_2)] \neq 0$ only if $s_2 = s_1$ when it has the value $\mathrm{Cov}[dP(s_1), dP(s_1)] = \mathrm{Var}[dP(s_1)]$. Consequently,

$$\mathrm{Cov}[dP(s_1), dP(s_2)] \overset{\text{gen}}{=} \lambda(s_1)\delta(s_1 - s_2)ds_1 ds_2 \tag{1.48}$$

for arbitrary Δt, so the inner integral of (1.47) will be

$$\int_t^{t+\Delta t} \mathrm{Cov}[dP(s_1), dP(s_2)] = \lambda(s_1)ds_1\,, \tag{1.49}$$

and (1.47) yields the same answer as (1.46).

- *The **temporal Poisson differential process distribution for** $dP(t)$ to precision-dt is*

$$\begin{aligned}\Phi_{dP(t)}(k; d\Lambda(t)) &= \text{Prob}[dP(t) = k] \\ &= p_k(d\Lambda(t)) \stackrel{\text{dt}}{=} (1 - \lambda(t)dt)\delta_{k,0} + \lambda(t)dt\delta_{k,1} \,, \quad (1.50)\end{aligned}$$

which simply follows from (1.43) for sufficiently small Δt and the corresponding simple process zero-one law result (1.37), if $\lambda(t) > 0$.

- *The **inter-jump times for the non-stationary Poisson process are exponentially distributed**. The increasing property of $\Lambda(t)$ ($d\Lambda(t) > 0$) means that it can be used as a substitute "clock" in place of t, but for $\Lambda(t)$ to be a full range clock it is necessary that $\Lambda(t)$ be unbounded, i.e., $\Lambda(t) \to +\infty$ as $t \to +\infty$. Let T_j be the jth jump time of the temporal $P(t)$ for $j \geq 1$ ($T_0 \equiv 0$ is the initial time) and $\Delta T_{j-1} \equiv T_j - T_{j-1}$ be the inter-jump time (also called inter-arrival time) for $j \geq 1$, so $T_1 = \Delta T_0$. Slightly modifying the arguments for the exponential distribution of ΔT_j for the stationary $P(t)$ in (1.24), the non-stationary distribution and its corresponding density conditioned on the most recent jump time T_{j-1} are given by:*

Theorem 1.22. *Provided that $\Lambda(t) \to \infty$ as $t \to \infty$, then the non-stationary distribution of the Poisson inter-jump time ΔT_{j-1} for $j = 1, 2, \ldots$ inter-jump times is*

$$\Phi_{\Delta T_{j-1}|T_{j-1}}(\Delta t) = 1 - \exp\left(-\int_{T_{j-1}}^{T_{j-1}+\Delta t} \lambda(t)dt\right) \quad (1.51)$$

with density

$$\phi_{\Delta T_{j-1}|T_{j-1}}(\Delta t) = \lambda(T_{j-1} + \Delta t)\exp\left(-\int_{T_{j-1}}^{T_{j-1}+\Delta t} \lambda(t)dt\right) \quad (1.52)$$

or alternatively in terms of the jump times T_j for $j \geq 1$,

$$\phi_{T_j|T_{j-1}}(t) = \lambda(t)\exp\left(-\int_{T_{j-1}}^{t} \lambda(s)ds\right). \quad (1.53)$$

Proof. *The modified stationary proof is as follows:*

$$\begin{aligned}
\Phi_{\Delta T_{j-1}|T_{j-1}}(\Delta t) &\equiv \text{Prob}[\Delta T_{j-1} \le \Delta t \,|\, T_{j-1}] \\
&= 1 - \text{Prob}[\Delta T_{j-1} > \Delta t \,|\, T_{j-1}] \\
&= 1 - \text{Prob}[\Delta P(T_{j-1}) \equiv P(T_{j-1} + \Delta t) - P(T_{j-1}) = 0 \,|\, T_{j-1}] \\
&= 1 - p_0(\Lambda(T_{j-1} + \Delta t) - \Lambda(T_{j-1})) = 1 - p_0(\Delta\Lambda(T_{j-1})) \\
&= 1 - e^{-\Delta\Lambda(T_{j-1})} = 1 - \exp\left(-\int_{T_{j-1}}^{T_{j-1}+\Delta t} \lambda(t)dt\right) \\
&= \Phi_e(\Delta\Lambda(T_{j-1}); 1)\,,
\end{aligned}$$

where $\Phi_e(\Delta\Lambda(T_{j-1}); 1)$ *is the exponential distribution (B.40) in* $\Delta\Lambda(T_{j-1})$ *with* $\Delta\Lambda(T_{j-1}) \equiv \Lambda(T_{j-1} + \Delta t) - \Lambda(T_{j-1})$ *and mean* $\mu = 1$*, i.e., still exponentially distributed but the distribution depends on* T_{j-1}*.* □

Caution*: if* $\Lambda(t)$ *is finite, then* $\Phi_{\Delta T_{j-1}|T_{j-1}}(\Delta t)$ *as derived is* ***not a proper*** *probability distribution since* $1 - \exp\left(-\int_{T_{j-1}}^{+\infty} \lambda(t)dt\right) < 1$ *with* $\Lambda(+\infty) < +\infty$*).*

- *For more general properties see Snyder and Miller [252] for extended information or Çinlar [55].*

1.10 Poisson Process Expectations Conditioned on Past

Example 1.23. ***Illustration of Independent Increments and Markov Properties for Poisson Process:***

- $\text{E}[P(t)|P(r), 0 \le r \le s] = P(\min[s,t]) + \lambda(t-s)H(t-s)$,
  where $H(X)$ *is the Heaviside step function (B.158). The techniques are similar to those for the Wiener process, except that there is no zero mean, but the mean increment is the same as the increment variance, i.e.,* $\text{E}[\Delta P(t)] = \lambda\Delta t = \text{Var}[\Delta P(t)]$*. Also,* $P(0)$ *is zero by definition with probability one.*

$$\begin{aligned}
\text{E}[P(t)|P(r), 0 \le r \le s] &= \left\{ \begin{array}{l} P(t),\ 0 \le t \le s \\ \text{E}[(P(t) - P(s)) + (P(s) - P(0))|P(r), 0 \le r \le s], \\ \qquad\qquad\qquad\qquad 0 \le s < t \end{array} \right\} \\
&= \left\{ \begin{array}{l} P(t),\ 0 \le t \le s \\ \lambda(t-s) + P(s),\ 0 \le s < t \end{array} \right\} \\
&= P(\min[s,t]) + \lambda(t-s)H(t-s).
\end{aligned}$$

When $0 \le s < t$ then the above formula symmetrized using the Poisson deviation process, $(P(t) - \lambda t)$, having zero mean, with $H(t-s) = 1$ for $s < t$, has the form

$$\mathrm{E}[P(t) - \lambda t | P(r), 0 \le r \le s] = P(s) - \lambda s, \tag{1.54}$$

signifies that for the deviation the average information conditioned on the past data, $\{P(r), r \in [0, s]\}$, is given by the most recent past deviation $P(s) - \lambda s$, which may imply a significant reduction in uncertainty for the present data, $P(t)$.

*The form of the result (1.54) is again the principal characteristic form for a **martingale** as was (1.18) with $X(t) = f(P(t))$ or (1.18) for $W(t)$, i.e., an abstract model of a fair game (see the beginning preliminary chapter of Mikosch [209] for a clear description of martingales, but in an elementary abstract presentation; martingales will be described at the end of this book).*

- $\mathrm{E}[P^2(t)|P(r), 0 \le r \le s] = P^2(\min[s,t]) + \lambda(t-s)(1+2P(s)+\lambda(t-s))H(t-s)$. *The derivation is similar to that for the conditional mean above.*

$$\begin{aligned}
\mathrm{E}[P^2(t)|P(r), 0 \le r \le s] &= \left\{ \begin{array}{l} P^2(t), \ 0 \le t \le s \\ \mathrm{E}[((P(t) - P(s)) + (P(s) - P(0)))^2 | P(r), 0 \le r \le s], \\ \qquad\qquad 0 \le s < t \end{array} \right\} \\
&= \left\{ \begin{array}{l} P^2(t), \ 0 \le t \le s \\ \mathrm{E}[(P(t) - P(s))^2] + 2P(s)\mathrm{E}[(P(t) - P(s))] + P^2(s), \\ \qquad\qquad 0 \le s < t \end{array} \right\} \\
&= \left\{ \begin{array}{l} P^2(t), \ 0 \le t \le s \\ \lambda(t-s)(1 + \lambda(t-s)) + 2P(s) \cdot \lambda(t-s) + P^2(s), \\ \qquad\qquad 0 \le s < t \end{array} \right\} \\
&= P^2(\min[s,t]) + \lambda(t-s)(1 + 2P(s) + \lambda(t-s))H(t-s).
\end{aligned}$$

Table 1.2 has to be used for $\mathrm{E}[(\Delta P)^2(s)]$ with $\Delta t = (t-s)$

Similar to the techniques used previously for the Wiener process with conditioning on the past, the general technique for powers $P^m(t)$, when $s < t$ with conditioning on $P(s)$, is to use the decomposition into independent increments $P(t) = P(s) + (P(t) - P(s))$ and then expand the power of m by the binomial expansion (B.152)

$$(P(s) + (P(t) - P(s)))^m = \sum_{k=0}^{m} \left(\begin{array}{c} m \\ k \end{array} \right) P^k(s)(P(t) - P(s))^{m-k},$$

and then use independence of the increments and conditioning to calculate for each term,

$$\begin{aligned}
&\mathrm{E}\left[\left(\begin{array}{c} m \\ k \end{array} \right) P^k(s)(P(t) - P(s))^{m-k} \middle| P(r), 0 \le r \le s \right] \\
&\qquad = \left(\begin{array}{c} m \\ k \end{array} \right) P^k(s)\mathrm{E}\left[(P(t) - P(s))^{m-k} \right],
\end{aligned}$$

relying on Table 1.2 for the remaining expectations.

1.11 Exercises

1. Show formally that

$$\phi_{dW(t)}(w) \stackrel{\mathrm{dt}}{=} \delta(w) + \frac{1}{2} dt \delta''(w) \,, \tag{1.55}$$

i.e., has a **delta-density** in the generalized sense, by showing that

$$\mathrm{E}[f(dW(t))] = \int_{\infty}^{+\infty} \phi_{dW(t)}(w) f(w) dw \stackrel{\mathrm{dt}}{=} f(0) + \frac{1}{2} dt f''(0) \,,$$

i.e., to precision-dt, neglecting terms $\mathrm{o}(dt)$. Also, show that the integral of the delta-density on the right hand side of (1.55) has the same effect as the integral of the left hand side. Assume that $f(w)$ is three times continuously differentiable and with $f(w)$ and its derivatives vanishing sufficiently at infinity.
{*Hint: Only a formal expansion of $f(w)$ should be needed here. The exponential properties of $\phi_{dW(t)}(w)$ ensure uniformity to allow expansion inside the integral, so that Laplace's or higher order asymptotic method should not be needed.*}

2. Let $\{t_i : t_{i+1} = t_i + \Delta t_i, i = 0 : n, t_0 = 0; t_{n+1} = T\}$ be a variably-spaced partition of the time interval $[0, T]$ with $\Delta t_i > 0$. Show the following properties and justify by giving a reason for every step, such as a property of the process or a property of expectations.

   (a) Let $G(t) = \mu_0 t + \sigma_0 W(t)$ and $\Delta G(t_i) \equiv G(t_i + \Delta t_i) - G(t_i)$ with μ and $\sigma_0 > 0$ constants, then show

$$\mathrm{Cov}[\Delta G(t_i), \Delta G(t_j)] = \sigma_0^2 \Delta t_i \, \delta_{i,j} \,,$$

   for $i, j = 0 : n$, where $\delta_{i,j}$ is the Kronecker delta.

   (b) Let $H(t) = \nu_0 P(t)$ and $\Delta H(t_i) \equiv H(t_i + \Delta t_i) - H(t_i)$, with $\lambda_0 > 0$ and $\nu_0 > -1$ constants, then show

$$\mathrm{Cov}[\Delta H(t_i), \Delta H(t_j)] = \nu_0^2 \lambda_0 \Delta t_i \delta_{i,j} \,,$$

   for $i, j = 0 : n$.

   (c) Let $\Delta W(t_i) \equiv W(t_i + \Delta t_i) - W(t_i)$, but $\Delta^\theta W(t_i) \equiv W(t_i + \theta \Delta t_i) - W(t_i)$ with $0 \le \theta \le 1$, then show

$$\mathrm{Cov}[\Delta W(t_i), \Delta^\theta W(t_j)] = \theta \Delta t_i \, \delta_{i,j} \,,$$

   for $i, j = 0 : n$.

3. (a) Verify the $m = 3 : 4$ entries in Table 1.1 of the text for $\mathrm{E}[|\Delta W(t)|^m]$.

   (b) Verify the $m = 3 : 4$ entries in Table 1.2 of the text for $\mathrm{E}[(\Delta P(t))^m]$ and $\mathrm{E}[(\Delta P(t) - \lambda \Delta t)^m]$.

4. (a) Show that when $0 \le s \le t$,

$$\mathrm{E}[W^3(t)|W(r), 0 \le r \le s] = W^3(s) + 3(t-s)W(s)\,,$$

justifying every step with a reason, such as a property of the process or a property of conditional expectations.

(b) Use this result to verify the martingale form (1.18)

$$\mathrm{E}[W^3(t) - 3tW(t)|W(r), 0 \le r \le s] = W^3(s) - 3sW(s)\,.$$

{*Hint: The general technique is to seek the expectation of mth power in the separable form,*

$$\mathrm{E}[M_m(W(t), t)|W(r), 0 \le r \le s] = M_m(W(s), s)\,,$$

where

$$M_m(W(t), t) = W^m(t) + \sum_{k=0}^{m-1} \alpha_k(t) W^k(t)\,,$$

satisfied for the sequence of coefficient functions $\{\alpha_0(t), \ldots, \alpha_{m-1}(t)\}$ *for the separable form, so that the conditional expectations of the lower order powers*

$$\mathrm{E}[W^k(t)|W(r), 0 \le r \le s]$$

can be recursively obtained in the order $k = 0 : m-1$.}

5. (a) Show that when $0 \le s \le t$,

$$\mathrm{E}[W^4(t)|W(r), 0 \le r \le s] = W^4(s) + 6(t-s)W^2(s) + 3(t-s)^2\,,$$

justifying every step with a reason, such as a property of the process or a property of conditional expectations.

(b) Use this result to verify the martingale form (1.18)

$$\mathrm{E}[W^4(t) - 6tW^2(t) + 3t^2|W(r), 0 \le r \le s] = W^4(s) - 6sW^2(s) + 3s^2\,,$$

together with the form for similar conditional expectation of $W^2(t)$ or that for $W^2(t) - t$.

{*See the Hint in Exercise 4 above.*}

6. Show that

$$E[(\Delta P)^m(t)] = \lambda \Delta t (1 + \mathrm{O}(\lambda \Delta t)) \tag{1.56}$$

for $\lambda \Delta t \ll 1$, by induction for $m \ge 1$.

7. Show that for the Poisson increment process, $\Delta P(t)$, the expectation can be expanded as

$$\mathrm{E}[f(\Delta P(t))] = \sum_{k=0}^{\infty} \frac{(\lambda \Delta t)^k}{k!} \Delta^k [f(0)] \,,$$

assuming that $f(p)$ is a bounded function so that the sum converges. The kth order finite difference is defined inductively such that

$$\Delta^{k+1}[f(i)] \equiv \Delta[\Delta^k[f(i)]]$$

starting from $\Delta^0[f(i)] = f(i)$ and $\Delta^1[f(i)] = \Delta[f(i)] \equiv f(i+1) - f(i)$.
{*Hint: Use the zero-step* $I_0[f(i)] \equiv f(i)$ *and one-step* $I_1[f(i)] \equiv f(i+1)$ *operators, so that* $\Delta = I_1 - I_0$ *and* $\Delta^k = (I_1 - I_0)^k$*, for which the binomial expansion can be used.*}

8. Show that the temporal Poisson process increment distribution, $p_k(\Delta\Lambda(t))$, satisfies the following differential-difference equation (DDE),

$$\frac{d}{dt}\left[p_k(\Delta\Lambda(t))\right] = \lambda(t)\left(p_k(\Delta\Lambda(t)) - p_{k-1}(\Delta\Lambda(t))\right) \,, \tag{1.57}$$

i.e., differential in t, but difference equation in k.

Show the following characteristic function (Fourier transform) formulas in the constant coefficient case, (you need only assume that the imaginary unit $i \equiv \sqrt{-1}$ is a constant with $i^2 = -1$ when integrating for the expectation or that $\zeta = i \cdot z$ can be treated the same as a real variable):

(a) for the *Gaussian process* with time-linear drift, $G(t) = \mu_0 t + \sigma_0 W(t)$, where μ_0 and $\sigma_0 > 0$ are constants,

$$C[G](z) \equiv \mathrm{E}[\exp(izG(t))] = \exp\left(iz\mu_0 t - z^2\sigma_0^2 t/2\right) \,;$$

(b) for the Poisson process, $\nu_0 P$, with constant jump rate $\lambda_0 > 0$ and constant jump amplitude ν_0,

$$C[\nu_0 P](z) \equiv \mathrm{E}[\exp(iz\nu_0 P(t))] = \exp\left(\lambda_0 t\left(\exp(iz\nu_0) - 1\right)\right) \,;$$

(c) and finally for the jump-diffusion process assuming that $W(t)$ and $P(t)$ are independent processes,

$$C[X](z) \equiv \mathrm{E}[\exp(izX(t))] = \exp\left(iz\mu_0 t - z^2\sigma_0^2 t/2 + \lambda_0 t\left(\exp(iz\nu_0) - 1\right)\right) \,.$$

9. (a) Show that when $0 \le s < t$ and constant jump rate λ_0 (*see the general result in section 1.10, but verify independently this special result*) that

$$\begin{aligned}\mathrm{E}[P^2(t)|P(r), 0 \le r \le s] = {} & P^2(s) + 2\lambda(t-s)P(s) \\ & + \lambda_0(t-s)(1 + \lambda_0(t-s)) \,,\end{aligned}$$

justifying every step with a reason for its validity.

(b) Find the time polynomials $\alpha_1(t)$ and $\alpha_0(t)$ so that

$$MP_2(t) = P^2(t) + \alpha_1(t)P(t) + \alpha_0(t)$$

is a martingale. Assume $\alpha_k(0) = 0$ for $k = 0 : 1$.
{*Remarks: The primary martingale property is that* $\mathrm{E}[X(t)|X(r), 0 \leq r \leq s] = X(s)$ *for some process* $X(t)$ *and in this case* $X(t) = f(P(t))$, *but there are also additional technical conditions to define a martingale form. Also, by a simple form of the principle of separation of variables, if* $f(t) = g(s)$ *for arbitrary values of* t *and* s, *then* $f(t) = C = g(s)$ *where* C *is a constant.*}

10. (a) Show that when $0 \leq s < t$ that

$$\begin{aligned}\mathrm{E}[P^3(t)|P(r), 0 \leq r \leq s] \\ = P^3(s) + 3\lambda(t-s)P^2(s) + 3\lambda(t-s)(1+\lambda(t-s))P(s) \\ +\lambda(t-s)(1+3\lambda(t-s)+\lambda^2(t-s)^2)\,,\end{aligned}$$

justifying every step with a reason, such as a property of the process or a property of conditional expectations.

(b) Use this result to verify the martingale form (1.18)

$$\begin{aligned}\mathrm{E}[P^3(t) - 3\lambda t P^2(t) - 3\lambda t(1-\lambda t)P(t) - \lambda t(1-3\lambda t+\lambda^2 t^2)|P(r), 0 \leq r \leq s] \\ = P^3(s) - 3\lambda s P^2(s) - 3\lambda s(1-\lambda s)P(s) - \lambda s(1-3\lambda s+\lambda^2 s^2).\end{aligned}$$

{*Hint: See the Hint in Exercise 4 in this section for* $W^3(t)$ *conditional expectation.*}

Suggested References for Further Reading

- Arnold, 1974 [13].
- Çinlar, 1975 [55].
- Gard, 1988 [91].
- Jazwinski, 1970 [154].
- Karlin and Taylor, 1981 [162].
- Klebaner, 1998 [164].
- Mikosch, 1998 [209].
- Øksendal, 1998 [222].
- Schuss, 1980 [244].
- Snyder and Miller, 1991 [252].

- Steele, 2001 [256].
- Taylor and Karlin, 1998 [265].
- Tuckwell, 1995 [270].

Chapter 2

Stochastic Integration for Diffusions

My major aim in this was to find facts which would guarantee as much as possible the existence of atoms of definite finite size.
—Albert Einstein (1879-1955) in the first of four "Annus Mirabilis" papers in the Annals der Physik during 1905, concerning Brownian motion.

Brownian motion, as described by Bachelier in 1900 and Einstein in 1905, was provided a rigorous mathematical definition by Wiener (1984-1964) in Wiener (1923, 1930) by proving the existence of an appropriate measure on a space of functions-of-time.
—Harry M. Markowitz in the forward to [245].

Jump-diffusion stochastic differential equations (SDEs) with initial conditions are of the form,

$$dX(t) = f(X(t),t)dt + g(X(t),t)dW(t) + h(X(t),t)dP(t)\ ,\ X(0) = x_0\ , \qquad (2.1)$$

where the Poisson process $dP(t)$ supplies the jumps and the Wiener process $dW(t)$ supplies the diffusion. Initial value problem (2.1), unlike the ordinary differential equations (ODEs) with initial conditions, are symbolic equations. They are not fully defined until the method of integration for solving a stochastic differential equation (SDE) is specified, given the coefficient functions $\{f(x,t), h(x,t), g(x,t)\}$. More precisely, the SDE (2.1) is not fully specified until the methods of integration for the three types of integrals in the formal integral solution,

$$X(t) = x_0 + \int_0^t f(X(s),s)ds + \int_0^t g(X(s),s)dW(s) + \int_0^t h(X(s),s)dP(s)\ , \quad (2.2)$$

with respect to t, $W(t)$ and $P(t)$, respectively, have been defined. Until then, the stochastic integral equation or SIE (2.2) is as symbolic as the SDE in (2.1), since the evaluation of the second and third integrals in (2.2) is very sensitive to the method of integration used due to the random and singular properties of $dW(t)$ and $dP(t)$. It will be necessary to re-examine the foundations for ordinary or Riemann integration to motivate the inclusion of integrands with randomness, non-smoothness and jump discontinuities contributed by the stochastic processes $W(t)$ and $P(t)$ to the state process $X(t)$. This re-examination of integration will also be useful for subsequent numerical approximations of the new definitions as well as providing a basis for new types of integrals that will arise.

In this chapter, the integrals of the second type in (2.2), i.e.,

$$\int_0^t g(X(s), s)dW(s) \ ,$$

where the integration is with respect to the diffusion process $W(t)$, will be treated primarily. However, the short treatment of ordinary integration will be sufficient for integrals of the first type, i.e.,

$$\int_0^t f(X(s), s)ds$$

where the integration is with respect to the time t and the stochastic process $X(t)$ is only in the integrand. The third type of integral will be treated in the next chapter.

When considering higher approximations or other difficult behavior in the numerical solution of ordinary differential equations, it is often necessary to work with the corresponding integral equation. Similarly, the proper form for solving stochastic differential equations (which can be, in general, considered as a symbolic concept anyway) is the exact and numerical analysis of the corresponding stochastic integral equation.

Once the foundations for stochastic integrations have been made, as they would be for ordinary integration in a good calculus course, and the definition is illustrated for few simple examples, then some simpler formal chain rules will be developed that will make calculations of integral, where possible, much easier. This chapter on stochastic integration of diffusions, and a similar one on jumps that follows, presents the basis for the stochastic differential equation models of this book. Although the level of analysis is much higher than would be expected for an applied text, it is important to have a good reference source when treating new types of problems that do not fit the current models or theories to facilitate the modification of the current theories.

2.1 Ordinary or Riemann Integration

The theory of ordinary or Riemann integration is quickly reviewed as an intermediate step to build up the treatment of stochastic integration. Let the ordinary

integral be symbolically defined as

$$I[f](t) = \int_0^t f(s)ds \ , \tag{2.3}$$

where $f(t)$ is a continuous function on $0 \le t \le T$, but continuity is really more than what would be needed in general here. For general functions f, the integral interval $[0, t]$ is partitioned into $n+1$ subintervals, $[t_i, t_{i+1}]$ of width $\Delta t_i \equiv t_{i+1} - t_i > 0$ for $i = 0 : n$, i.e., a grid of $n+2$ points such that

$$0 = t_0 < t_1 < t_2 < \cdots < t_n < t_{n+1} = t \ . \tag{2.4}$$

On each subinterval an approximation point $t_i^* \equiv t_{i+\theta_i} \equiv t_i + \theta_i \Delta t_i$ is selected with $0 \le \theta_i \le 1$ provided that the θ_is are chosen so that the t_is are distinct as in (2.4), and the area on the subinterval is approximated by the simplest geometry, a rectangle of width Δt_i and height $f_i^* \equiv f_{i+\theta_i} \equiv f(t_{i+\theta_i})$, with area $f(t_{i+\theta_i})\Delta t_i$. Next let the grid size be specified as $\delta t_n \equiv \max_{i=0:n}[\Delta t_i]$ such that $\delta t_n \to 0^+$ as $n \to \infty$ to insure that all subintervals shrink to zero in the limit as $n \to \infty$. Finally, let

$$I_n^{(\theta)}[f](t) \equiv \sum_{i=0}^{n} f_{i+\theta_i} \Delta t_i \tag{2.5}$$

be the discrete approximation of the integral and define constructively the **Riemann integral** as

$$I[f](t) = \lim_{\substack{n\to\infty \\ \delta t_n \to 0}} \left[I_n^{(\theta)}[f](t) \right] \ , \tag{2.6}$$

provided the limit exists. It is important to note that the limit is independent of θ_i, $0 \le \theta_i \le 1$.

Usually, only a constant value of θ_i is used in practice, so let $\theta_i = \theta$. Also, for simplicity, the grid partition will be assumed to be evenly spaced, so that $\Delta t_i = \Delta t$, with nodes starting at t_0 and successive nodes at $t_{i+1} = t_i + \Delta t$, but integrand approximation points at $t_{i+\theta} = t_i + \theta \Delta t$, for $i = 0 : n$. Also, $t_i = i * \Delta t$ for $i = 0 : (n+1)$. Since the step size is constant, then

$$\delta t_n = \Delta t = (t_{n+1} - t_0)/(n+1) = t/(n+1) \to 0^+ \ ,$$

as $n \to +\infty$, so the extra condition that $\delta t_n \to 0^+$ is not needed.

Fortunately, the limiting definition (2.6) does not have to be used much in ordinary calculus, but the **Riemann sum** (2.5) can be used for simply numerically approximating integrals. When $\theta = 0$ and $t_{i+\theta} = t_i$, the left hand endpoint of the ith subinterval, the numerical forward integration rule is called the **left rectangular rule** or **Euler's explicit method** or tangent-line method for ordinary differential equations. When $\theta = 1$ and $t_{i+\theta} = t_{i+1}$, the right hand endpoint of the ith subinterval, the numerical backward integration rule is called the **right rectangular rule** or implicit **backward Euler's method** for ordinary differential equations. When $\theta = 1/2$ and $t_{i+\theta} = (t_i + t_{i+1})/2$, the midpoint of the ith subinterval, the numerical

integration rule is called the **midpoint rectangular rule**, more accurate by an order of magnitude in δt_n provided $f(t)$ is sufficiently differentiable.

Since the process $W(t)$ is continuous with probability one, then integrals of composite functions $f(W(t),t)$ with respect to t can be defined by Riemann integration, i.e.,

$$\int_0^t f(W(s),s)ds = \lim_{n\to\infty}\left[\sum_{i=0}^{n} f(W(t_i),t_i)\Delta t_i\right] , \tag{2.7}$$

choosing $\theta = 0$ here, though other values would be suitable. Similarly, when the integrand is for the composite process $X(t)$ with implied dependence on the diffusion $W(t)$ and also the jump process $P(t)$ through (2.2), the integral will be defined by Riemann integration, i.e.,

$$\int_0^t f(X(s),s)ds = \lim_{n\to\infty}\left[\sum_{i=0}^{n} f(X(t_i),t_i)\Delta t_i\right] . \tag{2.8}$$

The Poisson jump process, while discontinuous, is right continuous with left limits, i.e., it is also a piece-wise continuous step function, so fits nicely in the framework of the use of forward integration, which is effectively a sequence of step-function approximations. However, the jumps are stochastic and not predictable, though once a jump is generated through simulation or observation, it will be known.

Sometimes, a deterministic integration is needed with respect to the position on the path $X(t)$. In this case, let the $f(s)ds$ in (2.3) be replaced by $f(X(s),s)dX(s)$, which could also come from the form $f(X(s),s)X'(s)ds$ provided the velocity $v(s) = X'(s)$ or $dX(s) = X'(s)ds$ exists, then this leads to the **Stieltjes integral**, or Riemann-Stieltjes integral, constructive definition:

$$\int_0^t f(X(s),s)dX(s) = \lim_{n\to\infty}\left[\sum_{i=0}^{n} f(X(t_{i+\theta}),t_{i+\theta})(X(t_{i+1}) - X(t_i))\right] , \tag{2.9}$$

provided $X(t)$ is continuous and has bounded variation [168], i.e.,

$$\sum_{i=0}^{n} |X(t_{i+1}) - X(t_i)| < B ,$$

for some constant $B > 0$ for all partitions (2.4) of $[0,t]$ and $f(X(t),t)$ is continuous. (These conditions are stronger than needed and Mikosch [209] gives weaker but more complicated conditions.) Another example is the Stieltjes form for the expectation in terms of the probability distribution $\Phi_X(x)$ in the random variable X,

$$\mathrm{E}_X[f(X)] = \int_{-\infty}^{\infty} f(x)d\Phi_X(x) ,$$

sometimes used to permit the use of more general distributions than would be possible under the usual Riemann integration conditions. The Stieltjes integration form will be modified for the stochastic integration relative to $W(t)$ in the next section.

2.2 Stochastic Integration in $W(t)$: The Foundations

As in elementary calculus, the presentation starts with a fairly simple example. The integral that forms the basis for the formulation that follows is the stochastic Stieltjes integral

$$I[W](t) = \int_0^t W(s)dW(s) \ , \tag{2.10}$$

which have a stochastic correction for the simple deterministic calculus Stieltjes integral,

$$I^{((det))}[X](t) = \int_0^t X(s)dX(s) = \frac{1}{2}\int_0^t d(X^2)(s) = \frac{1}{2}\left(X^2(t) - X^2(0)\right). \tag{2.11}$$

This follows from the ordinary calculus chain rule, $d(X^2)(s) = 2X(s)dX(s)$, for differentials, to form an exact differential.

However, in the case of the stochastic integral (2.10), $W(t)$ is a random process, is nowhere differentiable and it can be shown to have unbounded variation. Note that for even spacing $\delta t_n = \Delta t = (t-0)/(n+1)$ for $i = 0:n$, so that the expected variation, from Table 1.1, is

$$\mathrm{E}\left[\sum_{i=0}^{n} |\Delta W_i|\right] = \sum_{i=0}^{n} \sqrt{2\Delta t/\pi} = (n+1)\sqrt{2t/(\pi(n+1))} = \sqrt{2t(n+1)/\pi} \to +\infty \ ,$$

as $n \to +\infty$, so the variation must be unbounded since the expected variation must not exceed the supremum of the variation and the supremum must be unbounded as well. (See Mikosch [209] for another justification.)

In the first step in finding a constructive definition for the stochastic integral (2.10), with K. Itô [149], a left endpoint rectangular or forward integration rule ($\theta = 0$) is initially used to approximate the integral so that the independent increment property of $W(t)$ is preserved,

$$I_n^{(0)}[W](t) = \sum_{i=0}^{n} W(t_i)\Delta W(t_i) = \sum_{i=0}^{n} W_i \Delta W_i \ , \tag{2.12}$$

with W_i independent of ΔW_i as intended, where the simplifying numerical notations $W_i \equiv W(t_i)$ and $\Delta W_i \equiv \Delta W(t_i) \equiv W(t_{i+1}) - W(t_i)$ have been used. The form (2.12) is not too useful for summing or approximation, but the following two general identities are very useful:

Lemma 2.1. *Let $\{x_i | i = 0 : n+1\}$ be **any** sequence of numbers, and let $\Delta x_i = x_{i+1} - x_i$ for $i = 0 : n$, then*

$$\sum_{i=0}^{n} \Delta x_i = x_{n+1} - x_0 \ , \tag{2.13}$$

$$\sum_{i=0}^{n} x_i \Delta x_i = \frac{1}{2}\left(x_{n+1}^2 - x_0^2 - \sum_{i=0}^{n} (\Delta x_i)^2\right) \ . \tag{2.14}$$

Proof. The first identity (2.13) is trivial, since adding two successive increments cancels the common value of those increments, i.e.,

$$\Delta x_i + \Delta x_{i+1} = (x_{i+1} - x_i) + (x_{i+2} - x_{i+1}) = x_{i+2} - x_i \ .$$

Verifying the second and important identity is much easier by expanding the summand on the right hand side of (2.14) to obtain the left hand side, than vice versa:

$$\begin{aligned}
\tfrac{1}{2}\left(x_{n+1}^2 - x_0^2 - \textstyle\sum_{i=0}^n (\Delta x_i)^2\right) &= \tfrac{1}{2}\left(x_{n+1}^2 - x_0^2 - \textstyle\sum_{i=0}^n (x_{i+1} - x_i)^2\right) \\
&= \tfrac{1}{2}\left(x_{n+1}^2 - x_0^2 - \textstyle\sum_{i=0}^n (x_{i+1}^2 - 2x_i x_{i+1} + x_i^2)\right) \\
&= \tfrac{1}{2}\left(x_{n+1}^2 - x_0^2 - \textstyle\sum_{i=0}^n x_{i+1}^2 \right. \\
&\qquad \left. +2\textstyle\sum_{i=0}^n x_i x_{i+1} - \textstyle\sum_{i=0}^n x_i^2\right) \qquad (2.15) \\
&= \tfrac{1}{2}\left(x_{n+1}^2 - x_0^2 - \left(x_{n+1}^2 + \textstyle\sum_{j=0}^n x_j^2 - x_0^2\right)\right. \\
&\qquad \left. +2\textstyle\sum_{i=0}^n x_i x_{i+1} - \textstyle\sum_{i=0}^n x_i^2\right) \\
&= \textstyle\sum_{i=0}^n x_i \Delta x_i \ ,
\end{aligned}$$

where

$$\sum_{i=0}^{n} x_{i+1}^2 = \sum_{j=1}^{n+1} x_j^2$$

has been transformed by **change of index** to combine with a similar sum. □

The benefit of the form (2.14) when used as $x_i = W_i$, then the end points are explicitly given by $W_{n+1} = W(t)$ and $W_0 = 0$ with probability one, so the discrete approximation to stochastic integral of $W(t)$ becomes

$$I_n^{(0)}[W](t) = \frac{1}{2}(W^2(t) - \sum_{i=0}^{n} (\Delta W_i)^2) \ . \qquad (2.16)$$

Using Table 1.1 again, the expectation of $I_n^{(0)}[W](t)$ is

$$\mathrm{E}\left[I_n^{(0)}[W](t)\right] = \frac{1}{2}(t - \sum_{i=0}^{n} \Delta t_i) = \frac{1}{2}(t - t) = 0 \ ,$$

returning to more general spacing Δt_i, where the (2.13) identity $\sum_{i=0}^n \Delta t_i = t_{n+1} - t_0 = t$ has also been used. This result suggests that a reasonable form for the stochastic integral (2.10) corresponds to ($\approx$)

$$I[W](t) \approx \frac{1}{2}(W^2(t) - t) \ , \qquad (2.17)$$

where the term $(-\frac{1}{2}t)$ is the correction to the ordinary calculus or Riemann integration answer. However, since the proposed answer is not a true equality, another

condition is appropriate for the stochastic nature of the problem and that condition is the mean square limit or mean square convergence:

Definition 2.2. ***Mean Square Limit or Convergence:***

The random variable $I_n^{(0)}(t)$ ***converges in the mean square*** *to the random variable* $I(t)$ *if*

$$\mathrm{E}\left[\left(I_n^{(0)}(t) - I(t)\right)^2\right] \to 0 \tag{2.18}$$

as $n \to \infty$*, assuming that both random variables have bounded mean squares, i.e.*

$$\mathrm{E}\left[(I_n^{(0)})^2(t)\right] < \infty \quad \text{and} \quad \mathrm{E}\left[I^2(t)\right] < \infty \,.$$

If the limit (2.18) exists, then denote the ***mean square limit*** *as*

$$I(t) = \operatorname*{ms\,lim}_{n\to\infty}\left[I_n^{(0)}(t)\right] \,.$$

As an abbreviation, sometimes "$\overset{ims}{=}$*" will be used for "*$= \lim_{n\to\infty}^{\mathrm{ms}}$*", where "*$\overset{ims}{=}$*" means "Itô mean square equals".*

Some related **general stochastic convergence principles**:

Definition 2.3. ***Convergence in Probability:***
The random variable $I_n^{(0)}(t)$ ***converges in probability*** *to the random variable* $I(t)$ *if for every* $\epsilon > 0$*,*

$$\mathrm{Prob}\left[\left|I_n^{(0)}(t) - I(t)\right| \geq \epsilon\right] \to 0 \tag{2.19}$$

as $n \to \infty$*. If the limit (2.19) exists, then denote the* ***limit in probability*** *as*

$$I(t) = \operatorname*{\overset{prob}{lim}}_{n\to\infty}[I_n^{(0)}(t)] \,.$$

Definition 2.4. ***Convergence in Mean:***
The random variable $I_n^{(0)}(t)$ ***converges in the mean*** *to the random variable* $I(t)$ *if for every* $\epsilon > 0$*,*

$$\mathrm{E}\left[\left|I_n^{(0)}(t) - I(t)\right|\right] \to 0 \tag{2.20}$$

as $n \to \infty$*. If the limit (2.20) exists, then denote the* ***limit in the mean*** *as*

$$I(t) = \operatorname*{mean\,lim}_{n\to\infty}[I_n^{(0)}(t)] \,.$$

Theorem 2.5. ***Convergence in Mean Square*** $\Longrightarrow$ ***Convergence in Probability:***

$$I(t) = \lim_{n\to\infty}^{\text{ms}} [I_n^{(0)}(t)] \quad \Longrightarrow \quad I(t) = \lim_{n\to\infty}^{prob} [I_n^{(0)}(t)] \; . \tag{2.21}$$

Similarly:
Convergence in Mean Square $\Longrightarrow$ ***Convergence in Mean:***

$$I(t) = \lim_{n\to\infty}^{\text{ms}} [I_n^{(0)}(t)] \quad \Longrightarrow \quad I(t) = \lim_{n\to\infty}^{\text{mean}} [I_n^{(0)}(t)] \; . \tag{2.22}$$

Proof. Let $\epsilon > 0$. Tacitly the mean square expectation of the limit $I(t)$ and the approximation is assumed as conditions for mean square convergence, which implies that $E[|I(t)-I_n^{(0)}(t)|^2] \to 0^+$ as $n \to \infty$. The theorem follows from the **Chebyshev inequality** (B.191) of Exercise 4 on Page B71 which is written in a simplified but convenient form,

$$\text{Prob}[|X| \geq \epsilon] \leq \text{E}[|X|^2]/\epsilon^2 \; , \tag{2.23}$$

where $\epsilon > 0$. Let $X = I(t) - I_n^{(0)}(t)$ and thus

$$\text{E}[|I(t) - I_n^{(0)}(t)|^2] \geq \epsilon^2 \text{Prob}[|I(t) - I_n^{(0)}(t)| \geq \epsilon] \; .$$

Hence, as $n \to \infty$, $\text{Prob}[|I(t) - I_n^{(0)}(t)| \geq \epsilon] \to 0^+$ by being squeezed from above by the mean square deviation as it goes to zero, i.e., $I_n^{(0)}(t) \to I(t)$ in probability if $I_n^{(0)}(t) \to I(t)$ in the mean square.

Similarly, the Schwartz (Cauchy-Schwartz) inequality (B.192) of Exercise 5 on Page B72, truncated to one variable,

$$\text{E}^2[X] \leq \text{E}[X^2]$$

can be used to show that convergence in the mean square implies convergence in the mean, i.e., $I_n^{(0)}(t) \to I(t)$ in the mean if $I_n^{(0)}(t) \to I(t)$ in the mean square. □

The expectation of the proposed random variable answer is

$$\text{E}\left[I[W](t)\right] = \text{E}\left[\frac{1}{2}(W^2(t) - t)\right] = \frac{1}{2}(t - t) = 0 \; ,$$

the same as for the approximation.

In order to focus on the crucial term and to simplify the demonstration of the mean square limit, which is conjectured to be t, consider the following lemma:

Lemma 2.6. *Let*

$$J_n^{(0)}(t) \equiv \sum_{i=0}^{n} (\Delta W_i)^2 \; , \tag{2.24}$$

then

$$t = \operatorname*{ms\,lim}_{n\to\infty} [J_n^{(0)}(t)] \; . \tag{2.25}$$

Proof. The mean t of $J_n^{(0)}(t)$ is absorbed into the summation by (2.13) with $x_i = t_i$, the square of the mean square argument leads to a double sum which is separated into diagonal parts ($j = i$) and off-diagonal parts ($j \neq i$), allowing the splitting of the expectations using the independent increment property, so

$$\begin{aligned}
\mathrm{E}\left[\left(J_n^{(0)}(t) - t\right)^2\right] &= \mathrm{Var}\left[J_n^{(0)}(t)\right] \\
&= \mathrm{E}\left[\left(\sum_{i=0}^{n} (\Delta W_i)^2 - t\right)^2\right] \\
&= \mathrm{E}\left[\left(\sum_{i=0}^{n} \left((\Delta W_i)^2 - \Delta t_i\right)\right)^2\right] \\
&= \mathrm{E}\left[\sum_{i=0}^{n} \left((\Delta W_i)^2 - \Delta t_i\right) \sum_{j=0}^{n} \left((\Delta W_j)^2 - \Delta t_j\right)\right] \\
&= \sum_{i=0}^{n} \mathrm{E}\left[\left((\Delta W_i)^2 - \Delta t_i\right)^2\right] \\
&\qquad + \sum_{i=0}^{n} \mathrm{E}\left[(\Delta W_i)^2 - \Delta t_i\right] \sum_{\substack{j=0 \\ j\neq i}}^{n} \mathrm{E}\left[(\Delta W_j)^2 - \Delta t_j\right] \\
&= \sum_{i=0}^{n} \mathrm{Var}\left[(\Delta W_i)^2\right] + 0 \cdot 0 = \sum_{i=0}^{n} \left(\mathrm{E}\left[(\Delta W_i)^4\right] - E^2\left[(\Delta W_i)^2\right]\right) \\
&= \sum_{i=0}^{n} \left(3(\Delta t_i)^2 - (\Delta t_i)^2\right) = 2\sum_{i=0}^{n} (\Delta t_i)^2 \; ,
\end{aligned}$$

the last couple of steps relying on the results of Table 1.1. Since $\Delta t_i \leq \delta t_n = \max_j[\Delta t_j]$, then

$$\mathrm{E}\left[\left(J_n^{(0)}(t) - t\right)^2\right] = 2\sum_{i=0}^{n} (\Delta t_i)^2 \leq 2\delta t_n \sum_{i=0}^{n} \Delta t_i = 2t\delta t_n \to 0$$

as $n \to \infty$ showing that

$$t = \operatorname*{ms\,lim}_{n\to\infty} [J_n^{(0)}(t)] \; .$$

Clearly both $J_n^{(0)}(t)$ and t have bounded mean squares for bounded t. Hence, $J_n^{(0)}(t) = I_n^{(0)}[dW](t)$, in our functional notation. □

Lemma 2.7.

$$\frac{1}{2}\left(W^2(t)-t\right) = \operatorname*{ms\,lim}_{n\to\infty}\left[I_n^{(0)}[W](t)\right] \tag{2.26}$$

where $t<\infty$ *and*

$$I_n^{(0)}[W](t) = \sum_{i=0}^{n} W_i \Delta W_i \ .$$

Proof. Note that

$$\mathrm{E}[((W^2(t)-t)/2)^2] = \mathrm{E}[W^4(t)-2tW^2(t)+t^2]/4 = (3t^2-2t^2+t^2)/4 = t^2/2 \ ,$$

again using the convenient Table 1.1, so $(W^2(t)-t)/2$ has a bounded mean square so long at t is bounded. Similarly, one can show that $I_n^{(0)}[W](t)$ has a bounded mean square. The mean square convergence of $I_n^{(0)}[W](t)$ is obvious since $J_n^{(0)}(t)$ converged in the mean square to t and $J_n^{(0)}(t)$ is the only term that depends on the grid variable n. In fact,

$$\mathrm{E}\left[\left(I[W](t)-I_n^{(0)}[W](t)\right)^2\right] = \frac{1}{4}\mathrm{E}\left[\left(t-J_n^{(0)}(t)\right)^2\right] \to 0^+ \ ,$$

as $n\to\infty$, so converges for the same reason that $J_n^{(0)}(t)$ did in the mean square. This mean square relation follows due to the affine difference in forms $I_n^{(0)}[W](t) = (W^2(t)-J_n^{(0)}(t))/2$ in (2.16) with (2.24) and $I[W](t) \overset{ims}{=} (W^2(t)-t)/2$ in (2.17), no longer a proposed answer. In more general cases the decomposition of $I_n^{(0)}[W](t)$ will not be so simple as that between $I_n^{(0)}[W](t)$ and the part $J_n^{(0)}(t)$. □

Definition 2.8. *Denote the* ***Itô mean square (ims) limit stochastic integral*** *corresponding to the stochastic integral form*

$$I[g](t) = \int_{t_0}^{t} g(W(s),s)dW(s)$$

with associated forward integration (left rectangular rule or Euler's method) approximation

$$I_n^{(0)}[g](t) \equiv \sum_{i=0}^{n} g(W(t_i),t_i)(W(t_{i+1})-W(t_i))$$

by

$$I^{(ims)}[g](t) = \operatorname*{ms\,lim}_{n\to\infty}\left[I_n^{(0)}[g](t)\right] \tag{2.27}$$

where $0\le t_0\le t$*, assuming the integrand process* $g(W(t),t)$ *has a bounded mean integral of its square, i.e.,*

$$\mathrm{E}\left[\int_{t_0}^{t} g^2(W(s),s)ds\right] < \infty \ ,$$

and the grid partitioning satisfies

$$0 \le t_0 < t_1 < \cdots < t_{n+1} = t \tag{2.28}$$

with

$$\delta t_n = \max_{i=0:n}[\Delta t_i \equiv t_{i+1} - t_i] \ll 1$$

as $n \to \infty$.

Provided the Itô mean square limit (2.27) exists,

$$I[g](t) \stackrel{ims}{=} I^{(ims)}[g](t) \; . \tag{2.29}$$

In addition, the definition holds, since the independent increments property remains valid in a more general case, namely, if the function g depends on the past and present history of the Wiener process,

$$\mathcal{W}(t) = \{W(r), 0 \le r \le t\} \; ,$$

i.e., $g = g(\mathcal{W}(t), t)$*, in which case, g is called* ***non-anticipatory*** *or* ***adapted*** *to the process set* $\mathcal{W}(t)$.

Remarks 2.9.

- *For most of the sequel, general functions with dependence on* $W(t)$ *and* t*, i.e.,* $g(W(t), t)$*, will be used in stochastic diffusion integrals, but the reader can easily extend results to functions of the type* $g(\mathcal{W}(t), t)$ *adapted to* $\mathcal{W}(t)$.

- *If the Itô mean square limit (2.27),*

  $$I_n^{(0)}[g](t) \to I[g](t) = I^{(ims)}[g](t)$$

  in the mean square *as* $n \to \infty$*, exists, then by Theorem 2.5*

  $$I_n^{(0)}[g](t) \to I[g](t)$$

  in probability *as* $n \to \infty$.

- *In our notation,* $I[g](t) = I^{(ims)}[g](t)$ *denotes the mean square limit of the Itô forward integration approximation* $I_n^{(0)}[g](t)$ *with* $\theta = 0$ *meaning that the integral g is evaluated at* t_i *on the* i*th step. They denote particular evaluations or approximations of the purely symbolic* $I[g](t)$ *integral representation which can also have other evaluations using other integration rules with values of* θ *or using other rules relying on non-rectangular approximations.*

Thus, summarizing the results for the crucial simple example when $g(W(t), t) = W(t)$ is the following theorem:

Theorem 2.10. ***Itô Fundamental Mean Square Stochastic Integrals:***

$$\int_0^t (dW)^2(s) \stackrel{ims}{=} t \; . \tag{2.30}$$

and

$$\int_0^t W(s)dW(s) \overset{ims}{=} I^{(ims)}[W](t) = \frac{1}{2}\left(W^2(t) - t\right) \ . \tag{2.31}$$

Sketch of Proof. Some more heuristic justification is given here.

- In ordinary deterministic integral calculus, the symbol $\int_0^t (dx)^2(s)$ would be considered nonsense, but in Itô stochastic integration the symbol

$$\int_0^t (dW)^2(s) \overset{ims}{=} \operatorname*{ms\,lim}_{n\to\infty} \left[\sum_{i=0}^{n} (\Delta W)^2(t_i)\right] = t \ ,$$

  makes perfect sense, since the Itô mean square (IMS) limit is well defined and leads to the Itô correction to the ordinary calculus rule for the differential of $x^2(t)$, i.e., $x(t)dx(t) = \frac{1}{2}d(x^2)(t)$.

- In fact, this leads to a corresponding symbolic **Itô mean square** "$\overset{ims}{=}$" version for differentials,

$$(dW)^2(t) \overset{\mathrm{dt}}{=}_{ms} dt \ . \tag{2.32}$$

  and

$$W(t)dW(t) \overset{\mathrm{dt}}{=}_{ms} \frac{1}{2}(d(W^2)(t) - dt). \tag{2.33}$$

  Formally, we might rewrite (2.33) with the symbol "$\overset{\mathrm{dt}}{=}_{ms}$" for "equals in dt-precision mean square", or simply "$\overset{\mathrm{dt}}{=}$" for in "dt-precision", denoting a commutative operation,

$$d(W^2)(t) \overset{\mathrm{dt}}{=} 2W(t)dW(t) + dt \ . \tag{2.34}$$

  Using the formal increment definition of the differential (1.3), $dW(t) \equiv W(t+dt) - W(t)$ or the alternate form $W(t+dt) = W(t) + dW(t)$, then a quick calculation leads to

$$\begin{aligned} d(W^2)(t) &\equiv W^2(t+dt) - W^2(t) = (W+dW)^2(t) - W^2(t) \\ &= (W^2 + 2WdW + (dW)^2 - W^2)(t) \\ &\overset{\mathrm{dt}}{=}_{ms} 2WdW(t) + dt \ , \end{aligned} \tag{2.35}$$

  using a little bit of algebra and the symbolic fact that $(dW)^2 \overset{\mathrm{dt}}{=}_{ms} dt$, formally justifying (2.34), demonstrating a fast technique that would be useful when fast answers are needed.

□

Remarks 2.11.

- *The Itô mean square result symbolized by* $(dW)^2(t) \overset{\text{dt}}{=\!=}_{ms} dt$ *represents a remarkable paradox, since the differential* $(dW)^2(t)$ *is deterministic because* dt *is deterministic, but* $dW(t)$ *is stochastic or random.*

- *In the deterministic continuously differential case, the corresponding quadratic of a differential,* $(dx)^2(t)$ *would be negligible relative to terms of order* dt*. If the integral of such a term were consider the limit of its finite difference approximation would be zero:*

$$\begin{aligned}\int_0^t (dx)^2(s) &= \lim_{n\to\infty}\left[\sum_0^n (\Delta x_i)^2\right] \\ &= \lim_{n\to\infty}\left[\sum_0^n (x_{i+1}-x_i)\Delta x_i\right] \\ &= \lim_{n\to\infty}\left[\sum_0^n x_{i+1}\Delta x_i\right] - \lim_{n\to\infty}\left[\sum_0^n x_i \Delta x_i\right] \\ &= \lim_{n\to\infty}\left[I_n^{(1)}[x](t)\right] - \lim_{n\to\infty}\left[I_n^{(0)}[x](t)\right] \\ &= I[x](t) - I[x](t) = 0,\end{aligned}$$

  since the regular integral of $\int_0^t x(s)dx(s)$ *is independent in the limit of the particular approximation parameter used, whether* $\theta = 1$ *or* $\theta = 0$ *as in the above final lines.*

- *See also Exercise 1 which is to demonstrate that the the density,* $\phi_{dW(t)}(w)$*, for* $dW(t)$ *is the sum of two delta functions in the generalized sense that considerably constrains functions of* $dW(t)$*.*

- *Computational confirmation of the Itô's fundamental mean square stochastic integrals is the subject of Exercise 3 for the* $(dW)^2(t)$ *integrand in (2.30) and Exercise 4 for the* $(WdW)(t)$ *integrand in (2.31). For example, Fig. 2.1 is an illustration of the computational confirmation of the Itô fundamental forward integration approximating sum*

$$\int_0^t (dW)^2(s) \overset{ims}{=} t \simeq \sum_{i=0}^n (\Delta W_i)^2 \ ,$$

  with $n = 10^4$*. The confirmation is remarkable considering it is a pointwise comparison of the approximating sum with the exact Itô answer* t*, and not a demonstration of convergence in the Itô mean square limit. The sample size has to be sufficiently large, else the approximating sum tends away from the*

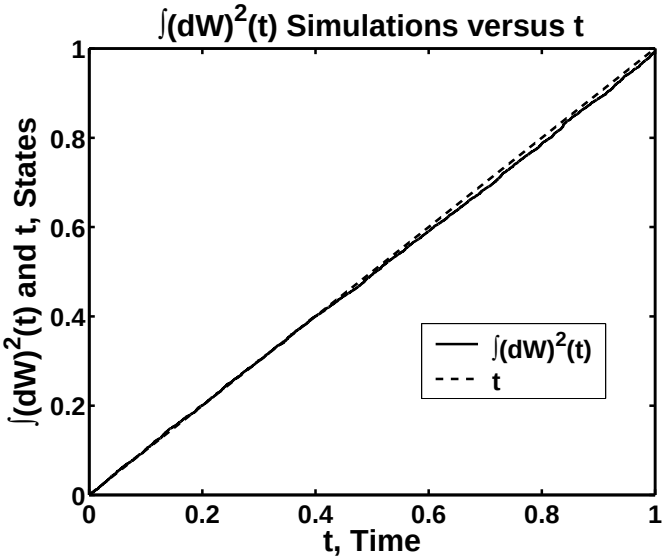


Figure 2.1. *Simulated sample path for the Itô forward integration approximating sum of* $\int (dW)^2(t) \stackrel{ims}{=} t \simeq \sum_i (\Delta W_i)^2$ *for* $n = 10^4$ *MATLAB* `randn` *sample size.*

t answer due to the slope of the tangent line bias, that is also a feature of deterministic ODE applications of Euler's method.

- *The general code for simulating the stochastic diffusion integral with integrand* $g(W(t), t)$ *by the Itô forward integration approximation*

$$I[g](t) = \int_{t_0}^{t} g(W(s), s)ds \simeq \sum_{i=0}^{n} g_i \Delta W_i \ ,$$

in an abbreviated MATLAB fragment might be

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function intdwdw
% Example MATLAB code for integral of (dW)^2.
clc % clear variables;
t0 = 0.0; tf = 1.0;
n = 1.0e+4; nf = n + 1; % set time grid: (n+1) subintervals
dt = (tf-t0)/nf;  %       and (n+2) points;
% replace these particular values according the application;
t(1) = t0; % set initial time at i = 1 for MATLAB;
W(1) = 0.0; % set initial diffusion noise condition;
sqrtdt = sqrt(dt); % dW(i) noise time scale so E[dW] = 0;
sumdw2(1) = 0.0; % set initial sum variable;
```

```
kstate = 1; randn('state',kstate); % Set randn state
%    for repeatability;
dW = sqrtdt*randn(nf,1); % simulate (n+1)-dW(i)'s sample;
t = t0:dt:tf; % get time vector t;
for i = 1:nf % simulate integral sample path.
    W(i+1) = W(i) + dW(i); % sum diffusion noise;
    sumdw2(i+1) = sumdw2(i) + (dW(i))^2; % sum whole integrand;
end
plot(t,sumdw2,'k-',t,t,'k--','LineWidth',2); % plot sum;
title('\int(dW)^2(t) Simulations versus t');
ylabel('\int(dW)^2(t) and t, States');
xlabel('t, Time');
legend('\int(dW)^2(t)','t',0);
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

- *The form for the simulation of the Wiener increment process $\Delta W(t)$ by a standard normal distribution Z scaled by $\sqrt{\Delta t}$ in the above code fragment is based upon the following change of variables (or change of measure) result, showing that both $\Delta W(t)$ and $\sqrt{\Delta t}Z$ have the same distribution:*

Theorem 2.12. Wiener Simulations by Standard Normal:
Let Z be a random variable with a standard normal distribution, $\Phi_Z(z) = \Phi_n(z;0,1)$, then

$$\Phi_{\Delta W(t)}(w) = \Phi_{\sqrt{\Delta t}Z}(w)\,, \tag{2.36}$$

where $\Delta t > 0$.

Proof. *From properties of the normal distribution,*

$$\Phi_Z(z) = \text{Prob}[Z \le z] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-y^2/2}dy$$

and

$$\begin{aligned}\Phi_{\Delta W(t)}(w) = Prob[\Delta W(t) \le w] &= \frac{1}{\sqrt{2\pi\Delta t}}\int_{-\infty}^{w} e^{-v^2/(2\Delta t)}dv \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w/\sqrt{\Delta t}} e^{-y^2/2}dy = Prob\left[z \le w/\sqrt{\Delta t}\right] \\ &= \text{Prob}\left[\sqrt{\Delta t}Z \le w\right] = \Phi_{\sqrt{\Delta t}Z}(w)\,,\end{aligned}$$

since $\text{Prob}[aZ \le w] = \text{Prob}[Z \le w/a]$ *provided* $a > 0$. □

- *See also the full version of this MATLAB code in Section C.11 of the Appendix C for actual type-set figure.*
- *See also Fig. 4.1 in Chapter 4 illustrating the application to* $g(W(t),t) = \exp(W(t) - t/2)$ *that yields an exact differential in the Itô mean square sense.*
- *Computational simulation is another way to get fast answers when they are needed.*

However, the Itô stochastic integration of exact differentials is easy as the following theorem shows.

Theorem 2.13. ***Fundamental Theorem of Itô Stochastic Diffusion Calculus:***
Let $g(w)$ *be continuous and* $G(w)$ *be continuously differentiable. Then*

(a)

$$d\left[\int_0^t g(W(s))dW(s)\right] \stackrel{ims}{=} g(W(t))dW(t) \tag{2.37}$$

and

(b)

$$\int_0^t dG(W(s)) \stackrel{ims}{=} G(W(t)) - G(0)\ , \tag{2.38}$$

for $0 \le t$.

Proof. The first part of the fundamental theorem (a) benefits from the Itô forward integration approximation and continuity of g, but mostly from the continuity of W. Consider the increment version for sufficiently small increments Δt,

$$\begin{aligned}\Delta\left[\int_0^t g(W(s))dW(s)\right] &= \left(\int_0^{t+\Delta t} - \int_0^t\right) g(W(s))dW(s)\\ &= \int_t^{t+\Delta t} g(W(s))dW(s)\\ &\simeq g(W(t))\Delta W(t)\\ &\to g(W(t))dW(t)\end{aligned}$$

as $\Delta t \to 0^+$, using the continuity of both g and W.

For second part of the fundamental theorem (b), using the Itô stochastic

integration Definition 2.8,

$$\begin{aligned}\int_0^t dG(W(s)) &\overset{ims}{=} \operatorname*{ms\,lim}_{n\to\infty}\left[\sum_{i=0}^{n}(G(W(t_{i+1})) - G(W(t_i)))\right] \\ &= \operatorname*{ms\,lim}_{n\to\infty}\left[\sum_{i=0}^{n}(\Delta G(W(t_i)))\right] = \operatorname*{ms\,lim}_{n\to\infty}[G(W(t_{n+1})) - G(W(t_0))] \\ &= \operatorname*{ms\,lim}_{n\to\infty}[G(W(t)) - G(0)] = G(W(t)) - G(0)\ ,\end{aligned}$$

upon using the facts that $t_0 = 0$, $t_{n+1} = t$, and for any sum over all increments is the total increment from (2.13) of Lemma 2.1. Assuming that $G(W(t))$ is bounded on $[0, t]$ should be all that is needed. Thus, for exact derivatives, Itô stochastic integration and ordinary deterministic or Riemann integration agree. See Kolmogorov and Fomin [168] or Protter [232] about the importance of bounded variation as well, but these details are beyond the scope of this book. □

Remarks 2.14.

- *The first part (a) relates the integral to the differential formulation and the second part (b) is useful since it is one of the main ways of finding stochastic integrals which are not often found in closed form. Usually, part (b) is used to reduce a more complicated stochastic integral to a closed form plus a simpler, perhaps Riemann, integral.*
- *Note that in the proof of part (a), there is a difference in the exact* ***increment of an integral*** *and its approximate increment for small Δt. Using a more general form in some process $X(t)$ for the integral, the exact increment has the form*

$$\Delta[I[G]](t) \equiv I[G](t + \Delta t) - I[G](t) = \int_t^{t+\Delta t} G(X(s), s)dX(s)$$

that holds for arbitrary Δt as long as the integral can be defined, while the approximate integral has the form

$$\Delta[I[G]](t) \simeq G(X(t), t)\Delta X(t)\ ,$$

for sufficiently small Δt. So which form is used in an application depends on the application and the size of the time increment Δt.

When dealing with Itô stochastic integrals more general functions of Markov stochastic processes such as $g(W(t))$, $g(W(t), t)$ or $g(X(t), t)$, where $X(t)$ may itself be a stochastic process that is a functional of $W(t)$ and also $P(t)$, some more information may be needed. In particular, some more assumptions or some theorems beyond the scope of this applied book may be be needed to demonstrate the mean

square convergence of the stochastic integrals. Typically, the usual assumptions [13, 165, 209, 232] require that the integrand function, say $Y(t) = g(X(t), t)$, has a piece-wise-constant, right-continuous approximation that is compatible with the Itô forward summation approximation and that permits satisfaction of the mean square limit criterion. Such assumptions are unnecessary when there is a explicit function of $W(t)$ since, as will be seen, the mean square limit property can be verified directly. However, when a general function is considered with little information then this extra piece-wise constant assumption will be necessary.

Assumption 2.15. ***Piece-Wise-Constant Approximations (i-PWCA) in the Itô Sense for General Mean Square Limits***

1. *Let $Z(t)$ be a* ***piece-wise-constant***, ***right-continuous*** *stochastic process such that*

$$Z_n(s) = \{\zeta_i : \ \tau_i \le s < \tau_{i+1}; \ \text{for } i = 0 : n\} \ , \tag{2.39}$$

   where the times τ_i belong to a partition of $[0, t]$ such that $\tau_0 = 0$ and $\tau_{n+1} = t$, so $Z_n(t) = \zeta_{n+1}$ if needed, but does not contribute to the integral. The ζ_i are a sequence of discrete stochastic processes depending on the past Wiener processes $\mathcal{W}_i = \{W(s) \mid 0 \le s \le \tau_i\}$, i.e., ***adapted to $\mathcal{W}_i$ for $i = 0 : n + 1$, in the Itô sense****. Let $\mathcal{W}$ be the set of all $\mathcal{W}_i$.*

2. *Let $Y(t)$ be a stochastic process depending on $\mathcal{W}$ where $Y(s)$ can be approximated in the Itô sense by the piece-wise-constant, right-continuous stochastic process $Z_n(s)$ for $0 \le s \le t$ such that*

$$\lim_{n\to\infty} \mathrm{E}\left[\int_0^t \left(Y(s) - Z_n(s)\right)^2 ds\right] \to 0 \tag{2.40}$$

   as $n \to +\infty$.

Remark 2.16. *An example of an approximation by i-PWCA is the function on $[t_0, t]$,*

$$\mathcal{G}_n(s) = \{g(W_i, t_i) : \ t_i \le s < t_{i+1}; \ \text{for } i = 0 : n\} \ , \tag{2.41}$$

where $g(w, t)$ is a continuous function of (w, t), $W_i = W(t_i)$ and the set

$$\{t_i : \ t_{i+1} = t_i + \Delta t_i, \ \Delta t_i > 0, \ \text{for } i = 0 : n; \ t_{n+1} = t\} \tag{2.42}$$

is the time partition of $[t_0, t]$.

Theorem 2.17. ***Mean of Itô Stochastic Integral:***

$$\mathrm{E}\left[\int_{t_0}^t g(W(s), s) dW(s)\right] \overset{ims}{=} 0 \ , \quad 0 \le t_0 \le t \ , \tag{2.43}$$

assuming the mean square integrability condition

$$\mathrm{E}\left[\int_{t_0}^{t} g^2(W(s),s)ds\right] < \infty \,, \tag{2.44}$$

and the i-PWCA Mean Square Limits Assumption 2.15 for $Y(t) = g(W(t),t)$.

Proof. Only heuristic justification will be given here to keep this presentation simple. For more elaborate justification using sequences of approximate step function sums, consult the works of Arnold [13], Schuss [244],Øksendal [222], Mikosch [209] or Steele [256].

- Using the Itô mean square limit (2.27), then we have the formal finite sum approximation using partition (2.42),

$$\int_{t_0}^{t} g(W(s),s)dW(s) \simeq \sum_{i=0}^{n} g(W(t_i),t_i)(W(t_{i+1}) - W(t_i)) = \sum_{i=0}^{n} g_i \Delta W_i \,,$$

  where $g_i = g(W(t_i),t_i)$ from $\mathcal{G}_n(s)$ (2.41) and $\Delta W_i = W(t_{i+1}) - W(t_i)$. Since the right hand side sum is finite, the operations of expectation and summation can be interchanged, so

$$\begin{aligned}\mathrm{E}\left[\int_{t_0}^{t} g(W(s),s)dW(s)\right] &\simeq \sum_{i=0}^{n} \mathrm{E}[g_i \Delta W_i] = \sum_{i=0}^{n} \mathrm{E}[g_i]\mathrm{E}[\Delta W_i] \\ &= \sum_{i=0}^{n} \mathrm{E}[g_i] \cdot 0 = 0 \,,\end{aligned}$$

  the last line using the independent increments and zero mean properties.

  Note that the forward integration rule of Itô is not used, then the mean zero result of (2.43) will not be true in general (see Exercise 10 on page 62 for a θ-rule counterexample).

- The final justification requires justifying the interchange of the expectation operator, a Riemann integral, and the mean square limit operator. The underlying integrability assumption can be rewritten using Itô's forward integration choice leads to the approximation,

$$\mathrm{E}\left[\int_{t_0}^{t} g^2(W(s),s)ds\right] = \int_{t_0}^{t} \mathrm{E}\left[g^2(W(s),s)\right] ds \simeq \sum_{i=0}^{n} \mathrm{E}[g_i^2]\Delta t_i \,.$$

- This approximation can be compared with the expected absolute value of original Itô approximated sum of interest followed by a one-component Schwarz's inequality $\left(\overset{\text{csi}}{\leq}\right)$, to put it into a usable quadratic form and rearrangement

into independent increments $\left(\overset{ind}{=}_{inc}\right)$,

$$\begin{aligned}
&\mathrm{E}\left[\left|\int_{t_0}^{t} g(W(s),s)dW(s)\right|\right] \simeq \mathrm{E}\left[\left|\sum_{i=0}^{n} g_i \Delta W_i\right|\right] \\
&\overset{\text{csi}}{\leq} \sqrt{\mathrm{E}\left[\sum_{i=0}^{n} g_i \Delta W_i \cdot \sum_{j=0}^{n} g_j \Delta W_j\right]} \\
&= \sqrt{\mathrm{E}\left[\sum_{i=0}^{n} g_i^2 (\Delta W_i)^2 + \sum_{i=0}^{n} g_i \Delta W_i \left(\sum_{j=0}^{i-1} + \sum_{j=i+1}^{n}\right) g_j \Delta W_j\right]} \\
&\overset{ind}{=}_{inc} \left[\sum_{i=0}^{n} \mathrm{E}\left[g_i^2\right] \mathrm{E}\left[(\Delta W_i)^2\right]\right. \\
&\left. + \sum_{i=0}^{n} \left(\sum_{j=0}^{i-1} \mathrm{E}\left[g_i g_j \Delta W_j\right] \mathrm{E}\left[\Delta W_i\right] + \sum_{j=i+1}^{n} \mathrm{E}\left[g_i g_j \Delta W_i\right] \mathrm{E}\left[\Delta W_j\right]\right)\right]^{0.5} \\
&= \sqrt{\sum_{i=0}^{n} \mathrm{E}\left[g_i^2\right] \Delta t_i + 0}\,,
\end{aligned}$$

where the zero mean and Δt_i variance properties of ΔW_i were used in the last step. The **expectation Schwarz (Cauchy-Schwarz) inequality**

$$\mathrm{E}[|XY|] \leq \sqrt{\mathrm{E}[X^2] \cdot \mathrm{E}[Y^2]} \tag{2.45}$$

has been used with $X = \sum_{i=0}^{n} g_i \Delta W_i$ and $Y = 1$ to relate the magnitude of the sum to the square root of the sum of squares. Hence, in the mean square sense as $n \to +\infty$, we formally have the expected absolute value of the stochastic diffusion integral is majorized by the square root of the integral of the expected square of the integrand,

$$\mathrm{E}\left[\left|\int_{t_0}^{t} g(W(s),s)dW(s)\right|\right] \leq \sqrt{\int_{t_0}^{t} E\left[g^2(W(s),s)\right] ds}\,. \tag{2.46}$$

It has been assumed that the sums are bounded on the bounded interval $[t_0, t]$, so that, in absence of stochasticity, we can expect uniform convergence of the sums and that the operations of expectation and mean square limit can be interchanged.

- Note, that this mean zero (2.43) for the Itô stochastic integral result depends heavily on the Itô forward or left endpoint integration choice, and as will be seen later, the mean zero result will not hold for other rectangular integration rule choice.
- Under similar conditions, a quadratic or "ims-covariance" version of this theorem holds for interchanging expectation and mean square limit.

□

Theorem 2.18. ***Itô-Covariance of Stochastic Integral***

$$\begin{aligned}
&\mathrm{E}\left[\int_{t_0}^{t} f(W(s),s)dW(s) \int_{t_0}^{t} g(W(r),r)dW(r)\right] \\
&\overset{ims}{=} \int_{t_0}^{t} \mathrm{E}\left[f(W(s),s)g(W(s),s)\right] ds\,,
\end{aligned} \tag{2.47}$$

for $0 \leq t_0 \leq t$, assuming that $f(W(t),t)$ and $g(W(t),t)$ satisfy the mean square integrability condition (2.44) and the i-PWCA Mean Square Limits Assumption 2.15 for $Y(t) = g(W(t),t)$.

Proof. Again, heuristic justifications are presented here. Replacing the expectation of the Itô integral product with that of the corresponding product of finite sum approximations leads to

$$J_2(t) = \mathrm{E}\left[\int_{t_0}^{t} f(W(s),s)dW(s)\int_{t_0}^{t} g(W(r),r)dW(r)\right] \simeq \sum_{i=0}^{n}\sum_{j=0}^{n}\mathrm{E}[f_i\Delta W_i g_j \Delta W_j] \,,$$

but the independent increments are inter-mingled in the sums and the argument of the expectation of $f_i\Delta W_i g_j \Delta W_j$. However, if $j < i$ then the increment ΔW_i will be independent of f_i, g_j and ΔW_j, while if $j > i$ then ΔW_j will be independent of f_i, g_j and ΔW_i, and for $i = j$ the usual independent increment form is obtained. Thus, taking these independence properties to split the double sum three ways and using independent increment properties leads to

$$\begin{aligned} J_2(t) \;&\simeq\; \sum_{i=0}^{n}\mathrm{E}[f_i g_i]\mathrm{E}[(\Delta W_i)^2] + \sum_{i=0}^{n}\sum_{j=0}^{i-1}\mathrm{E}[f_i g_j \Delta W_j]\mathrm{E}[\Delta W_i] \\ &\qquad s + \sum_{i=0}^{n}\sum_{j=i+1}^{n}\mathrm{E}[f_i g_j \Delta W_i]\mathrm{E}[\Delta W_j] \\ &= \sum_{i=0}^{n}\mathrm{E}[f_i g_i]\Delta t_i \\ &\xrightarrow{ims} \int_{t_0}^{t}\mathrm{E}\left[f(W(s),s)g(W(s),s)\right]ds \,, \end{aligned}$$

giving the desired conclusion except for replacing the approximately equals ($\simeq$) by the mean square limit as $n \to \infty$

Upon replacing the function f by g, leads to the immediate corollary for the "ims-variance" of the Itô stochastic integral in the following. □

Corollary 2.19. ***Itô-Variance of Stochastic Integral:***

$$\mathrm{E}\left[\left(\int_{t_0}^{t} g(W(s),s)dW(s)\right)^2\right] \overset{ims}{=} \int_{t_0}^{t}\mathrm{E}\left[g^2(W(s),s)\right]ds \,, \tag{2.48}$$

for $0 \leq t_0 \leq t$, assuming that $g(W(t),t)$ satisfies the mean square integrability condition (2.44).

Result (2.48) is also called *Itô isometry* or *martingale isometry*.

Theorem 2.20. ***Itô Stochastic Integral Simple Rules:***
Let g, g_1 and g_2 satisfy the mean square integrability condition (2.44) on $0 \le t_0 \le t$ and the i-PWCA Mean Square Limits Assumption 2.15, while letting c_1 and c_2 be constants.

- ***Operator Linearity:***

$$\int_{t_0}^{t} [c_1 g_1(W(s), s) \ + \ c_2 g_2(W(s), s)] dW(s)$$
$$\stackrel{ims}{=} c_1 \int_{t_0}^{t} g_1(W(s), s) dW(s) + c_2 \int_{t_0}^{t} g_2(W(s), s) dW(s) \ .$$

- ***Additivity over Subintervals:***

$$\int_{t_0}^{t} g(W(s), s) dW(s) \stackrel{ims}{=} \int_{t_0}^{r} g(W(s), s) dW(s) + \int_{r}^{t} g(W(s), s) dW(s)$$

  for $0 \le t_0 \le r \le t$.

- ***Continuity of Sample Paths for***

$$I[g](t) = \int_{t_0}^{t} g(W(s), s) dW(s) \ ,$$

  with probability one.

Proof. The first two are clearly true by examining the forward integration approximation. For the last item note that

$$\Delta I[g](t) = I[g](t+\Delta t) - I[g](t) = \int_{t}^{t+\Delta t} g(W(s), s) dW(s) \stackrel{ims}{=} g(W(t), t) \Delta W(t) \to 0$$

with probability one as $\Delta t \to 0^+$. □

For later use in formal stochastic calculations, it will be helpful to know how to handle powers of $dW(t)$ greater than square powers. The critical problem is to know when to truncate a differential expansion, such as that for $exp(dW(t))$, at a power of $dW(t)$ beyond which the higher powers are zero in the sense of the Itô mean square limit. For example, $\exp(dW(t))$ can be formally expanded by Taylor series as

$$\exp(dW(t)) = 1 + dW(t) + (dW)^2(t)/2! + (dW)^3(t)/3! + (dW)^4(t)/4! + \cdots$$

and it turns out we can justify stopping at the quadratic term for the mean square limit. The consequence will be the famous Itô stochastic chain rule discussed for jump-diffusions in Chapter 4 and will lead to more rapid calculations. The main purpose of the current chapter is setting up the foundational justification for this chain rule.

Lemma 2.21. ***Powers of*** $dW(t)$***:***
Let the integer $m \geq 3$.

$$\int_0^t (dW)^m(s) \overset{ims}{=} 0 \tag{2.49}$$

or in symbolic differential notation

$$(dW)^m(t) \underset{ms}{\overset{\mathrm{dt}}{=}} 0 \, . \tag{2.50}$$

Proof. Let $m \geq 3$ and

$$I[(dW)^{m-1}](t) = I(t;m) \equiv \int_0^t (dW)^m(s) \simeq I_n^{(0)}(t;m) = \sum_{i=0}^{n} (\Delta W_i)^m. \tag{2.51}$$

The expectation of the Itô approximate sum $I_n^{(0)}(t;m)$ yields different formulae for odd values, $m = 2k-1$ for $k \geq 2$,

$$\mathrm{E}[I_n^{(0)}(t;2k-1)] = \sum_{i=0}^{n} \mathrm{E}\left[(\Delta W_i)^{2k-1}\right] = 0 \, ,$$

while for even values, $m = 2k$ for $k \geq 2$,

$$\begin{aligned} \mathrm{E}[I_n^{(0)}(t;2k)] = \sum_{i=0}^{n} \mathrm{E}\left[(\Delta W_i)^{2k}\right] &= (2k-1)!! \sum_{i=0}^{n} (\Delta t_i)^k \\ &\leq (2k-1)!! t (\delta t_n)^{k-1} \to 0 \, , \end{aligned}$$

as $n \to \infty$, where $(2k-1)!!$ is the double factorial function (1.15). Odd or even m, $m \geq 3$, the results suggest that the Itô mean square value is given by

$$I(t;m) \overset{ims}{=} I^{(ims)}(t;m) \equiv \lim_{n\to\infty} [I_n^{(0)}(t;m)] = 0 \, .$$

The justification requires confirmation of mean square convergence,

$$\lim_{n\to\infty} \mathrm{E}\left[(I_n^{(0)}(t;m) - I^{(ims)}(t;m))^2\right] = \lim_{n\to\infty} \mathrm{E}\left[(I_n^{(0)})^2(t;m)\right] \, .$$

For odd values, $m = 2k-1$, separating out the diagonal part of the quadratic to separate the independent increments,

$$\begin{aligned} \mathrm{E}\left[(I_n^{(0)})^2(t;2k-1)\right] &= \sum_{i=0}^{n} \mathrm{E}\left[(\Delta W_i)^{2(2k-1)} + \sum_{j\neq i} (\Delta W_i)^{2k-1} (\Delta W_j)^{2k-1}\right] \\ &= (4k-3)!! \sum_{i=0}^{n} (\Delta t_i)^{2k-1} \\ &\leq (4k-3)!! t (\delta t_n)^{2k-2} \to 0 \, , \end{aligned}$$

as $n \to \infty$, off-diagonal odd power terms do not contribute. Here $(4k-3)!!$ is the double factorial function (1.15). For even values, $m = 2k$, the off-diagonal terms contribute since they are products of even powers of increments in i and j, so upon completing the double sum over $j \neq i$ and subtracting the completed amount from the single sum,

$$\begin{aligned}
\mathrm{E}\left[(I_n^{(0)})^2(t;2k)\right] &= \sum_{i=0}^{n} \mathrm{E}\left[(\Delta W_i)^{4k} + \sum_{j\neq i}(\Delta W_i)^{2k}(\Delta W_j)^{2k}\right] \\
&= ((4k-1)!! - ((2k-1)!!)^2)\sum_{i=0}^{n}(\Delta t_i)^{2k} \\
&\quad +((2k-1)!!)^2 \sum_{i=0}^{n}(\Delta t_i)^k \sum_{j=0}^{n}(\Delta t_j)^k \\
&\leq (4k-1)!!t(\delta t_n)^{2k-1} + ((2k-1)!!)^2 t(\delta t_n)^{2k-2}(t-\delta t_n) \to 0\ ,
\end{aligned}$$

as $n \to \infty$. Thus, denoting the conclusion symbolically, $(dW)^m(t) \overset{\mathrm{dt}}{=\!=}_{ms} 0$, provided $m \geq 3$ to an accuracy with error $\mathrm{o}(dt)$. □

Another differential product whose Itô mean square limit will be useful is $dt\ dW(t)$ since it arises in the expansions of functions of stochastic differentials:

Lemma 2.22. ***Differential Product*** *$dt\ dW(t)$:*

$$\int_0^t ds\ dW(s) \overset{ims}{=} 0 \tag{2.52}$$

or in symbolic notation

$$dt\ dW(t) \overset{\mathrm{dt}}{=\!=}_{ms} 0\ . \tag{2.53}$$

Proof. Let

$$I[dt](t) = \int_0^t ds\ dW(s) \simeq I_n^{(0)}[dt](t) \equiv \sum_{i=0}^{n} \Delta t_i \Delta W_i, \tag{2.54}$$

with some abuse of the notation by replacing functional argument g by dt. The expectation of the sum $I_n^{(0)}[dt](t)$ yields

$$\mathrm{E}[I_n^{(0)}[dt](t)] = \sum_{i=0}^{n} \mathrm{E}\left[\Delta t_i \Delta W_i\right] = 0\ .$$

The result suggests that the Itô mean square value is given by

$$I[dt](t;m) \overset{ims}{=} \lim_{n\to\infty}[I_n^{(0)}[dt](t;m)] = 0\ .$$

The justification requires confirmation of mean square convergence, separating out the diagonal part of the quadratic to separate the independent increments,

$$\begin{aligned}\mathrm{E}\left[(I_n^{(0)})^2[dt](t)\right] &= \sum_{i=0}^{n} \mathrm{E}\left[(\Delta t_i)^2(\Delta W_i)^2 + \sum_{j\neq i} \Delta t_i \Delta t_j \Delta W_i \Delta W_j\right] \\ &= \sum_{i=0}^{n} (\Delta t_i)^3 \leq t(\delta t_n)^2 \to 0 \ ,\end{aligned}$$

as $n \to \infty$, off-diagonal do not contribute. Thus, $dt\ dW(t) \stackrel{\mathrm{dt}}{=}_{ms} 0$ to an accuracy with error $\mathrm{o}(dt)$. □

Remarks 2.23.

- *Of the Itô differentiable forms that have zero limit in the mean square, $dtdW(t)$ is one of the most marginable to approximate due to the randomness of $dW(t)$, even though we know $\mathrm{E}[dtdW(t)] = 0$ and $\mathrm{E}[|dW(t)|] = \sqrt{2\Delta t/\pi}$ from convenient Table 1.1. Hence, the justification of $\int_0^t dsdW(s) \stackrel{ims}{=} 0$ by showing the mean square limit is especially important. Note that for even spacing of time increments, the root mean square of the bound of the mean square approximation above is $\sqrt{t(\delta t_n)^2} = t\sqrt{t}/(n+1) \to 0$ as $n \to \infty$. However, see Exercise 2 for a more cutting-edge example.*
- *See Exercise 5 for how to computationally confirm the above Lemma 2.22.*

The mean square limits to an accuracy with error $\mathrm{o}(dt)$ are summarized in the following Table 2.1.

Table 2.1. *Some Itô stochastic diffusion differentials with an accuracy with error $\mathrm{o}(dt)$ as $dt \to 0^+$.*

Differential Diffusion Form	Itô Mean Square Limit
$dW(t)$	$dW(t)$
dt	dt
$dt\ dW(t)$	0
$(dW)^2(t)$	dt
$(dW)^m(t)$	0, $m \geq 3$
$(dt)^\alpha (dW)^m(t)$	0, $\alpha > 0,\ \ m \geq 1$

The more general form,

$$(dt)^p (dW)^q(t) \stackrel{\mathrm{dt}}{=}_{ms} \delta_{2p+q,0} + dW(t)\delta_{2p+q,1} + dt\ \delta_{2p+q,2} \ , \tag{2.55}$$

when p and q are non-negative integers, is left as Exercise 1 on Page 59.

Remark 2.24. *In using Table 2.1, the differential entries are just symbols of the underlying integral basis and care should be taken when applying them to find the mean square representation of differentials, especially when they appear in multiplicative combinations. For instance, one might be tempted to replace* $(dW)^4(t)$ *by* $(dW)^2(t)(dW)^2(t)$*, then replace those terms with* $(dW)^2(t) \overset{\text{dt}}{=\!=}_{ms} dt$ *and getting to* $(dt)^2 \overset{\text{dt}}{=\!=}_{ms} 0$*, which is the correct but crudely found answer for* $(dW)^4(t)$*. Note that for finite increments,* $\mathrm{E}[(\Delta W_i)^4] = 3(\Delta t_i)^2$ *while* $\mathrm{E}^2[(\Delta W_i)^2] = (\Delta t_i)^2$*, differing by a factor of three.*

2.3 Stratonovich and other Stochastic Integration Rules

In this section, other definitions of stochastic integration rules, other than Itô's choice of the forward left endpoint rule, are explored for the purpose of comparison and understanding Itô's choice. This comparison will be illustrated by the simple stochastic integral of $W(t)$.

Let the integration θ-rule approximation point be

$$t_{i+\theta} \equiv t_i + \theta \Delta t_i \ , \tag{2.56}$$

where $0 \le \theta \le 1$, so the Itô's rule is when $\theta = 0$ with $\Delta t_i \equiv t_{i+1} - t_i$. Let the interval of integration be $[0, t]$ with partition (2.4). Let the approximate integrand be $W_{i+\theta} \equiv W(t_{i+\theta})$. The technique of splitting terms into independent increments is similar to that for Itô's rule, except that there are extra independent increments,

$$\Delta^\theta W_i \equiv W_{i+\theta} - W_i \tag{2.57}$$

and its complement

$$\Delta_c^\theta W_i \equiv \Delta W_i - \Delta^\theta W_i = W_{i+1} - W_{i+\theta} \tag{2.58}$$

for intermediate approximation points when $\theta > 0$, such that $\Delta^\theta W_i + \Delta_c^\theta W_i = \Delta W_i$. We also reuse (2.14) of the reduction Lemma 2.1 for the Itô case in the more general case here:

$$\begin{aligned} I[W](t) = \int_0^t W(s) dW(s) \simeq I_n^{(\theta)}[W](t) &\equiv \sum_{i=0}^n W_{i+\theta} \Delta W_i \\ &= \sum_{i=0}^n (W_i + \Delta^\theta W_i)(\Delta^\theta W_i + \Delta_c^\theta W_i) \\ &= \sum_{i=0}^n \left(W_i \Delta W_i + (\Delta^\theta W_i)^2 + \Delta^\theta W_i \Delta_c^\theta W_i \right) \\ &= \frac{1}{2} \left(W_{n+1}^2 - \sum_{i=0}^n (\Delta W_i)^2 \right) + \sum_{i=0}^n (\Delta^\theta W_i)^2 + \sum_{i=0}^n \Delta^\theta W_i \Delta_c^\theta W_i \ . \end{aligned}$$

Since $W_{n+1} = W(t)$ with this $[0,t]$ partition and the mean square limit of $\sum_{i=0}^{n}(\Delta W_i)^2$ has been shown to be t, similarly the mean square limit of $\sum_{i=0}^{n}(\Delta^\theta W_i)^2$ will be the expected value θt, and the last sum will not contribute in the mean being the product of independent increments, the mean square limit corresponding to the Itô Lemma 2.7 can be stated:

Lemma 2.25.

$$\begin{aligned}\int_0^t W(s)dW(s) \underset{ms}{\overset{\theta}{=}} I^{(\theta)}[W](t) &= \frac{1}{2}W^2(t) - \left(\frac{1}{2}-\theta\right)t \\ &= \underset{n\to\infty}{\overset{\text{ms}}{\lim}} \left[I_n^{(\theta)}[W](t)\right] ,\end{aligned} \tag{2.59}$$

Proof. The mean square limit justifications are quite lengthy and somewhat tangent to our goals here, so only the general end result is given with the details left to the reader:

$$\begin{aligned}\mathrm{E}\left[\left(I_n^{(\theta)}[W](t) - I^{(\theta)}[W](t)\right)^2\right] &= 2\left|\frac{1}{2}-\theta\right|\sum_{i=0}^{n}(\Delta t_i)^2 \\ &\le 2\left|\frac{1}{2}-\theta\right| t\delta t_n \to 0 ,\end{aligned}$$

where $\delta t_n = \max_{i=0:n}[\Delta t_i] \to 0^+$ as $n \to \infty$. □

Remark 2.26. ***Stratonovich and Other Stochastic Integration Rules:*** *The mean square limit is exact, no limit $n \to \infty$, required, in the case $\theta = 1/2$ where $t_{i+0.5} = (t_i + t_{i+1})/2$ is the midpoint of $[t_i, t_{i+1}]$ and the integration rule is called the midpoint rule or* ***Stratonovich stochastic integration*** *[260]. For Stratonovich integration,*

$$\int_0^t W(s)dW(s) \underset{ms}{\overset{\theta}{=}} I^{(0,5)}[W](t) = W^2(t)/2 ,$$

which is the deterministic integral answer, containing no correction as in the case of Itô's rule. This deterministic property might offer some benefit in some applications, but at the expense of more complicated overlapping dependence of increments in time.

Lemma 2.27.

$$\mathrm{E}\left[I^{(\theta)}[W](t)\right] = \mathrm{E}\left[\frac{1}{2}W^2(t) - \left(\frac{1}{2}-\theta\right)t\right] = \theta t. \tag{2.60}$$

Proof. The result is immediate since $\mathrm{E}[W^2(t)] = t$ from Table 1.1 when $n = 2$ with $|\Delta W|^2(t)$ replaced by $W^2(t)$ and Δt by t. □

Remarks 2.28.

- *When $\theta \neq 0$, then the useful Itô* ***expectation-integration interchange property****,*

$$\mathrm{E}\left[\int_0^t f(W(s),s)dW(s)\right] \overset{ims}{=} \int_0^t \mathrm{E}[f(W(s),s)]\mathrm{E}[dW(s)] = 0$$

  is no longer valid as implied by (2.43). This is a quite nice concrete property, but for abstract analysis it is more crucial since it means, with appropriate qualification on $f(W(t),t)$, that the Itô integral is a martingale.

- *Decades ago, there was a larger controversy as to*

  whether Itô or Stratonovich stochastic integration

  should be used. The question sometimes centered about what was more appropriate for the application at hand (see for instance, Turelli [271] for a discussion involving biological applications), but the benefits of Itô's choice of forward integration facilitating the use of independent increments of the processes and the fact that many Stratonovich properties were derived by Itô stochastic calculus have made the Itô calculus dominant.

2.4 Conclusion

In this chapter, the foundations have been laid for the integrals of the second type in the integrated SDE (2.2), i.e., using the stochastic diffusion integral of Itô of Definition 2.8 extended to the more general case:

Definition 2.29. ***Stochastic Diffusion Integration:***

$$\int_0^t g(X(s),s)dW(s) \overset{ims}{=} \operatorname*{ms\,lim}_{n\to\infty}\left[\sum_{i=0}^{n} g(X(t_i),t_i)dW(t_i)\right] , \tag{2.61}$$

where $X(t)$ in the integrand function g has an implied dependence on the diffusion process $W(t)$, but also depends on the jump process $P(t)$. The integrand process $g(X(t),t)$ is also assumed to have a bounded mean square,

$$\mathrm{E}\left[\int_0^t g^2(X(s),s)ds\right] < \infty ,$$

and satisfy the i-PWCA Mean Square Limits Assumption 2.15 for $Y(t) = g(X(t),t)$.

However, as previously explained, the Poisson jump process fits within the framework of Itô stochastic integration since it is piece-wise continuous. The stochastic diffusion integration rule (2.61) has been motivated and illustrated by a number of examples using functions and powers of the diffusion process $W(t)$.

2.5 Exercises

In all computational exercises, *Mathematica*, MATLAB, Maple or other programming may be used where appropriate, but both figures and codes should be submitted for evaluation.

1. Justify the general form (2.55) by mean square convergence,

$$(dt)^p (dW)^q(t) \overset{\text{dt}}{\underset{ms}{=}} \delta_{2p+q,0} + dW(t)\delta_{2p+q,1} + dt\ \delta_{2p+q,2}\ ,$$

   when p and q are non-negative integers.
   {*Remark: It may be assumed that the cases* $2p+q = 0:2$ *are well-known, so need to show mean square convergence results for* $2p+q \geq 3$ *in general.*}

2. Show the limit in the mean square for

$$I\left[(dt)^\alpha\right](t) \equiv \int_0^t (ds)^\alpha dW(s) \overset{ims}{=} 0\ ,$$

   provided $\alpha > 0$ and is real.
   {*Hint: See Lemma 2.22 for the case* $\alpha = 1$.}

3. Computationally confirm the mean square limit for Itô's most fundamental stochastic integral given as

$$\int_0^t (dW)^2(s) \overset{ims}{=} t\ ,$$

   by demonstrating that the Itô forward integration approximating sum

$$I_n^{(0)}[dW](t) = \sum_{i=0}^{n} (\Delta W_i)^2$$

   gives a close approximation to t for sufficiently large n. Apply a modification of the algorithm of Program C.7 in Appendix C, used in generating Figure 1.1, to the approximation $I_n^{(0)}[dW](t)$, submitting your modification of the code. Use $n_1 = 1000$ and $n_2 = 10000$ sample sizes, plotting the $I_n^{(0)}[dW](t)$ with the limit t versus t for $t \in [0,2]$. Plot separately the errors for each n between the approximation sum and the exact IMS answer. Also report the standard deviation (`std` in MATLAB) of the errors for each n. Characterize the convergence on the average by assuming that the standard deviation satisfies the simple rule $std_j \simeq C/n_j^\beta$ as $n_j \to \infty$ for $j = 1:2$, and find the average convergence rate β from the two sample step sizes n_j for $j = 1:2$.
   {*Caution: In this problem and the next two, you are not asked to verify the mean square limit, but to verify that the forward approximation comes close in this example.*}

4. Computationally confirm the mean square limit for Itô's other very fundamental stochastic integral given as

$$\int_0^t W(s)dW(s) \stackrel{ims}{=} I^{(\mathrm{ims})}[W](t) = \frac{1}{2}\left(W^2(t) - t\right)$$

by demonstrating that the Itô forward integration approximating sum

$$I_n^{(0)}[W](t) = \sum_{i=0}^{n} W_i \Delta W_i$$

gives a close approximation to $W^2(t) - t)/2$ for sufficiently large n. Apply a modification of the algorithm of Program C.7 in Appendix C, used in generating Figure 1.1, to the approximation $I_n^{(0)}[W](t)$. Use $n_{1,k} = 100$ and $n_{2,k} = 10000$ sample sizes and for $k = 1 : 4$ different states or seeds. Plot the approximation $I_n^{(0)}[W](t)$ with $(W^2(t) - t)/2$ and the error $E_n[W](t) = I_n^{(0)}[W](t) - (W^2(t) - t)/2$ versus t for $t \in [0, 2]$. Plot separately the errors for each n between the approximation sum and the exact IMS answer. Also report the standard deviation (`std` in MATLAB) of the errors for each nj, k for $j = 1 : 2$ sample sizes and $k = 1 : 4$ states. From these values compute the common rate β_k for both $j = 1 : 2$ sizes and for each fixed state $k = 1 : 4$, assuming $std_{j,k} = C_k/n_{j,k}^{\beta_k}$ for each $j = 1 : 2$ and $k = 1 : 4$, finally computing the average β_k over $k = 1 : 4$. Does the larger value of n make Itô's stochastic integration model more convincing than the smaller value?

5. Computationally confirm the mean square limit for another of Itô's more obvious fundamental stochastic integrals:

$$\int_0^t dsdW(s) \stackrel{ims}{=} I^{(\mathrm{ims})}[dt](t) = 0$$

by demonstrating that the Itô forward integration approximating sum

$$I_n^{(0)}[dt](t) = \sum_{i=0}^{n} \Delta t_i \Delta W_i$$

gives a close approximation to 0 for sufficiently large n. Apply a modification of the algorithm of Program C.7 in Appendix C , used in generating Figure 1.1, to the approximation $I_n^{(0)}[dt](t)$. Use $n_1 = 1000$ and $n_2 = 10000$ sample sizes, plotting the common value of the approximation and error $I_n^{(0)}[dt](t) = E_n[dt](t)$ and the noise $W(t)$ for $t \in [0, 2]$. Plot separately the errors for each n between the approximation sum and the exact IMS answer. Also report the standard deviation (`std` in MATLAB) of the errors for each nj, k for $j = 1 : 2$ sample sizes and $k = 1 : 4$ states. From these values compute the common rate β_k for both $j = 1 : 2$ sizes and for each fixed state $k = 1 : 4$, assuming $std_{j,k} = C_k/n_{j,k}^{\beta_k}$ for each $j = 1 : 2$ and $k = 1 : 4$, finally computing the average β_k over $k = 1 : 4$. Does the larger value of n make Itô's stochastic integration model more convincing than the smaller value?

6. Computationally check the Itô mean square limit for convergence of the Itô approximating sum of the stochastic integral of $(dW)^2(t)$ to the limit t by directly computing the K-sample mean square

$$S_{i,n}^{(K)} = \frac{1}{K}\sum_{k=1}^{K}\left(\sum_{j=1}^{i}\left(\left(\Delta W_j^{(k)}\right)^2 - \Delta t_j\right)\right)^2 ,$$

where the identity $t = t_{n+1} = \sum_{i=0}^{n}\Delta t_i$ has been used to merge t into the approximating sum. Select $K = 5$ random states or seeds, $n = 10^m$ for $m = 2:5$ sample sizes, constant $\Delta t_i = \Delta t$, $i = n$ and $t = 1$, as an example. Plot $\log_{10}(S_{n,n}^{(K)})$ versus $m = \log_{10}(n)$. What rate of convergence is suggested by this graph?
{*Hint: If* $\Delta t = 10^m$ *and* $S \sim C\cdot(\Delta t)^a$ *then* $\log_{10}(S) \sim a\cdot m + \log_{10}(C)$. *In MATLAB for instance, recall that* `randn('state',k);` *sets the* k *normal random number* `state`.}

7. Show that the non-Itô, approximate backward integration rule ($\theta = 1$) for the stochastic integral

$$\int_{t_0}^{t} W(s)dW(s) \simeq I_n^{(1)}(t) = \sum_{i=0}^{n} W_{i+1}\Delta W_i$$

differs from the Itô rule ($\theta = 0$) by a deterministic factor of t in the mean square limit, i.e.,

$$I_n^{(1)}(t) - I_n^{(0)}(t) \xrightarrow{ims} t\ .$$

{*Hint: The mean square limit is not needed if the approximate integral is related to the Itô integral for* $(dW)^2(t)$.}

8. Show that the non-Itô, approximate trapezoidal integration rule, a variant of the Stratonovich integral, for the stochastic integral

$$\int_{t_0}^{t} W(s)dW(s) \simeq I_n^{(trap)}(t) = \frac{1}{2}\sum_{i=0}^{n}(W_i + W_{i+1})\Delta W_i$$

differs from the Itô rule ($\theta = 0$) by a deterministic factor of $t/2$ in the mean square limit, i.e.,

$$I_n^{(trap)}(t) - I_n^{(0)}(t) \xrightarrow{ims} \frac{1}{2}t\ .$$

{*Hint: The mean square limit is not needed if the approximate integral is related to the one for* $(dW)^2(t)$.}

9. Demonstrate that the **trapezoidal rule** leads to Stratonovich or regular calculus by approximating the stochastic integral example

$$\int_{t_0}^{t} W^2(s)dW(s)$$

with

$$I_n^{(0)}(t) = \frac{1}{2}\sum_{i=0}^{n}(W_i^2 + W_{i+1}^2)\Delta W_i \ .$$

In particular, show that

$$I_n^{(0)}(t) = \frac{1}{3}(W^3(t_{n+1}) - W^3(t_0)) + \frac{1}{6}\sum_{i=0}^{n}(\Delta W_i)^3 \ ,$$

by forming convenient powers of independent increments. Formally, justify that mean square limit is just the first term using elementary mean square properties for the powers of increments $(\Delta W)^p(t_i)$. You are not required to rigorously show mean square convergence, unless you want to show it.

Remark 2.30. *In numerical integration of deterministic integrands, both the midpoint rectangular rule and the trapezoidal rule yield the same order of error estimate when the integrand is sufficiently continuous.*

10. Formally show that the θ-rule expansion (no mean square convergence justification requested) leads to

$$\mathrm{E}\left[\int_0^t g(W(s))dW(s)\right] \underset{ms}{\overset{\theta}{=}} \mathrm{E}\left[I^{(\theta)}[g(W)](t)\right] = \theta\int_0^t \mathrm{E}\left[g'(W(s))\right]ds,$$

where $0 \le \theta \le 1$, assuming the basic θ-rule approximation for the stochastic integral is

$$\int_0^t g(W(s))dW(s) \simeq I_n^{(\theta)}[g(W)](t) \equiv \sum_{i=0}^{n} g_{i+\theta}\Delta W_i,$$

where g has a bounded mean square expectation (2.44), bounded derivatives of all orders, $g_{i+\theta} = g(W_{i+\theta}) = g(W(t_{i+\theta})) = g(W_i + \Delta^\theta W_i)$ with $t_{i+\theta} = t_i + \theta\Delta t_i$ from (2.56), and assuming that g satisfies the 2nd-order Taylor approximation with 3rd-order error,

$$g(w_0 + \Delta W) = g(w_0) + g'(w_0)\Delta W + \frac{1}{2}g''(w_0)(\Delta W)^2 + (\Delta W)^3\mathrm{O}(1),$$

sufficiently uniform with respect to the density $\phi_{\Delta W(t)}(w)$ on $(-\infty, +\infty)$ to allow termwise expectations, provided you can show that $\mathrm{E}[(\Delta^\theta W_i)^m] = \mathrm{O}^2(\theta\Delta t_i)$ for $m \ge 3$ and sufficiently small Δt_i. See also the θ-decomposition (2.57-2.58) of ΔW_i.

Remark 2.31. *Thus, this demonstrates that the Itô sense Theorem 2.17 is generally limited to $\theta = 0$.*

Suggested References for Further Reading

- Arnold, 1974 [13].
- Gard, 1988 [91].
- Itô, 1951 [149].
- Karlin and Taylor, 1981 [162].
- Kloeden and Platen, 1999 [165].
- Kolmogorov and Fomin, 1970 [168].
- Mikosch, 1998 [209].
- Øksendal, 1998 [222].
- Protter, 1990 [232].
- Schuss, 1980 [244].
- Taylor and Karlin, 1998 [265].

Chapter 3

Stochastic Integration for Jumps

A unique feature of this chapter is the greater emphasis on the importance of the lack of continuity that leads to deviations from the chain rule of regular calculus, namely, the discontinuity of Poisson jumps in time and the non-smooth behavior of Wiener. The Poisson jump processes are given in terms of special right-continuous step and impulse functions. Unless otherwise stated, a fixed jump rate λ is assumed. The Poisson jump calculus is also formulated in terms of finite difference algebraic recursions.

3.1 Stochastic Integration in $P(t)$: The Foundations

In this chapter, foundations will be laid for the integrals of the third type in the integrated SDE (2.2), i.e., using the notion of Itô stochastic integral of Definition 2.8 (p. 40) by extending it to the jump case:

Definition 3.1. ***Poisson Jump Stochastic Integration:***

$$\int_0^t h(X(s),s)dP(s) \overset{ims}{=} \operatorname*{ms\,lim}_{n\to\infty} \left[\sum_{i=0}^{n} h(X(t_i),t_i)\Delta P(t_i)\right] , \tag{3.1}$$

where $X(t)$ in the integrand function h has an implied dependence on the diffusion process $W(t)$, but also depends on the jump process $P(t)$. The integrand process $h(X(t),t)$ is also assumed to have a ***bounded mean integral of squares****,*

$$\mathrm{E}\left[\int_0^t h^2(X(s),s)ds\right] < \infty , \tag{3.2}$$

and to satisfy the Piece-Wise Constant Approximations (i-PWCA) Mean Square Limits Assumption 2.15 (p. 48) for $Y(t) = h(X(t),t)$, with the usual grid partition specifications on $[0,t]$.

For most problems encountered in practice, there will not be a need for this

elaborate but fundamental mean square definition. The definition may be needed as a reference for unusual applications with stochastic jumps.

For instance, if an exact differential in $P(t)$ can be formed, then as with stochastic diffusion integration, i.e., when the variable of integration is the random diffusion process $W(t)$, there will be no need for mean square justification. Since much of the work of stochastic integration was performed in the previous chapter, with some very general results, it will be possible to move though this chapter faster.

Theorem 3.2. ***Fundamental Theorem of Poisson Jump Calculus:*** *Let $h(p)$ be continuous and $\mathcal{H}(p)$ be continuously differentiable. Then*

(a)

$$d\left(\int_0^t h(P(s))dP(s)\right) \stackrel{ims}{=} h(P(t))dP(t) \tag{3.3}$$

and

(b)

$$\int_0^t d\mathcal{H}(P(s)) \stackrel{ims}{=} \mathcal{H}(P(t)) - \mathcal{H}(0)\ ,\ 0 \le t\ . \tag{3.4}$$

Proof. The proof is almost the same as for the analogous result (2.37, 2.38), except for change in names from $W(t)$ to $P(t)$ and that the issue of unbounded variation need not be considered.

However, the right continuity property of $P(t)$ is essential to account for a jump at t for part (a). Consider the increment version for sufficiently small increments Δt,

$$\begin{aligned}\Delta\left(\int_0^t h(P(s))dP(s)\right) &= \left(\int_0^{t+\Delta t} - \int_0^t\right) h(P(s))dP(s)\\ &= \int_t^{t+\Delta t} h(P(s))dP(s)\\ &\simeq h(P(t))\Delta P(t) = h(P(t))(P(t+\Delta t) - P(t))\\ &\to h(P(t))dP(t)\end{aligned}$$

as $\Delta t \to 0^+$, using the increment definition, subinterval additivity (see (3.23) later in this chapter), the continuity h and piece-wise continuity of P, such that any last minute jump is captured in $\Delta P(t)$ or $dP(t)$.

See the proof of the diffusion part (b) (2.38) for the jump part (b). □

First, consider the most basic jump integral, the integral of $P(t)$ with respect to $P(t)$, namely,

$$I[P](t) = \int_0^t P(s)dP(s)\ ,$$

which will be evaluated directly through precision-dt calculus and indirectly by showing that the defining mean square limit is satisfied.

Theorem 3.3. ***Jump Integral of*** $\int PdP$***:***

$$I[P](t) = \int_0^t P(s)dP(s) \stackrel{ims}{=} I^{(ims)}[P](t) \equiv \frac{1}{2}(P(P-1))(t) \ , \tag{3.5}$$

is ***mean square limit integral****,*

$$I^{(ims)}[P](t) \stackrel{ims}{=} \lim_{n\to\infty}^{\text{ms}} \left[I_n^{(0)}[P](t) \right] \ , \tag{3.6}$$

where the forward integration approximation is

$$I_n^{(0)}[P](t) = \sum_{i=0}^{n} P(t_i)\Delta P(t_i) \ . \tag{3.7}$$

Proof. Starting with the Poisson increment and the square $P^2(t)$, as in the diffusion case since $d(x^2) = 2xdx$ in smooth deterministic calculus,

$$\begin{aligned} \Delta(P^2) &\equiv P^2(t+\Delta t) - P^2(t) = \left((P+\Delta P)^2 - P^2 \right)(t) \\ &= \left(2P\Delta P + (\Delta P)^2 \right)(t) \ . \end{aligned}$$

Taking the limit $\Delta t \to 0^+$, replacing ΔP by dP, and using the zero-one jump law (1.35) to let $(dP)^2 \stackrel{\text{dt}}{=} dP$ **with probability one** upon neglect of smaller order terms, leads to

$$d(P^2)(t) \stackrel{\text{dt}}{=} (2PdP + dP)(t)$$

in probability. Solving for the integrand-differential while forming an exact differential yields in probability

$$(PdP)(t) \stackrel{\text{dt}}{=} \frac{1}{2} d\left(P^2 - P \right)(t) \ .$$

Therefore, integration by the fundamental theorem of stochastic jump integration (3.3)

$$\int_0^t (PdP)(s) \stackrel{ims}{=} \frac{1}{2} \int_0^t (d\left(P^2 - P \right)(t))(s) = \frac{1}{2}\left(P^2 - P \right)(t) = I^{(\text{ims})}[P](t) \ ,$$

where the initial Poisson condition $P(0) = 0$ **with probability one** has been used to eliminate the initial value of the integral. That takes care of the first part of the proof, but the technique is general enough for other powers.

For the second part, the forward integration approximation can be simplified by the useful finite difference identity (2.14),

$$I_n^{(0)}[P](t) = \sum_{i=0}^{n} P_i \Delta P_i = \frac{1}{2}\left(P^2(t) - \sum_{i=0}^{n} (\Delta P_i)^2 \right)$$

for the partition

$$0 = t_0 < t_1 < \cdots < t_{n+1} = t$$

and using the fact (2.13) that

$$P(t) = P_{n+1} = \sum_{i=0}^{n} \Delta P_i \ ,$$

the difference between the approximation and the limit reduces to

$$I_n^{(0)}[P](t) - I^{(\mathrm{ims})}[P](t) = \frac{1}{2} \sum_{i=0}^{n} \left(\Delta P_i - (\Delta P_i)^2 \right) \ .$$

The mean square again is reduced by splitting up the sums due to the square into independent increments prior to term-wise passing the mean over the sums,

$$\begin{aligned}
\mathrm{E}\left[\left(I_n^{(0)}[P](t) - I^{(\mathrm{ims})}[P](t) \right)^2 \right] &= \frac{1}{4} \mathrm{E}\left[\left(\sum_{i=0}^{n} \left(\Delta P_i - (\Delta P_i)^2 \right) \right)^2 \right] \\
&= \frac{1}{4} \sum_{i=0}^{n} \mathrm{E}\left[\left(\Delta P_i - (\Delta P_i)^2 \right)^2 \right] \\
&\quad + \frac{1}{4} \sum_{i=0}^{n} \sum_{j \neq i} \mathrm{E}\left[\left(\Delta P_i - (\Delta P_i)^2 \right) \cdot \left(\Delta P_j - (\Delta P_j)^2 \right) \right] \\
&= \frac{1}{4} \sum_{i=0}^{n} \mathrm{E}\left[(\Delta P_i)^2 - 2(\Delta P_i)^3 + (\Delta P_i)^4 \right] \\
&\quad + \frac{1}{4} \sum_{i=0}^{n} \mathrm{E}\left[\Delta P_i - (\Delta P_i)^2 \right] \sum_{j \neq i} \mathrm{E}\left[\Delta P_j - (\Delta P_j)^2 \right] \\
&= \frac{1}{4} \sum_{i=0}^{n} \left(\lambda \Delta t_i (1 + \lambda \Delta t_i) - 2\lambda \Delta t_i (1 + 3\lambda \Delta t_i + (\lambda \Delta t_i)^2) \right. \\
&\quad \left. + \lambda \Delta t_i (1 + 7\lambda \Delta t_i + 6(\lambda \Delta t_i)^2 + (\lambda \Delta t_i)^3) \right) \\
&\quad + \frac{1}{4} \sum_{i=0}^{n} \left(\lambda \Delta t_i - \lambda \Delta t_i (1 + \lambda \Delta t_i) \right) \\
&\quad \cdot \sum_{j \neq i} \left(\lambda \Delta t_j - \lambda \Delta t_j (1 + \lambda \Delta t_j) \right) \\
&\leq \frac{1}{4} \left(\sum_{i=0}^{n} (\lambda \Delta t_i)^2 \left(2 + 4\lambda \Delta t_i \right) + \sum_{i=0}^{n} (\lambda \Delta t_i)^2 \sum_{j=0}^{n} (\lambda \Delta t_j)^2 \right) \\
&\leq \frac{1}{4} \left(\lambda t (2\lambda \delta t_n + 4(\lambda \delta t_n)^2) + (\lambda t^2)(\lambda \delta t_n)^2 \right) \quad \longrightarrow \quad 0
\end{aligned}$$

as $n \to \infty$ and bounded t. For the evaluation of the expectations of powers of Poisson increments, the convenient Table 1.2 has been frequently used. Therefore, the mean square limit has been proven. □

Remarks 3.4.

- *The main result (3.5),*

$$\int_0^t P(s)dP(s) \overset{ims}{=} \frac{1}{2}(P(P-1))(t) \ ,$$

for this basic integral has an interesting mathematical interpretation. Since $P(t)$ is integer valued, the answer is the Pythagorean $(P(t)-1)th$ ***triangular number*** *given by the successive sum of $n = P(t) - 1$ integers,*

$$S_n^{(1)} = \sum_{k=0}^{n} k = n(n+1)/2 \ . \tag{3.8}$$

The interpretation is not a coincidence, since when $P(t)$ jumps instantaneously by one unit and adds it to its count, $dP(t)$ jumps by one only momentarily so that the integral in (3.5) serves as a ***triangular number counter****. The forward integration approximation serves to keep the count short of the last jump, e.g., the forward approximation is zero when $P(t) = 1$.*

- *The derivation of (3.8) by finite differences gives useful techniques for calculating and interpreting other Poisson jump integrals. The basic lemma for the difference inversion ("discrete integration") is given by*

Lemma 3.5. *If*

$$\Delta[a_n] = \Delta[b_n] \ ,$$

for two sequences and any integer n, then

$$a_n = b_n + C$$

where C is an arbitrary constant.

The proof is obvious since a constant sequence is the only sequence elements that produces zero difference.

- *Since $\Delta[S_n^{(1)}] = S_{n+1}^{(1)} - S_n^{(1)} = (n+1)$, $\Delta[n] = 1$ and $\Delta[n^2] = 2n+1 = 2n + \Delta[n]$ or $n = \frac{1}{2}\Delta[n^2 - n]$, then $\Delta[S_n^{(1)}] = \Delta[(n^2-n)/2 + n]$ and $S_n^{(1)} = n(n+1)/2$, upon elimination the constant of discrete integration by the initial condition $S_0^{(1)} = 0$. This proves the first triangular number sum (3.8) by finite differences using Lemma 3.5.*

The first few Poisson power integrals are listed with an accuracy with error $o(dt)$ in the Table 3.1:

Remarks 3.6.

Table 3.1. *Some stochastic jump integrals of powers with an accuracy with error* $\mathrm{o}(dt)$ *as* $dt \to 0^+$.

m	precision-dt: $\int_0^t (P^m dP)(s)$
0	$P(t)$
1	$(P(P-1))(t)/2$
2	$(P(P-1)(2P-1))(t)/6$
3	$(P^2(P-1)^2)(t)/4$

- *The proofs of the formulas for* $m = 2$ *and* $m = 3$ *are left as an exercise for the reader in Exercise 1 on Page 80.*
- *The integral results of Table 3.1 are all in the form of generalized or super-triangular numbers of order* m *when* $n = P(t) - 1$*:*

  Definition 3.7. *The* ***super-triangular numbers*** *of order* m *for the first* $n+1$ *non-negative integers are defined as*

  $$S_n^{(m)} = \sum_{k=0}^{n} k^m \ ,$$

  for integers $m \geq 0$ *and* $n \geq 0$*.*

  The summation form of a pure Poisson integral is generalized in the following theorem:

Theorem 3.8. ***Pure Poisson Integral as Sum Form:*** *Let* $h(p)$ *be a continuous function and let the process* $h(P(t))$ *have a bounded mean integral of squares (3.2). Then,*

$$\int_0^t h(P(s))dP(s) \overset{ims}{=} \sum_{k=0}^{P(t)-1} h(k) \ , \tag{3.9}$$

with the usual summation convention for irregular forms that

$$\sum_{k=0}^{-1} h(k) \equiv 0 \tag{3.10}$$

for the case that $P(t) = 0$*.*

Proof. It is only necessary to confirm that both sides of Eq. (3.9) satisfy the same differential. The tools used will be the Fundamental Theorem of Stochastic Calculus (3.3) and the idea of Zero-One Jump Power Law (1.36). By the fundamental

theorem, the differential of the left hand side of (3.9),

$$d\left(\int_0^t h(P(s))dP(s)\right) \stackrel{\text{dt}}{=} h(P(t))dP(t) \ .$$

Then, by using the incremental definition of the differential for the right hand side of (3.9),

$$d\left(\sum_{k=0}^{P(t)-1} h(k)\right) = \sum_{k=0}^{P(t)+dP(t)-1} h(k) - \sum_{k=0}^{P(t)-1} h(k) \stackrel{\text{dt}}{=} h(P(t))dP(t) \ ,$$

where the last step is due to the zero-one jump law since the difference in the two sums in the first line is zero if $dP(t) = 0$, else there is only one extra term in the first sum in the alternate case $dP(t) = 1$. Also $dP(t) = 1$ is used in the argument of h. Hence, the differential of both sides of (3.9) are the same. The final result then follows for the reasons:

1. both sides satisfy the same initial condition,
2. the vanishing of the jump integral in the limit,

$$\lim_{t\to 0^+} \int_0^t h(P(s))dP(s) = 0,$$

3. the vanishing of the Poisson sum in the limit.

$$\lim_{t\to 0^+} \sum_{k=0}^{P(t)-1} h(k) = \sum_{k=0}^{-1} h(k) \equiv 0 \ ,$$

4. $P(0^+) = 0$ and
5. the irregular summation convention (3.10).

The argument is analogous to that of mathematical induction, since we have shown that both sides of (3.9) satisfy the same initial condition and the same changes so lead to the same result hypothesized in the theorem. □

Remarks 3.9.

- *Note that in this theorem the sum is over all $P(t)$ jump amplitudes for $k+1 = 1 : P(t)$ jumps, but that the jump amplitude h is evaluated at the pre-jump value $h(k)$ for $k = 0 : P(t) - 1$ by the definition of the Poisson jump with amplitude determined by the function h. This jump amplitude evaluation is consistent with the Itô forward integral approximation,*

$$\Delta \int_0^t h(P(s))dP(s) \simeq h(P(t))\Delta P(t)$$

for a single, sufficiently small time step Δt, picking the prior value of h at $P(t)$ in the case $\Delta P(t) > 0$, though it is not that obvious for the simple jump amplitude dependence $h(P(t))$, the picking of the pre-jump value is also a consequence of the right continuity property of the Poisson process (1.20).

Corollary 3.10.

$$\int_{t=0}^{t} P^m(s)dP(s) \stackrel{ims}{=} S^{(m)}_{P(t)-1} = \sum_{k=0}^{P(t)-1} k^m \ , \tag{3.11}$$

for $m \geq 0$ and the irregular summation convention (3.10) is applicable.

Remark 3.11. *A simple consistency check on (3.11) is to verify the simplest case when $m = 0$ and the integral of $(P^m dP)(t) = dP(t)$ on $[0, t]$ must be $P(t)$ by the fundamental theorem. The right hand side of (3.11), with $k^m = 1$, is*

$$\sum_{k=0}^{P(t)-1} 1 = (P(t) - 1 + 1) \cdot 1 = P(t) \ .$$

Theorem 3.12. ***General Poisson Stochastic Integral:***
Let $h(x, t)$ be a continuous function and let the process $h(X(t), t)$ have a bounded mean integral of squares (3.2) and satisfy the i-PWCA Mean Square Limits Assumption 2.15 for $Y(t) = h(X(t), t)$. Then,

$$\int_0^t h(X(s), s)dP(s) \stackrel{ims}{=} \sum_{k=1}^{P(t)} h(X(T_k^-), T_k^-) \ , \tag{3.12}$$

where T_k is the kth jump of Poisson process $P(t)$.

Proof. Here, we rely explicitly on both the Itô forward integration rule ($\theta = 0$) and the right-continuity property of $P(t)$. It is sufficient to examine the processes $P(t)$, $\Delta P(t)$ and $h(X(t), t)$ in the very neighborhood of the kth jump at time T_k, such that Δt is small enough that we can exclude the prior jump at T_{k-1} and the next jump at T_{k+1} with $T_{k-1} < t < T_{k+1}$. After all, the Poisson process is a rare event process. Thus, the Poisson process has the simple, right-continuous form

$$P(t) = \left\{ \begin{array}{ll} k-1, & T_{k-1} < t \leq T_k^- \\ k, & T_k = T_k^+ \leq t < T_{k+1} \end{array} \right\} \ ,$$

where $1 \leq k \leq P(t)$. However, the increment $\Delta P(t_i) = P(t_i + \Delta t) - P(t_i)$ is a function of both t_i and Δt for $i = 1 : n$, but we are interested in the limit as $\Delta t \to 0^+$ with t_i fixed in (T_{k-1}, T_{k+1}), so there are three case, both t_i and $t_i + \Delta t$

to the left of T_k, T_k between t_i and $t_i + \Delta t$ and both on the right of T_k, i.e.,

$$
\begin{aligned}
&h(X(t_i), t_i)\Delta P(t_i) \\
&\qquad = \left\{ \begin{array}{ll} 0, & T_{k-1} < t_i < t_i + \Delta t \le T_k^- \\ h(X(t_i), t_i), & T_{k-1} < t_i \le T_k^- < T_k = T_k^+ \le t_i + \Delta t < T_{k+1} \\ 0, & T_k = T_k^+ \le t_i < t_i + \Delta t < T_{k+1} \end{array} \right\} \\
&\qquad \to \left\{ \begin{array}{ll} 0, & T_{k-1} < t_i < T_k^- \\ h(X(T_k^-), T_k^-), & T_{k-1} < t_i = T_k^- \\ 0, & T_k = T_k^+ \le t_i < T_{k+1} \end{array} \right\} ,
\end{aligned}
$$

as $\Delta t \to 0^+$ with t_i fixed in (T_{k-1}, T_{k+1}) and this is valid for $1 \le k \le P(t)$. Thus, the Itô approximate sum is

$$
\begin{aligned}
\int_0^t h(X(s), s) dP(s) &\simeq \sum_{i=0}^{n} h(X(t_i), t_i)\Delta P(t_i) \\
&\to \sum_{k=1}^{P(t)} h(X(T_k^-), T_k^-) ,
\end{aligned}
$$

as $n \to +\infty$ and $\delta t_n = \max_j[\Delta t_j] \to 0^+$, since for large n the $\Delta P(t_i)$ will be mostly zero and only the time intervals that straddle a jump T_k^- will be selected. The state process, different from the simple jump Poisson process, will in general undergo continuous changes between jumps of $P(t)$, but the right-continuity causes the immediate pre-jump value of the jump-amplitude at T_k^- to be chosen for each jump time T_k. □

Remark 3.13. *Obviously, if $h(X(t), t) = 1$, then $\sum_{k=1}^{P(t)} 1 = P(t)$. Another simple consistency check on (3.12) is to verify the case when $h(X(t), t) = P(t)$ and the integral of $(PdP)(t)$ on $[0, t]$ must be $(P(P-1))(t)/2$ by (3.5). The right hand side of (3.11), with $h(X(t), t) = P(t)$, $P(T_k^-) = k - 1$, is*

$$
\sum_{k=1}^{P(t)} P(T_k^-) = \sum_{k=1}^{P(t)} (k-1) = P(t)(P(t)-1)/2 ,
$$

using the standard triangular number summation. Hence, Eq. (3.12) is consistent with Eq. (3.9).

Definition 3.14. ***Jump Function*** $[X](t)$***:***
The jump value of the state X at the pre-jump time T_k^- is defined as

$$
[X](T_k) \equiv X(T_k^+) - X(T_k^-) , \tag{3.13}
$$

when the kth jump is at time T_k. For finite discontinuities, the jump function includes all the change of the function, the ***zeroth change or discrete derivative*** *of the state $X(t)$.*

Example 3.15. *Let*

$$Y(t) = \int_0^t h(X(s), s) dP(s)$$

and

$$\Delta Y(t) = \int_t^{t+\Delta t} h(X(s), s) dP(s) \simeq h(X(t), t) \Delta P(t)$$

for $0 < \Delta t \ll 1$*, so*

$$[Y](t) \equiv Y(t^+) - Y(t^-) = \int_{t^-}^{t^+} h(X(s), s) dP(s) = h(X(t^-), t^-) dP(t) \ , \quad (3.14)$$

since $dP(t) = dP(t^-)$ *with both being one when* $t = T_k^-$ *or* $t^- = T_k^-$ *but otherwise zero when* $T_{k-1} < t < T_k^-$ *or* $T_{k-1} < t^- < T_k^-$*.*

In the non-Itô integration approximation, $0 < \theta \leq 1$,

$$\Delta \int_0^t h(P(s)) dP(s) \simeq h(P(t + \theta \Delta t)) \Delta P(t) \ ,$$

so if the last jump is T_k and the next one is T_{k+1}, such that $T_k < t < T_{k+1} < t + \Delta t$, i.e., within the single time step, then $P(t) = k$ and we get the jump amplitude is $h(k)$ if the jump is late, $t + \theta \Delta t < T_{k+1} < t + \Delta t$, since $P(t + \theta \Delta t) = k$, but we get the amplitude $h(k + 1)$ if the jump is early, $t < T_{k+1} < t + \theta \Delta t$, since $P(t + \theta \Delta t) = k + 1$. Thus, the Itô formulation has much less complexity and is more straight-forward to implement.

Some other jump differential products whose mean square limits will be useful are $dt \ dP(t)$ and $dP(t) \ dW(t)$, since they arise in the expansions of functions of stochastic differentials:

Lemma 3.16. ***Differential Products*** $dt \ dP(t)$ ***and*** $dP(t) \ dW(t)$***:***

$$\int_0^t ds \ dP(s) \stackrel{ims}{=} 0, \quad (3.15)$$

or in symbolic notation

$$dt \ dP(t) \stackrel{\mathrm{dt}}{=} 0 \ , \quad (3.16)$$

and

$$\int_0^t dP(s) \ dW(s) \stackrel{ims}{=} 0, \quad (3.17)$$

or in symbolic notation

$$dP(t) \ dW(t) \stackrel{\mathrm{dt}}{=} 0 \ , \quad (3.18)$$

where $W(t)$ and $P(t)$ are independent random variables.

Proof. The proofs are similar to the proof for $dt\ dW(t)$, with a minor change in argument due to the non-zero incremental mean

$$\mathrm{E}[\Delta P(t_i)] = \lambda \Delta t_i.$$

Let

$$I[dt](t) = \int_0^t ds\ dP(s) \simeq I_n[dt](t) \equiv \sum_{i=0}^{n} \Delta t_i \Delta P(t_i). \tag{3.19}$$

The expectation of the sum $I_n[dt](t)$ yields

$$\begin{aligned} \mathrm{E}[I_n[dt](t)] &= \sum_{i=0}^{n} \mathrm{E}\left[\Delta t_i \Delta P(t_i)\right] = \sum_{i=0}^{n} \lambda (\Delta t_i)^2 \\ &\leq \lambda t \delta t_n \to 0^+ , \end{aligned}$$

as $n \to +\infty$. The result suggests that the Itô mean square value is given by

$$I[dt](t) \overset{ims}{=} \lim_{n\to\infty}^{\mathrm{ms}} I_n[dt](t) = 0 .$$

This can be verified in the mean square limit by showing that the mean square limit is zero, while the splitting into independent increments is employed,

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{i=0}^{n} \Delta t_i \Delta P_i - 0\right)^2\right] &= \sum_{i=0}^{n} \left((\Delta t_i)^2 \mathrm{E}[(\Delta P_i)^2] + \sum_{j\neq i} \Delta t_i \Delta t_j \mathrm{E}[\Delta P_i]\mathrm{E}[\Delta P_i] \right) \\ &= \sum_{i=0}^{n} \left(\lambda (\Delta t_i)^3 (1 + \lambda \Delta t_i) + \sum_{j\neq i} \lambda^2 (\Delta t_i \Delta t_j)^2 \right) \\ &= \mathrm{O}^2(\delta t_n) \to 0 , \end{aligned}$$

as $n \to +\infty$. So,

$$dt\ dP(t) \overset{\mathrm{dt}}{=} 0 .$$

The cross product of differentials $dP(t)dW(t)$ works out similarly, except here we have the benefit of independence of processes as well as independence of respective process increments. Let

$$J(t) = \int_0^t dP(s)\ dW(s) \simeq J_n(t) \equiv \sum_{i=0}^{n} \Delta P(t_i) \Delta W(t_i). \tag{3.20}$$

The expectation of the sum $J_n(t)$ yields

$$\mathrm{E}[J_n(t)] = \sum_{i=0}^{n} \mathrm{E}\left[\Delta P(t_i) \Delta W(t_i)\right] = \sum_{i=0}^{n} \lambda (\Delta t_i) \cdot 0 = 0.$$

This result suggests that the Itô mean square value is given by

$$J(t) \overset{ims}{=} \overset{\text{ms}}{\lim_{n\to\infty}} [J_n(t)] = 0 \ ,$$

so that it is intuitively clear that the mean square limit will also behave like the cases $dt\ dW(t)$ and $dt\ dP(t)$, but the verification of the mean square limit is still needed and is left as Exercise 3 for the reader. □

Theorem 3.17. ***Mean Square Limit Form of the Zero-One Law:*** *Let m be a non-negative integer and $\mathrm{E}[dP(t)] = \lambda(t)dt$ with bounded maximum, $\lambda^* = \max_t[\lambda(t)]$, then*

$$\int_0^t (dP)^m(s) \overset{ims}{=} P(t), \tag{3.21}$$

or in symbolic notation

$$(dP)^m(t) \overset{\text{dt}}{=} dP(t) \ . \tag{3.22}$$

Proof. The mean square limit proof is left as an exercise for the reader in Exercise 4. □

The Itô mean square limits to an accuracy with error $\mathrm{o}(dt)$ in the case of the Poisson jump process are summarized in the Table 3.2:

Table 3.2. *Some Itô stochastic jump differentials with an accuracy with error $\mathrm{o}(dt)$ as $dt \to 0^+$.*

Differential Jump Form	Itô Mean Square Limit
$dP(t)$	$dP(t)$
dt	dt
$dt\ dP(t)$	0
$(dP)^m(t)$	$dP(t)$, $m \geq 1$
$dP(t)\ dW(t)$	0
$(dt)^k(dP)^m(t)$	0, $k \geq 1,\ \ m \geq 1$
$(dt)^k(dP)^m(t)(dW)^n(t)$	0, $\ k \geq 1,\ \ m \geq 1,\ \ n \geq 1$

Remarks 3.18.

- *In the use of Table 3.2, the differential entries are just symbols of the underlying integral basis and care should be taken when applying them to find*

the mean square representation of differentials, especially when they appear in multiplicative combinations.

- *The mean square limit justification of the power rule* $(dP)^m(t) \stackrel{\text{dt}}{=} dP(t)$ *is left as Exercise 4, along with Exercise 3 previously mentioned for* $dP(t)dW(t)$.

3.2 Stochastic Jump Integration Rules and Expectations:

Theorem 3.19. ***Itô Stochastic Jump Integral Simple Rules:***
Let h, h_1 *and* h_2 *satisfy the mean square integrability condition (2.44) on* $0 \le t_0 \le t$, *while letting* $X(t)$ *be a Markov process, along with lettimg* c_1 *and* c_2 *be constants.*

- ***Operator Linearity:***

$$\int_{t_0}^{t} [c_1 h_1(X(s), s) \; + \; c_2 h_2(X(s), s)] dP(s)$$
$$\stackrel{ims}{=} c_1 \int_{t_0}^{t} h_1(X(s), s) dP(s) + c_2 \int_{t_0}^{t} h_2(X(s), s) dP(s) \; .$$

- ***Additivity over Subintervals:***

$$\int_{t_0}^{t} h(X(s), s) dP(s) \stackrel{ims}{=} \int_{t_0}^{r} h(X(s), s) dP(s) + \int_{r}^{t} h(X(s), s) dP(s) \quad (3.23)$$

for $0 \le t_0 \le r \le t$.

Proof. These are clearly true by examining the forward integration approximation. □

Poisson jump processes may seem easier in terms of differentials, but they can lead to more difficulties when more complicated integral properties are considered.

Theorem 3.20. ***Some Mean Stochastic Jump Integrals:***
Let $h(X(t), t)$ *satisfy the mean square integrability condition on* $0 \le t_0 \le t$ *and* $X(t)$ *be a Markov process,*

$$\mathrm{E}\left[\int_{t_0}^{t} h^2(X(s), s) ds\right] < \infty \quad (3.24)$$

and the i-PWCA Mean Square Limits Assumption 2.15 for $Y(t) = h(X(t), t)$, *where* $\mathrm{E}[dP(t)] = \lambda(t)dt$, *then*

1. $\boldsymbol{E[\int h(X(s), s) dP(s)]}$*:*

$$\mathbf{E}\left[\int_{t_0}^{t} \boldsymbol{h(X(s), s) dP(s)}\right] \stackrel{ims}{=} \int_{t_0}^{t} \mathbf{E}[\boldsymbol{h(X(s), s)}] \boldsymbol{\lambda(s) ds}. \quad (3.25)$$

2. $\mathbf{E}[\int \boldsymbol{h}(\boldsymbol{X}(s),s)d\widehat{\boldsymbol{P}}(s)]$*:* *Letting*

$$d\widehat{\boldsymbol{P}}(t) \equiv d\boldsymbol{P}(t) - \boldsymbol{\lambda}(t)dt \tag{3.26}$$

be the simple ***mean-zero Poisson process****,*

$$\mathbf{E}\left[\int_{t_0}^{t} \boldsymbol{h}(\boldsymbol{X}(s),s)d\widehat{\boldsymbol{P}}(s)\right] \overset{ims}{=} \mathbf{0}\ . \tag{3.27}$$

3. $\mathbf{E}[|\int \boldsymbol{h}(\boldsymbol{X}(s),s)d\boldsymbol{P}(s)|]$ ***Estimate:***

$$\mathbf{E}\left[\left|\int_{t_0}^{t} \boldsymbol{h}(\boldsymbol{X}(s),s)d\boldsymbol{P}(s)\right|\right] \leq \int_{t_0}^{t} \mathbf{E}\left[|\boldsymbol{h}(\boldsymbol{X}(s),s)|\right]\boldsymbol{\lambda}(s)d\boldsymbol{s}\ , \tag{3.28}$$

where the inequality is in the mean square sense.

4. $\mathbf{E}[\int \boldsymbol{h_1}(\boldsymbol{X}(s),s)d\widehat{\boldsymbol{P}}(s) \int \boldsymbol{h_2}(\boldsymbol{X}(r),r)d\widehat{\boldsymbol{P}}(r)]$*:*
Let $\boldsymbol{h_1}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})$ *and* $\boldsymbol{h_2}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})$ *satisfy the same mean square integrability condition (2.44) as* $\boldsymbol{h}(\boldsymbol{X}(\boldsymbol{t}),\boldsymbol{t})$ *on* $\mathbf{0} \leq \boldsymbol{t_0} \leq \boldsymbol{t}$*, then the Itô-Covariance for jump stochastic integrals is*

$$\begin{array}{l}\mathbf{E}\left[\int_{t_0}^{t} \boldsymbol{h_1}(\boldsymbol{X}(s),s)d\widehat{\boldsymbol{P}}(s) \int_{t_0}^{t} \boldsymbol{h_2}(\boldsymbol{X}(r),r)d\widehat{\boldsymbol{P}}(r)\right] \\ \overset{ims}{=} \int_{t_0}^{t} \mathbf{E}\left[\boldsymbol{h_1}(\boldsymbol{X}(s),s)\boldsymbol{h_2}(\boldsymbol{X}(s),s)\right]\boldsymbol{\lambda}(s)d\boldsymbol{s}\ .\end{array} \tag{3.29}$$

5. $\mathbf{E}[(\int \boldsymbol{h}(\boldsymbol{X}(s),s)d\widehat{\boldsymbol{P}}(s))^{\mathbf{2}}]$*:*
The Itô-Variance for jump stochastic integrals is given by

$$\mathbf{E}\left[\left(\int_{t_0}^{t} \boldsymbol{h}(\boldsymbol{X}(s),s)d\widehat{\boldsymbol{P}}(s)\right)^{\mathbf{2}}\right] \overset{ims}{=} \int_{t_0}^{t} \mathbf{E}[\boldsymbol{h}^{\mathbf{2}}(\boldsymbol{X}(s),s)]\boldsymbol{\lambda}(s)d\boldsymbol{s}\ . \tag{3.30}$$

Sketch of Proof. Only fast heuristic or formal justification will be given here to keep this presentation simple, since many of the techniques have been given earlier for diffusion $W(t)$ and our interests are in applications.

1. Using the Itô mean square limit (2.27), we have the formal finite sum approximation using partition (2.28) with $h_i = h(X(t_i),t_i)$ for the expectation,

$$\begin{aligned}\mathrm{E}\left[\int_{t_0}^{t} h(X(s),s)dP(s)\right] &\simeq \sum_{i=0}^{n} \mathrm{E}[h_i\Delta P_i] = \sum_{i=0}^{n} \mathrm{E}[h_i]\mathrm{E}[\Delta P_i] \\ &= \sum_{i=0}^{n} \mathrm{E}[h_i]\lambda_i\Delta t_i\ ,\end{aligned}$$

the last line using the independent increments and mean properties. Hence (3.25) is formally justified.

2. The form (3.27) follows immediately by combining both sides of the mean square equation in part (a).

3. Again using the forward integration approximation, but with the triangular inequality, the expectation of the absolute value of the stochastic jump integral formally follows,

$$\begin{aligned}
\mathrm{E}\left[\left|\int_{t_0}^{t} h(X(s),s)dP(s)\right|\right] &\simeq \mathrm{E}\left[\left|\sum_{i=0}^{n} h_i \Delta P_i\right|\right] \leq \sum_{i=0}^{n} \mathrm{E}[|h_i|\Delta P_i] \\
&= \sum_{i=0}^{n} \mathrm{E}[|h_i|]\mathrm{E}[\Delta P_i] = \sum_{i=0}^{n} \mathrm{E}[|h_i|]\lambda_i \Delta t_i \\
&\xrightarrow{ims} \int_{t_0}^{t} \mathrm{E}[|h(X(s),s)|]\lambda(s)ds \ ,
\end{aligned}$$

as $n \to +\infty$, using the means square limit in the last step to get the desired limiting estimate.

4. Due to the mean zero property (3.27) of the stochastic jump integral with respect to the mean zero process $d\widehat{P}(t)$ (3.26), the Itô forward integration approximation to the covariance of the stochastic jump integral follows. However, the use of the mean zero process is critical, otherwise the independent increment property is not very helpful. As in the $W(t)$ diffusion case, the approximate finite difference double sum is split up into three parts, the diagonal $(j = i)$, lower diagonal $(j < i)$ and upper diagonal $(j > i)$ parts,

$$\begin{aligned}
&\mathrm{E}\left[\textstyle\int_{t_0}^{t} h_1(X(s),s)d\widehat{P}(s) \int_{t_0}^{t} h_2(X(r),r)d\widehat{P}(r)\right] \\
&\quad \simeq \textstyle\sum_{i=0}^{n}\sum_{j=0}^{n} \mathrm{E}[h_{1,i}\Delta\widehat{P}_i h_{2,i}\Delta\widehat{P}_j] \\
&\quad \simeq \textstyle\sum_{i=0}^{n} \mathrm{E}[h_{1,i}h_{2,i}]\mathrm{E}[(\Delta\widehat{P}_i)^2] + \sum_{i=0}^{n}\sum_{j=0}^{i-1} \mathrm{E}[h_{1,i}h_{2,j}\Delta\widehat{P}_j]\mathrm{E}[\Delta\widehat{P}_i] \\
&\qquad\qquad + \textstyle\sum_{i=0}^{n}\sum_{j=i+1}^{n} \mathrm{E}[h_{1,i}h_{2,j}\Delta\widehat{P}_i]\mathrm{E}[\Delta\widehat{P}_j] \\
&\qquad = \textstyle\sum_{i=0}^{n} \mathrm{E}[h_{1,i}h_{2,i}]\lambda_i\Delta t_i \\
&\qquad \xrightarrow{ims} \textstyle\int_{t_0}^{t} \mathrm{E}\left[h_1(X(s),s)h_2(X(s),s)\right]\lambda(s)ds \ ,
\end{aligned}$$

giving the desired conclusion except for replacing the approximately equals $(\simeq)$ by the mean square limit as $n \to \infty$

5. The Itô-variance stochastic jump integral follows immediately from part (d) for the Itô-covariance stochastic jump integral by replacing the functions h_1 and h_2 by h. This result (3.30) is also called *Itô isometry* or *martingale isometry* since $\widehat{P}(t)$ is a martingale.

□

3.3 Conclusion

In this chapter, the foundations have been laid for the integrals of the third type in the integrated SDE (2.2), i.e., using the stochastic jump integral of Itô of Definition 2.8 extended to the more general case and defined in Definition 3.1 at the beginning of this chapter:

$$\int_0^t h(X(s),s)dP(s) \overset{ims}{=} \operatorname*{ms\,lim}_{n\to\infty}\left[\sum_{i=0}^{n} h(X(t_i),t_i)dP(t_i)\right] ,$$
$$= \sum_{k=1}^{P(t)} h(X(T_k^-),T_k^-) \tag{3.31}$$

where $X(t)$ in the integrand function h has an implied dependence on the simple Poisson jump process $P(t)$, but also depends on the diffusion process $W(t)$. The integrand process $h(X(t),t)$ is also assumed to have a bounded mean integral of squares (3.2),

$$\mathrm{E}\left[\int_0^t h^2(X(s),s)ds\right] < \infty ,$$

with the usual grid partition specifications on $[0,t]$. However, as previously explained, the Poisson jump process fits within the framework of Itô stochastic integration since it is piece-wise continuous. The stochastic jump integration rule (3.31) has been motivated and illustrated by a number of examples using functions and powers of the jump process $P(t)$.

3.4 Exercises

1. Show that the power rules for stochastic integration for Poisson noise can be written as the recursions,

$$\int_0^t P^m(s)dP(s) = \frac{1}{m+1}\left(P^{m+1}(t) - \sum_{k=2}^{m+1}\binom{m+1}{k}\int_0^t P^{m+1-k}(s)dP(s)\right),$$

using the jump form of the stochastic chain rule and the binomial theorem.

   (a) Illustrate the application of the formulae for $P(t)$ to confirm the results for $m = 0{:}3$ in Table 3.1.

   (b) Alternatively, show the general result for $m \geq 1$.

2. Show that the partial sums of the **geometric series** can be summed as

$$S_n(x) \equiv \sum_{k=0}^{n} x^k = T_n(x) \equiv \left\{ \begin{array}{ll} \frac{1-x^{n+1}}{1-x}, & x \neq 1 \\ n+1, & x = 1 \end{array} \right\} , \tag{3.32}$$

for integers $n \geq 0$ by showing that the difference of the defined summation, $\Delta S_n(x)$, and the difference of the summed answer, $\Delta T_n(x)$, to the far right are the same and that the discrete initial conditions are the same at $n = 0$.

3. Show the mean square limit for the product of $dP(t)$ and $dW(t)$ in (3.17-3.18) by proving that

$$\mathrm{Var}\left[\sum_{i=0}^{n} \Delta P_i \Delta W_i\right] \to 0\,, \tag{3.33}$$

as $n \to +\infty$ and $\delta t_n \to 0^+$.

4. Show the mean square limit for the Poisson differential power $(dP)^m(t)$ version of the Zero-One jump law in Theorem 3.17 by showing that

   (a) Let $M_m(\Delta\Lambda_j) = \mathrm{E}[(\Delta P_j)^m]$ be the mth power of the jth Poisson increment for $\Delta\Lambda_j = M_1(\Delta\Lambda_j)$ and bounded maximum jump rate $\lambda^* = \max_t[\lambda(t)]$, with non-negative integers m and j, then $M_m(u)$ satisfies the recursion relation

$$M_{m+1}(u) = u \cdot (M_m(u) + M'_m(u)\,. \tag{3.34}$$

   (b) Let $M_m(u) = u + K_m(u)u^2$, then $K_m(u) \geq 0$, $K_m(u) = \mathrm{O}(1)$ and $K'_m(u) = \mathrm{O}(1)$, both as $u \to 0^+$.

   (c) Finally,

$$\mathrm{E}\left[\left(\sum_{i=0}^{n} ((\Delta P_i)^m - \Delta P_i)\right)^2\right] \to 0\,, \tag{3.35}$$

   as $n \to +\infty$ and the mesh $\delta t_n \to 0^+$ for $m \geq 1$. Hence,

$$(dP)^m(t) \stackrel{\mathrm{dt}}{=} dP(t),$$

   the symbolic version of the mean square limit form of the Zero-One law.

5. Show that

$$\int_0^t e^{aP(s)} dP(s) = \left\{ \begin{array}{cc} \frac{e^{aP(t)}-1}{e^a-1}, & e^a \neq 1 \text{ or } a \neq 0 \\ P(t), & e^a = 1 \text{ or } a = 0 \end{array} \right\}\,, \tag{3.36}$$

   for real constant a, in two ways, showing that they give the same answers,

   (a) Using the Poisson sum form $\sum_{k=0}^{P(t)-1} h(k)$ of Theorem 3.8 and the geometric series partial sum results in (3.32) of this Exercise section.

   (b) Using the Zero-One Jump Law and the Fundamental Theorem of Jump Calculus 3.4 (b) applied to $d\exp(aP(t))$ to evaluate the integral.

Suggested References for Further Reading

- Çinlar, 1975 [55].
- Protter 1990, [232].
- Snyder and Miller, 1991 [252].
- Tuckwell, 1995 [270].

Chapter 4

Stochastic Calculus for Jump-Diffusions: Elementary SDEs

In Chapter 2 for diffusions and Chapter 3 for jumps, the foundations of Itô stochastic jump-diffusion integrals have been given. In Table 2.1 of Chapter 2, the mean square differential forms for diffusions, powers of $dW(t)$ and dt, were summarized, such that higher order differential forms are zero symbolically in the Itô mean square sense to dt-precision, for example

$$(dW)^3(t) \stackrel{\text{dt}}{=} 0.$$

In Table 3.2 of of Chapter 3, the mean square differential forms for Poisson jumps, powers of $dP(t)$ and dt, were summarized. Different from diffusion differential forms, the powers of $(dP)(t)$ are generally non-zero except when multiplied by a positive power of dt, but have the Zero-One Jump law property that $(dP)^m(t) \stackrel{dt}{=}_{zol} dP(t)$ for integers $m > 0$.

Similar rules apply in the algebra of deterministic differentials and in constructing deterministic models, e.g., terms with the factor $(dt)^2$ are neglected compared to terms with just the factor dt in both deterministic and stochastic differential models. For stochastic differentials, the non-differentiability of $W(t)$ and the jump discontinuities of $P(t)$ produce notable exceptions from deterministic differential rules.

For the mean square limits of more general functions and their approximations where there is insufficient information for a proof, the mean square integrability assumption and the PWCA Mean Square Limits (2.44) Assumption 2.15 will be assumed to be satisfied. This is applicable to both diffusion and jump integrals and this will be an underlying assumption throughout this chapter. However, the primary focus of this chapter will be faster, efficient formal stochastic calculations.

4.1 Diffusion Process Calculus Rules

The most basic rule (2.32) for diffusions in the Itô mean square sense is

$$(dW)^2(t) \stackrel{\text{dt}}{=} dt\,, \tag{4.1}$$

symbolically, while the higher order differential forms are zero in the Itô mean square sense, beginning with

$$(dW)^3(t) \stackrel{\text{dt}}{=} 0, \quad dtdW(t) \stackrel{\text{dt}}{=} 0 \text{ and } (dt)^2 \stackrel{\text{dt}}{=} 0\,,$$

using summary Table 2.1.

Another basic rule or principle is the use of increments both for increments themselves in single steps of Δt,

$$\begin{aligned}\Delta G(W(t),t) &\equiv G(W(t+\Delta t),t+\Delta t) - G(W(t),t)\\ &= G(W(t)+\Delta W(t),t+\Delta t) - G(W(t),t)\,,\end{aligned} \tag{4.2}$$

with functions of the form $G(w(t),t)$ and $\Delta W(t) \equiv W(t+\Delta t) - W(t)$, as well as for differentials as increments,

$$\begin{aligned}dG(W(t),t) &\equiv G(W(t+dt),t+dt) - G(W(t),t)\\ &= G(W(t)+dW(t),t+dt) - G(W(t),t)\,,\end{aligned} \tag{4.3}$$

with $dW(t) \equiv W(t+dt) - W(t)$.

The increment (4.2) and differential (4.3) rules can be used, with the rest of Table 2.1, to develop a fast and efficient procedure for deriving stochastic formulas. When there are problems it is best to go back and check the result by more precise Itô stochastic integral procedures.

4.1.1 Functions of Diffusions Alone, $G(W(t))$

Some simple calculus-like examples are given below as an introduction. Although we could just as well work with differentials at the start, we will start with the increments at t and then get the differential form in the limit as $\Delta t \to 0^+$, but later switch to starting with the differential forms as increment forms in dt.

Examples 4.1.

- ***Cubic Integral:***

  $$\Delta\left[W^3\right](t) = (W+\Delta W)^3(t) - W^3(t) = \left(3W^2\Delta W + 3W(\Delta W)^2 + (\Delta W)^3\right)(t),$$

  using the cubic expansion. As $\Delta t \to 0$, $(\Delta W)^2(t) \to (dW)^2(t) \stackrel{\text{dt}}{=} dt$ *and* $(\Delta W)^3(t) \to (dW)^3(t) \stackrel{\text{dt}}{=} 0$, *so the corresponding differential form is*

  $$d\left[W^3\right](t) \stackrel{\text{dt}}{=} \left(3W^2dW + 3Wdt\right)(t)\,.$$

  The first term is the deterministic differential, since $d(w^3) = 3w^2dw$, *but with an Itô stochastic correction* $3W(t)dt$. *Solving for* $W^2(t)dW(t)$, *the Itô integral of the square of* $W(t)$ *yields*

  $$\int_{t_0}^{t} W^2(s)dW(s) \stackrel{ims}{=} \frac{1}{3}\left(W^3(t) - W^3(t_0)\right) - \int_{t_0}^{t} W(s)ds.$$

The Itô integral of w^2 is reduced to a stochastic-Riemann integral and the Itô correction to the Riemann integral of w^2 and looks simple, but cannot be Itô-integrated exactly and must be numerically simulated if needed.

- ***General Integer Power Integral:***
By using the full binomial theorem (B.152),

$$\begin{aligned}\Delta\left[W^{m+1}\right](t) &= (W+\Delta W)^{m+1}(t) - W^{m+1}(t) \\ &= \sum_{i=0}^{m} \left(\begin{array}{c} m+1 \\ i \end{array} \right) W^i(t) \Delta W^{m+1-i}(t)\,,\end{aligned}$$

where the passage to the limit as $\Delta t \to 0$ and the Itô mean square limit leading to the integral form

$$\int_0^t W^m(s) dW(s)$$

has been left as Exercise 5 on page 126 in Section 4.5.

- ***Exponential Integral:***
Using laws of exponents and the first few terms of the exponential expansion (B.53), going directly to the formal differential form and skipping the more general increment form to expedite applied stochastic calculations,

$$\begin{aligned}d\left[e^W\right](t) &= \left(e^{W+dW} - e^W\right)(t) = \left(e^W\left(e^{dW}-1\right)\right)(t) \\ &\stackrel{\mathrm{dt}}{=} \left(e^W\left(dW + \frac{1}{2}(dW)^2\right)\right)(t)\,,\end{aligned}$$

neglecting differential forms that are zero in the Itô mean square limit, such as $dW^3(t) \stackrel{\mathrm{dt}}{=} 0$, $dtdW(t) \stackrel{\mathrm{dt}}{=} 0$, $(dt)^2 \stackrel{\mathrm{dt}}{=} 0$ and higher powers with this zero mean square limit property.

Using the basic mean square limit differential form (4.1), $(dW)^2(t) \stackrel{\mathrm{dt}}{=} dt$, so

$$d\left[e^W\right](t) \stackrel{\mathrm{dt}}{=} \left(e^W\left(dW + \frac{1}{2}dt\right)\right)(t)\,. \tag{4.4}$$

This is almost like the deterministic differential, $d(e^w) = e^w dw$, but here with an Itô stochastic correction $e^{W(t)}dt/2$. Solving for $e^{W(t)}dW(t)$, the Itô integral of the exponential of $W(t)$ yields the implicit integration

$$\int_{t_0}^t e^{W(s)} dW(s) \stackrel{ims}{=} e^{W(t)} - e^{W(t_0)} - \frac{1}{2}\int_{t_0}^t e^{W(s)} ds. \tag{4.5}$$

As with the integral of w^2, the Itô integral of e^w cannot be Itô-integrated exactly and must be numerically simulated if needed. The simulations are presented in Fig. 4.2 for the Itô partial sums form

$$S_{i+1} = \sum_{j=0}^{i} \exp(W_i)\Delta W_i$$

for $t = t_{i+1} = (i+1)\Delta t$ for $t_0 = 0$ evenly spaced using $\Delta t_i = \Delta t$ where

$$W_{i+1} = \sum_{j=0}^{i} \Delta W_j$$

and the error

$$E_{i+1} = S_{i+1} - R_{i+1}$$

between the partial sums S_{i+1} and the difference approximation to the right hand side

$$R_{i+1} = \exp(W_{i+1}) - 1 - \frac{1}{2}\sum_{j=0}^{i} \exp(W_j)\Delta t$$

of (4.5), noting that $t_0 = 0$ so $\exp(W(t_0)) = 1$. Remember that the cumulative noise W_i must always be approximated by sums of simulated independent increments ΔW_j for $j = 0 : i-1$.

In the differential (4.4) of the pure exponential there is a clue to an exact differential in the Itô mean square sense, since the factor $(dW + dt/2)$ suggests subtracting $t/2$ from $W(t)$. In fact,

$$d\left[e^{W(t)-t/2}\right] \stackrel{\text{dt}}{=} e^{W(t)-t/2} dW(t)\,. \tag{4.6}$$

So

$$\int_0^t e^{W(s)-s/2} dW(s) \stackrel{ims}{=} e^{W(t)-t/2} - 1\,. \tag{4.7}$$

In forthcoming Example 4.1.3 on p. 90, a method for systematically finding general exact integrals is presented, provided they exist. The simulations are presented in Fig. 4.1 on p. 95 for the Itô partial sums form $S_{i+1} = \sum_{j=0}^{i} \exp(W_i - t_i/2)\Delta W_i$ and the error between the partial sums and the difference approximation of (4.7).

More general rules can be derived by the same techniques.

Rule 4.2. ***Chain Rule for $G(W(t))$:***
Let $G(w)$ be twice continuously differentiable, then the differential form of the Itô stochastic ***chain rule for $G(W(t))$*** *is*

$$dG(W(t)) \stackrel{\text{dt}}{=} G'(W(t))dW(t) + \frac{1}{2}G''(W(t))dt\,, \tag{4.8}$$

corresponding to the integral form of the Itô stochastic ***chain rule for $G(W(t))$****,*

$$G(W(t)) \stackrel{ims}{=} G(W(t_0)) + \int_{t_0}^{t} G'(W(s))dW(s) + \frac{1}{2}\int_{t_0}^{t} G''(W(s))ds\,, \tag{4.9}$$

for $0 \le t_0 \le t$.

Sketch of Proof. Assuming $G(w)$ is twice continuously differentiable in the argument w, then $G(W(t))$ has the differential:

$$\begin{aligned} dG(W(t)) &= G(W(t) + dW(t)) - G(W(t)) \\ &\stackrel{\text{dt}}{=} G'(W(t))dW(t) + \frac{1}{2}G''(W(t))(dW)^2(t)\,, \end{aligned}$$

Taking the Itô mean square limit neglecting error terms that are zero in the mean square limit, such as $dW^3(t)$, $dtdW(t)$ and $(dt)^2$, then using $(dW)^2(t) \stackrel{\text{dt}}{=} dt$ yields the differential form (4.8) of the Itô stochastic **chain rule for $G(W(t))$**. The last term in the second derivative is the Itô stochastic correction to the deterministic chain rule. Immediately, we have Itô stochastic integral form (4.9), which provides substantial meaning to the symbolic differential form. □

Rewriting (4.9) yields the **fundamental theorem of calculus** according to Itô [149] version:

Corollary 4.3. ***Itô's Fundamental Theorem of Calculus for Stochastic Diffusions***
Let $G(w)$ be twice continuously differentiable, then

$$\int_{t_0}^{t} G'(W(s))dW(s) \stackrel{ims}{=} G(W(t)) - G(W(t_0)) - \frac{1}{2}\int_{t_0}^{t} G''(W(s))ds\,, \quad (4.10)$$

Remark 4.4. *Recall the more elementary integral of a differential form of the Fundamental Theorem of Stochastic Diffusion Calculus in (2.38), which in fact leads to the exact part of the Itô version, using G in (2.38),*

$$\int_{t_0}^{t} dG(W(s)) \stackrel{ims}{=} G(W(t)) - G(W(t_0))\,.$$

4.1.2 Functions of Diffusions and Time, $G(W(t), t)$

Rule 4.5. ***Chain Rule for $G(W(t), t)$:***
Let $G(w, t)$ be twice continuously differentiable in w and once continuously differentiable in t, then the differential Itô stochastic ***chain rule for $G(W(t), t)$*** *is*

$$dG(W(t), t) \stackrel{\text{dt}}{=} \left(G_t + \frac{1}{2}G_{ww}\right)(W(t), t)dt + G_w(W(t), t)dW(t)\,, \quad (4.11)$$

corresponding to the integral form of Itô stochastic ***chain rule for*** *$G(W(t), t)$,*

$$\begin{aligned} G(W(t), t) &\stackrel{ims}{=} G(W(t_0), t_0) + \int_{t_0}^{t} G_w(W(s), s)dW(s) \\ &\quad + \int_{t_0}^{t} \left(G_t + \frac{1}{2}G_{ww}\right)(W(s), s)ds\,, \end{aligned} \quad (4.12)$$

for $0 \le t_0 \le t$.

Sketch of Proof. Assuming $G(w,t)$ is twice continuously differentiable in the argument w and once continuously differentiable in t, then, using a mean square order modification of the Taylor approximation in (B.183), $G(W(t),t)$ has the differential:

$$\begin{aligned} dG(W(t),t) &= G(W(t)+dW(t),t+dt) - G(W(t),t) \\ &\stackrel{\text{dt}}{=} G_t(W(t),t)dt + G_w(W(t),t)dW(t) + \frac{1}{2}G_{ww}(W(t),t)(dW)^2(t)\,, \end{aligned}$$

where the partial derivatives are denoted with subscripts, i.e.,

$$G_w(w,t) = \frac{\partial G}{\partial w}(w,t), \quad G_t(w,t) = \frac{\partial G}{\partial t}(w,t), \quad G_{ww}(w,t) = \frac{\partial^2 G}{\partial w^2}(w,t)\,.$$

Taking the Itô mean square limit with $(dW)^2(t) \stackrel{\text{dt}}{=} dt$ and neglecting the higher order differential forms that are zero in the Itô mean square sense, such as $dW^3(t)$, $dtdW(t)$ and $(dt)^2$ yields (4.11), which is called the Itô stochastic **chain rule for $G(W(t),t)$**. Again the last term in the second derivative is the Itô stochastic correction to the deterministic chain rule. Translating the symbolic differential form into the substantial Itô stochastic integral form gives (4.12). □

Remarks 4.6. ***Functions, Values and Partial Derivatives:***

- *For readers without much PDE background, there are certain concepts that are important and subtle differences in the function and its values $G(w,t)$, particularly when there are two or more independent variables, such as the $w = W(t)$ and t in $G(W(t),t)$, that do not arise when there is just one independent variable, such as x in $y = f(x)$. Another complication is the $W(t)$ is a non-differentiable function so we never form its derivative, but only compute its differential $dW(t)$, and that is best done formally by the increment form of the differential.*

- *The symbol G denotes a function specified by a set of rules for its calculation, while $G(w,t)$ is the value of that function with first argument evaluated at w and with the second argument at time t. Similarly, $G(W(t),t)$ is the value of G specified at the random variable $W(t)$ at time t in place of the realized or dummy variable w. Further, $X(t) = G(W(t),t)$ is the path of the state in time and is non-differentiable along with $W(t)$, i.e., $X(t)$ is a composite function in time through both arguments of G, implicitly through $W(t)$ and explicitly through the second argument t.*

- *Using limits of Newton's quotient for derivatives, the partial derivatives of $G(w,t)$ are defined, also giving several alternate notations, at (w,t) as*

$$G_w(w,t) = \frac{\partial G}{\partial w}(w,t) = \left(\frac{\partial G}{\partial w}\right)\Big|_{\substack{t \\ \text{fixed}}}(w,t) = \lim_{\Delta w \to 0} \frac{G(w+\Delta w,t) - G(w,t)}{\Delta w}$$

and

$$G_t(w,t) = \frac{\partial G}{\partial t}(w,t) = \left(\frac{\partial G}{\partial t}\right)\Big|_{\substack{w \\ \text{fixed}}}(w,t) = \lim_{\Delta t \to 0} \frac{G(w,t+\Delta t) - G(w,t)}{\Delta t}\,,$$

provided the limits exist. Hence, partial derivatives with one of the variables fixed are based on the definition of ordinary derivatives.

- *The partial derivatives G_w and G_t are defined as rules based upon the target function rule G. For the topics here, when the first argument is a random variable $w = W(t)$,*

$$\frac{\partial G}{\partial w}(W(t),t)$$

is just G_w evaluated at the first variable $w = W(t)$ after differentiation. We would ***never*** *write $G_{W(t)}$ due to the non-differentiable properties of $W(t)$. Anyway, the partial derivative is calculated first then it is evaluated. For example, $G_w(1,1)$ can be computed if we knew G_w and it had a unique value at $(1,1)$, but $(G(1,1))_w = 0$ since $G(1,1)$ has a fixed, constant value, presumably unique, at $(1,1)$, the order of partial differentiation and partial derivative function evaluation are very important.*

- *Another more relevant example illustrating the difference, is the differential to be multiplied by dt to avoid obtaining the singular derivative of $W(t)$, is*

$$dG(W(t),t) \stackrel{\text{dt}}{=} \left(G_t dt + G_w dW(t) + \frac{1}{2}G_{ww}dt\right)(W(t),t),$$

contains the partial derivative of the function G with respect to t evaluated at $(W(t),t)$,

$$\frac{\partial G}{\partial t}(W(t),t)dt$$

rather than partial derivative with respect to t is written as the derivative of the value $G(W(t),t)$,

$$\frac{\partial G(W(t),t)}{\partial t}dt,$$

which makes no sense since it would involve the derivative of the non-differentiable $W(t)$ in t with probability one (recall Theorem 1.9 on page 9).

Corollary 4.7. *Let $g(W(t),t)$ satisfy the conditions of Definition 2.8 for an Itô stochastic integral and be once continuously differentiable in w. Let $G(w,t)$ be the anti-derivative of $g(w,t)$ with respect to w, i.e., $G_w(w,t) = g(w,t)$, and let $G(w,t)$ be twice continuously differentiable in w, but only once in t. Then,*

$$\int_{t_0}^{t} g(W(s),s)dW(s) \stackrel{ims}{=} G(W(t),t) - G(W(t_0),t_0) - \int_{t_0}^{t}(G_t + 0.5 * g_w)(W(s),s)ds\,, \tag{4.13}$$

for $0 \leq t_0 \leq t$.

Proof. This follows directly from (4.12) by rearranging terms, since $G_w = g$ and $G_{ww} = g_w$. □

Remark 4.8. *Thus, the Itô stochastic diffusion integral of* $g(W(t), t)$ *can be reduced to an exact integral* $G(W(t), t) - G(W(t_0), t_0)$ *with respect to* w *less a quasi-deterministic Riemann integral over the diffusion shifted drift function* $(G_t + 0.5 * g_w)(W(t), t)$. *Thus, if the partial differential equation* $(G_t + 0.5 * g_w)(w, t) = 0$ *is valid with* $g_w(w, t) = G_{ww}(w, t)$, *then the integral of* $g(W(t), t)$ *is equal to the exactly integrated part* $G(W(t), t) - G(W(t_0), t_0)$ *in the Itô mean square sense. This idea can be the basis for constructing exact stochastic diffusion integrals.*

Example 4.9. ***Merton's Analysis of Black-Scholes Option Pricing Model:*** *At this point in the text, a good application in finance is the survey of Merton's [201] (Merton [203, Chapter 8]) analysis of the Black-Scholes [34] financial options pricing model in Section 10.2 of Chapter 10. This survey follows the tone of this book, although Merton's model has several state dimensions, the bond, the stock and the option. While multi-dimension SDEs will be covered in the next chapter, Chapter 5, this treatment will serve as motivation for the next chapter, which contains details not in Merton's paper.*

4.1.3 Itô Stochastic Natural Exponential Construction

From the differential of $\exp(W(t))$ in (4.4) it is seen that the stochastic exponential is not like the deterministic natural exponential, where the derivative is proportional to the original function, e.g., the natural exponential e^x in the natural base e has the differential property:

$$d\left(e^x\right) = e^x dx\,,$$

returning the original function times dx, and has the following inverse relationship to the natural logarithm

$$e^{\ln(x)} = x$$

for $x > 0$, whereas when $b > 0$ and in particular $b \neq e$ for the base b, then

$$d\left(b^x\right) = d\left(e^{x\ln(b)}\right) = b^x \ln(b) dx\,,$$

returning an additional factor $\ln(b)$.

For more generality, consider the deterministic model

$$d\left(e^{ax}\right) = ae^{ax} dx\,,$$

where the parameter a is a non-zero constant. The corresponding stochastic model is the process $X(t) = G(W(t), t)$ such that

$$dX(t) = dG(W(t), t) \stackrel{\text{dt}}{=} aG(W(t), t)dW(t) = aX(t)dW(t)\,. \tag{4.14}$$

The explicit t dependence is needed to avoid correction factors in dt. Applying the appropriate stochastic chain rule (4.11) to illustrate a technique for inverting the chain rule to get the desired model in terms of the composite function G,

$$\begin{aligned} aG(W(t),t)dW(t) &\stackrel{\text{dt}}{=} dG(W(t),t) \\ &\stackrel{\text{dt}}{=} \left(G_t(W(t),t) + \frac{1}{2}G_{ww}(W(t),t)\right)dt + G_w(W(t),t)dW(t)\,. \end{aligned}$$

Since the differentials, $dW(t)$ and dt, can be independently varied in this equation, the coefficients of $dW(t)$ and dt can be separately set equal to their values on both sides of the equation (dropping the arguments of G for simplicity):

$$G_w = aG \;\text{ and }\; G_t + \frac{1}{2}G_{ww} = 0\,. \tag{4.15}$$

The solution of the first partial differential equation (PDE), $G_w = aG$, in (4.15), being effectively an ordinary differential equation (ODE) with t held fixed, is

$$G(w,t) = A(t)e^{aw}\,, \tag{4.16}$$

since $d(e^{-aw})/dw = -ae^{-aw}$ (differentiation is allowable for a regular continuous, i.e., non-stochastic, function) so

$$d\left(e^{-aw}G\right)_w = e^{-aw}\left(G_w - aG\right) = 0\,,$$

which shows that (4.16) satisfies the first PDE by substitution, $e^{-aw} \neq 0$. Here, $A(t)$ is a **function of integration** since the differential equation is only in w and t is arbitrary, although held fixed in the equation. Given a differentiable function $F(w,t)$, the notation $F_w(w,t) = 0$ is short hand for the partial deriviative

$$\left(\frac{\partial F}{\partial w}\right)_{\substack{t \\ \text{fixed}}}(w,t) = 0.$$

This means that $F(w,t) = A(t)$ for some function A of t, since t is held fixed in the partial differentiation with respect to w.

Upon substituting this current functional form into the second partial differential equation, $G_t + 0.5G_{ww} = 0$, using

$$\begin{aligned} \left(A(t)e^{aw}\right)_t &= e^{aw}(A(t))_t = A'(t)e^{aw}\,, \\ \left(A(t)e^{aw}\right)_{ww} &= A(t)(e^{aw})_{ww} = a^2A(t)e^{aw}\,, \end{aligned}$$

then

$$A'(t)e^{aw} + \frac{a^2}{2}A(t)e^{aw} = 0\,.$$

Upon cancelling out the common nonzero factor e^{aw},

$$A'(t) + \frac{a^2}{2}A(t) = 0\,, \tag{4.17}$$

and solving for the function of integration yields

$$A(t) = Ce^{-a^2t/2}\,, \tag{4.18}$$

where C is a **genuine constant of integration**.

Remark 4.10. *Note that an ultimate test of a solution of a differential equation solution is the* ***substitution test****, i.e., substituting the solution back into the equation and verifying that the equation and any conditions are satisfied.*

For (4.18), substitution into the ODE (4.17) leads to

$$A'(t) + \frac{a^2}{2}A(t) = Ce^{-a^2t/2} \cdot \left(-\frac{a^2}{2} + \frac{a^2}{2}\right) = 0\,.$$

Assembling the parts of the solution back together, we obtain the Itô general stochastic form of the natural exponential (exponential in the natural base e),

$$X(t) = G(W(t), t) = Ce^{aW(t)-a^2t/2}\,, \tag{4.19}$$

systematically deriving what previously was a guess in (4.6). The extra exponential term $(-a^2t/2)$ is the special Itô correction that forces the simple linear growth model $dX(t) = aX(t)dt$ for the exponential growth in the diffusion $W(t)$.

Since $W(0^+) = 0$ with probability one, $X(0^+) = G(0, 0^+) = C$, with probability one, is the initial value of the state $X(t)$, while a is a rate of growth. The basic moments of the state trajectory can be calculated by using the density $\phi_{W(t)}(w)$ for $W(t)$ in (1.7).

Some of the details are given to illustrate the use of the **completing the square** technique when computing exponential moments with respect to normal distributions. An illustration of the completing the square technique is presented for the expectation of an exponential whose exponent is linear (or affine) in $W(t)$, i.e., $\exp(a(t)W(t) + b(t))$.

Lemma 4.11. ***Completing the Square for*** $\mathbf{E}[\boldsymbol{K(t)\exp(a(t)W(t) + b(t))}]$***:*** *Let $a(t) \neq 0$, $b(t)$ and $K(t) \neq 0$ be bounded deterministic functions of t, then*

$$\mathrm{E}\left[K(t)e^{a(t)W(t)+b(t)}\right] = K(t)e^{a^2(t)t/2+b(t)} \tag{4.20}$$

Proof. Since the Wiener process density,

$$\phi_{W(t)}(w) = \frac{1}{\sqrt{2\pi t}}e^{-w^2/(2t)},$$

$-\infty < w < +\infty$, from (1.7), is essentially a function of the sampled dummy variable w and t is only a parameter that we can hold fixed during the integration, the deterministic functions of time are treated as constants. By the laws of exponents,

the exponent of the density and the exponent of the argument of the expectation with the dummy variable substitution $W(t) = w$ are added together and combined to obtain a complete square of all w terms,

$$-w^2/(2t) + a(t)w + b(t) = -(w - a(t)t)^2/(2t) + a^2(t)t/2 + b(t).$$

Thus,

$$\begin{aligned}
\mathrm{E}\left[K(t)e^{a(t)W(t)+b(t)}\right] &= K(t)\frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{+\infty} e^{-(w-a(t)t)^2/(2t)+a^2(t)t/2+b(t)}dw \\
&= K(t)e^{a^2(t)t/2+b(t)}\frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{+\infty} e^{-v^2/(2t)}dv \\
&= K(t)e^{a^2(t)t/2+b(t)}\mathrm{E}[1]\,, \\
&= K(t)e^{a^2(t)t/2+b(t)}\,,
\end{aligned}$$

where the fixed part of the integral with exponent $(a^2(t)t/2 + b(t))$ has been separated out and the change of variables $v = w - a(t)t$ with $dv = dw$, t being fixed, in the integral has been used to transform the completed square part of the expectation integral as one for conservation of probability $\mathrm{E}[1] = 1$ for the standard Wiener process. □

The mean state $X(t)$ using Lemma 4.11 is

$$\mathrm{E}\left[Ce^{aW(t)-a^2t/2}\right] = C = X(0^+)\,, \tag{4.21}$$

so the mean trajectory is a constant, at the initial level $X(0^+)$. However, the state variance, again using Lemma 4.11 but with $a(t)$ replaced by $2a$ following application of the variance-expectation identity (B.188), $Var[X] = \mathrm{E}[X^2] - \mathrm{E}^2[X]$, to use the expectation result (4.21), is

$$\begin{aligned}
\mathrm{Var}\left[Ce^{aW(t)-a^2t/2}\right] &= \mathrm{E}\left[\left(Ce^{aW(t)-a^2t/2}\right)^2\right] - \mathrm{E}^2\left[Ce^{aW(t)-a^2t/2}\right] \\
&= C^2\mathrm{E}\left[e^{2aW(t)-a^2t}\right] - C^2 \\
&= C^2\left(e^{a^2t} - 1\right).
\end{aligned}$$

Examining the standard deviation, or square root of the variance,

$$\sigma_{X(t)} = \sqrt{\mathrm{Var}[X(t)]} = C\sqrt{e^{a^2t} - 1} \sim Ce^{a^2t/2}$$

as $t \to \infty$, it is seen that the root mean square (RMS) of stochastic fluctuations grows exponentially with exponent $a^2t/2$ starting initially at $\sigma_{X(0^+)} = 0^+$.

In Fig. 4.1 is an illustration of the simulation of the integral of this natural exponential in the special case

$$I[g](t) = \int_0^t g(W(s), s)dW(s) = \int_0^t e^{W(s)-s/2}dW(s) \overset{ims}{=} e^{W(t)-t/2} - 1\,, \tag{4.22}$$

i.e., when $a = 1 = C$. Also, plotted is the diffusion process $W(t)$ for comparison and the error,

$$E_{i+1} = S_{i+1} - I_{i+1},$$

between the simulation of the integral by Itô finite difference partial sums,

$$S_{i+1} = \sum_{j=0}^{i} g_j \Delta W_j \, ,$$

and the simulation of the exact mean square integral value in (4.22)

$$I_{i+1} = g_{i+1} - 1 \, ,$$

for $i = 0 : n$, where the integrand is

$$g_i = \exp(W_i - t_i/2) \, ,$$

with $W_i = \sum_{j=0}^{i-1} \Delta W_j$ and $t_i = i * \Delta t$ for $i = 0 : n + 1$. Observe that the integral initially tracks the W_i simulated noise, but eventually diverges from it. Also, the error slowly degrades as the time t_i gets long (not shown) in this case for $n = 10,000$ (note that this is an approximate sample size since random sample size is $n + 1 = 10,001$ random increments) and $t = 2.0$. The MATLAB code for the exactly integrable $g(W(t), t)$ in the Itô mean square diffusion integral sense is given in Program C.12 in Appendix C.

In Fig. 4.2, the chain rule formulation of the Itô diffusion integral of the simple exponential $g(W(t), t) = \exp(W(t))$ of Example 4.1 is compared to the Itô partial sums $S_{i+1} = \sum_{j=0}^{i} g_j \Delta W_j$. Unlike the stochastic natural exponential $\exp(W(t) - t/2)$, the simple exponential is not exactly integrable in the Itô mean square sense since the stochastic chain rule introduces a quasi-deterministic regular type integral for the diffusion term

$$-0.5 G_w(w, t) = -0.5 g(w, t) = -0.5 \exp(w) \, .$$

The partially integrated chain rule form is thus

$$I_{i+1} = \exp(W_i) - 1 - 0.5 * \sum_{j=0}^{i} \exp(W_j) \Delta t \, , \tag{4.23}$$

with $G_t(w, t) = 0$. In the figure the error $E_{i+1} = S_{i+1} - I_{i+1}$ between the two approximations of the integral and the underlying diffusive noise $W(t)$. The error is very small for a sample size of $n = 10,000$. The integration significantly dampens the fluctuations in the original noise $W(t)$. The MATLAB code for this figure is given in Program C.13 of Appendix C.

4.1.4 Transformations of Linear Diffusion SDEs:

Consider the diffusion SDE, linear in the state process $X(t)$, with time-dependent coefficients,

$$dX(t) = X(t) \left(\mu(t) dt + \sigma(t) dW(t) \right) \, , \tag{4.24}$$

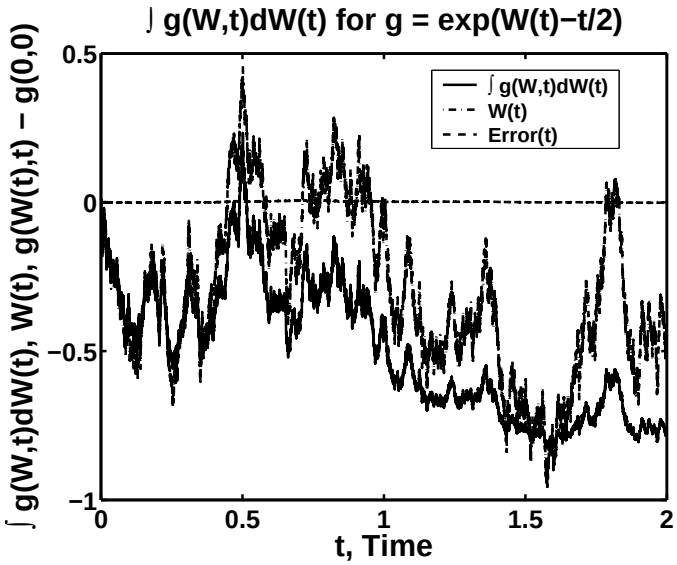


Figure 4.1. *Example of a simulated Itô discrete approximation to the stochastic diffusion integral* $I_n[g](t_{i+1}) = \sum_{j=0}^{i} g_j \Delta W_j$ *for* $i = 0 : n$*, using the MATLAB* ***randn*** *with sample size* $n = 10,000$ *on* $0 \leq t \leq 2.0$*. Presented are the simulated Itô partial sums* S_{i+1}*, the simulated noise* W_{i+1} *and the error* E_{i+1} *relative to the exact integral,* $I^{(\mathrm{ims})}[g](t_{i+1}) \stackrel{ims}{=} \exp(W_{i+1} - t_{i+1}/2) - 1$*, in the Itô mean square sense.*

where the initial condition is $X(t_0) = x_0 > 0$ with probability one, $\mu(t)$ is called the **drift** or deterministic coefficient and $\sigma(t)$ is called the **volatility** or standard deviation of the diffusion term. The **diffusion coefficient** is usually defined as $\mathcal{D} = \sigma^2(t)/2$, so $\sigma(t) = \sqrt{2\mathcal{D}}$. The linear form of (4.24) is sometimes called the **multiplicative noise** case, the state $X(t)$ multiplies the stochastic terms, and the word **noise** referring to the randomness or stochastic properties here. In the deterministic case, transforming the state variable to its logarithm makes the right hand side independent of the transformed state variable, so let

$$Y(t) = F(X(t)) \equiv \ln(X(t)) \, .$$

Since we have F depending on $X(t)$ rather than $W(t)$, we go back to the basic treatment of the change as an increment and expand the increment to second order,

$$\begin{aligned} dY(t) &= \log(X(t) + dX(t)) - \log(X(t)) \\ &\stackrel{\mathrm{dt}}{=} \frac{1}{X(t)} dX(t) - \frac{1}{2X^2(t)} (dX)^2(t) \\ &\stackrel{\mathrm{dt}}{=} (\mu(t)dt + \sigma(t)dW(t)) - 0.5\sigma^2(t)(dW)^2(t) \\ &\stackrel{\mathrm{dt}}{=} (\mu(t) - 0.5\sigma^2(t))dt + \sigma(t)dW(t) \, , \end{aligned} \tag{4.25}$$

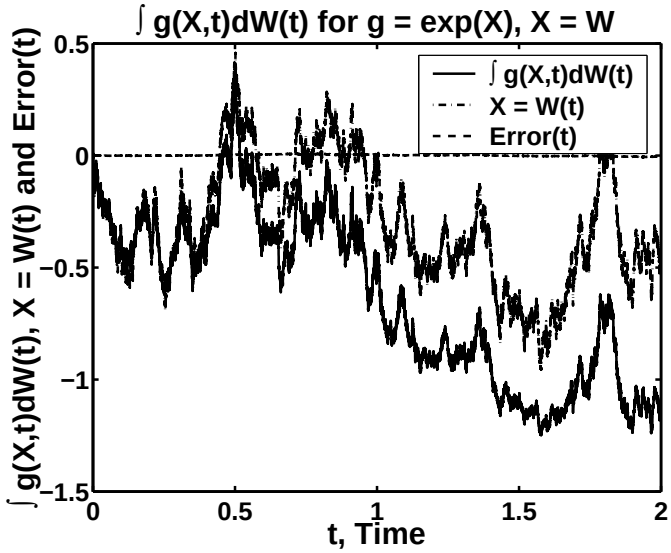


Figure 4.2. *Example of a simulated Itô discrete approximation to the stochastic diffusion integral* $I_n[g](t_{i+1}) = \sum_{j=0}^{i} g_j \Delta W_j$ *for* $i = 0 : n$*, using the MATLAB* `randn` *with sample size* $n + 1 = 10,001$ *on* $0 \leq t \leq 2.0$*. Presented are the simulated Itô partial sums* S_{i+1}*, the simulated noise* W_{i+1} *and the error* E_{i+1} *relative to the stochastic chain rule partially integrated form,* I_{i+1} *given in the text (4.23).*

where we again used $(dW)^2(t) \stackrel{\text{dt}}{=} dt$ and dropped terms zero in the mean square. Use has been made of the following partial derivatives,

$$F_t(X(t)) \equiv 0\,, \quad F_x(X(t)) = 1/X(t)\,, \quad F_{xx}(X(t)) = -1/X^2(t)\,.$$

The final line in (4.25) is also called **additive noise** since it just adds to the state value and can be immediately integrated, as opposed to the multiplicative noise in the original SDE in (4.24). In the above derivation, the Itô stochastic correction on the drift $\mu(t)$ is the negative of the **diffusion coefficient** $\sigma^2(t)/2$. The final right hand side (4.25) defines a differential simple Gaussian process (B.24) with infinitesimal mean $(\mu(t) - 0.5\sigma^2(t))dt$ and infinitesimal variance of $\sigma^2(t)dt$. The **infinitesimal mean** here is defined as

$$\mathrm{E}[dY(t)] \tag{4.26}$$

and the **infinitesimal variance** is defined as

$$\mathrm{Var}[dY(t)]\,, \tag{4.27}$$

in each case neglecting orders smaller than ord(dt). An alternate method of deriving (4.25) is to use the Itô stochastic chain rule for $G(W(t))$, but with $W(t)$ replaced by

$X(t)$, subsequently expanding the differentials $dX(t)$ and $(dX)^2(t)$, then replacing them by the SDE in (4.24) and neglecting any terms that are zero in the mean square.

Since the right hand side of (4.25) does not depend on the state $Y(t)$, we can immediately integrate for $Y(t)$ given the coefficient functions leading to

$$Y(t) = y_0 + \int_{t_0}^{t} (\mu(s) - 0.5\sigma^2(s))ds + \int_{t_0}^{t} \sigma(s)dW(s)\,, \tag{4.28}$$

where $y_0 = \ln(x_0)$. Exponentiation leads to the formal solution for the original state,

$$X(t) = x_0 \exp\left(\int_{t_0}^{t} (\mu(s) - 0.5\sigma^2(s))ds + \int_{t_0}^{t} \sigma(s)dW(s) \right)\,. \tag{4.29}$$

Linear Diffusion SDEs with Constant Coefficients:

If the SDE has **constant coefficients**, $\mu(t) = \mu_0$ and $\sigma(t) = \sigma_0$, while letting $t_0 = 0$, then the solution is simpler,

$$X(t) = x_0 \exp\left((\mu_0 - 0.5\sigma_0^2)t + \sigma_0 W(t) \right)\,. \tag{4.30}$$

Note that if $X(0^+) = x_0$ is initially positive as declared, then the solution $X(t)$ will never become negative by the property of the exponential for real arguments and the transformation $Y(t) = \ln(X(t))$ is proper with $X(t) > 0$. The state $X(t)$ positivity feature is very important in biological and financial applications. Aside from time dependence, this is just a shift by the drift of the exponent, as in the Itô stochastic exponential in (4.19).

In the **additive noise** case, borrowing the exponent form in (4.25), the relation between the new and old values of Y is computed by adding the noise,

$$Y(t + \Delta t) = Y(t) + (\mu_0 - 0.5\sigma_0^2)\Delta t + \sigma_0 \Delta W(t)\,, \tag{4.31}$$

or recursively in the time-step Δt_i from t_i to t_{i+1} and then summing the recursion,

$$\begin{aligned} Y_{i+1} &= Y_i + (\mu_0 - 0.5\sigma_0^2)\Delta t_i + \sigma_0 \Delta W_i \\ &= y_0 + \sum_{j=0}^{i} \left((\mu_0 - 0.5\sigma_0^2)\Delta t_j + \sigma_0 \Delta W_j \right)\,. \end{aligned}$$

So taking the expectation,

$$\begin{aligned} \mathrm{E}[Y_{i+1}] &= y_0 + \sum_{j=0}^{i} \left(\mu_0 - 0.5\sigma_0^2 \right) \Delta t_j \\ &= y_0 + (\mu_0 - 0.5\sigma_0^2) \sum_{j=0}^{i} \Delta t_j. \end{aligned} \tag{4.32}$$

This result should be compared to the corresponding deterministic additive or arithmetic recursion with constant a,

$$z_{i+1} = z_i + a, \quad \Longrightarrow \quad z_{i+1} = z_0 + (i+1) \cdot a,$$

so the corresponding additive parameter to the mean in (4.32)

$$\mathrm{E}[Y_{i+1}] = y_0 + (i+1)(\mu_0 - 0.5\sigma_0^2)\overline{\Delta t_i}^{(am)}, \tag{4.33}$$

where

$$\overline{\Delta t_i}^{(am)} = \frac{1}{i+1}\sum_{j=0}^{i} \Delta t_j$$

is the **arithmetic mean** of the first $(i+1)$ time steps Δt_j for $j = 0 : i$.

Whereas, the **multiplicative noise** property can be seen by rewriting (4.30) as a single step,

$$X(t+\Delta t) = X(t)\exp\left((\mu_0 - 0.5\sigma_0^2)\Delta t + \sigma_0 \Delta W(t)\right) \tag{4.34}$$

so the new noise exponential contribution from $\Delta W(t)$ multiplies the current value of the solution $X(t)$ to produce the new value $X(t+\Delta t)$. The corresponding recursive form in the time-step Δt_i from t_i to t_{i+1}, followed by a summing of the recursion,

$$\begin{aligned} X_{i+1} &= X_i \exp\left((\mu_0 - 0.5\sigma_0^2)\Delta t_i + \sigma_0 \Delta W_i\right) \\ &= x_0 \exp\left(\sum_{j=0}^{i}\left((\mu_0 - 0.5\sigma_0^2)\Delta t_j + \sigma_0 \Delta W_j\right)\right) \\ &= x_0 \prod_{j=0}^{i} \exp\left((\mu_0 - 0.5\sigma_0^2)\Delta t_j + \sigma_0 \Delta W_j\right) \end{aligned}$$

using the laws of exponents to turn the exponential of a sum into a product of exponentials. Thus, taking the expectation and using the completing the squares Lemma 4.11,

$$\mathrm{E}[X_{i+1}] = x_0 \prod_{j=0}^{i} \exp\left(\mu_0 \Delta t_j\right) \ . \tag{4.35}$$

This result should be compared to the corresponding deterministic multiplicative recursion or geometric progression with constant r,

$$x_{i+1} = r x_i = x_0 r^{i+1} \ ,$$

so the corresponding additive parameter to the mean in (4.32)

$$\mathrm{E}[X_{i+1}] = x_0 \left(\overline{\xi_i}^{(gm)}\right)^{i+1} \ ,$$

where

$$\overline{\xi}_i^{(gm)} = \left(\prod_{j=0}^{i} e^{\mu_0 \Delta t_j} \right)^{\frac{1}{i+1}}$$

is the **geometric mean** of the first $(i+1)$ growth steps $\xi = e^{\mu_0 \Delta t_j}$ for $j = 0 : i$.

Applications include stochastic population growth where $X(t)$ is the population size, $\mu(t)$ is an intrinsic growth rate (rate of growth in absence of stochastic or other effects in the environment) and the $\sigma(t)X(t)dW(t)$ denote the stochastic effects. The term $\sigma(t)X(t)dW(t)$ is called **demographic stochasticity** [271], since it looks like a stochastic perturbation from $\mu(t)$. Similarly, perturbations of nonlinear saturation terms are called **environmental stochasticity**. In biology, multiplicative or geometric noise is also called **density independent noise**, since $dX(t)/X(t)$ is independent of $X(t)$. See also Chapter 11 on **Biological Applications**.

Another application is financial engineering, where $X(t)$ is the investment return, $\mu(t)$ is the mean appreciation rate and $\sigma(t)$ is the investment volatility. In stochastic finance, the process $X(t)$ is called **geometric Brownian motion (GBM)** due the linear scaling on the right hand side for the $dX(t)$ and, in particular, due to the stochastic noise being multiplied by the state process $X(t)$, i.e., the multiplicative noise. In finance, one of the earliest stochastic stock models was from the thesis of Bachelier [16], in which additive noise was used, but this work did not attract much attention until Black-Scholes [34], Merton [203] and others began using multiplicative noise stock and options models. Multiplicative models are more appropriate in finance as well as in biology, since random effects are more likely to compound rather than add. See also Chapter 10 on **Financial Engineering Applications**.

For the constant coefficient case of the linear stochastic diffusion SDE, the solution can be shown to have a log-normal distribution.

Theorem 4.12. ***Solution of the Constant Coefficient, Linear Stochastic Diffusion SDE is Log-Normally Distributed:***
Let $X(t)$ satisfy

$$dX(t) = X(t) \left(\mu_0 dt + \sigma_0 dW(t) \right) , \tag{4.36}$$

$X(0) = x_0 > 0$ w.p.o., where μ_0 and $\sigma_0 > 0$ are constants. Then, the distribution of $X(t)$,

$$\Phi_{X(t)}(x) = \Phi_n \left(\ln(x); \ln(x_0)\mu_n(t), (\sigma_n)^2(t) \right) , \tag{4.37}$$

where Φ_n is the general normal distribution defined in (B.18),

$$\mu_n(t) = \ln(x_0) + \left(\mu_0 - 0.5\sigma_0^2 \right) t$$

and

$$(\sigma_n)^2(t) = \sigma_0^2 t \, .$$

Proof. Using the **probability inversion** Lemma B.19, the distribution for the solution $X(t)$ in (4.30) can be derived by reducing the distribution for $X(t)$ to the known one for the Wiener process $W(t)$ by inverting $X(t)$ in favor of $W(t)$. It is important here that $x_0 > 0$, $\sigma_0 > 0$ and that the natural logarithm $\ln(x)$ is an increasing function to preserve the direction of an inequality.

$$\begin{aligned}
\Phi_{X(t)}(x) &= \text{Prob}[X(t) \leq x] \\
&= \text{Prob}\left[x_0 \exp\left((\mu_0 - 0.5\sigma_0^2)t + \sigma_0 W(t)\right) \leq x\right] \\
&= \text{Prob}\left[\left(\mu_0 - 0.5\sigma_0^2\right) t + \sigma_0 W(t) \leq \ln(x/x_0)\right] \\
&= \text{Prob}\left[W(t) \leq \left(\ln(x/x_0) - (\mu_0 - 0.5\sigma_0^2)t\right)/\sigma_0\right] \\
&= \Phi_{W(t)}\left(\left(\ln(x/x_0) - (\mu_0 - 0.5\sigma_0^2)t\right)/\sigma_0; 0, t\right) \\
&= \Phi_n\left(\ln(x); \ln(x_0) + \left(\mu_0 - 0.5\sigma_0^2\right) t, \sigma_0^2 t\right) \ .
\end{aligned}$$

The last step follows from the conversion identity from the standard Wiener distribution $\Phi_{W(t)}$ in (B.22) to the general normal distribution Φ_n, given for Φ_n in Exercise 9 on Page B72. Thus, the probability distribution of the solution $X(t)$ is the general **lognormal distribution** of Section B.1.6, where the exponent has the normal distribution mean

$$\mu_n(t) = \ln(x_0) + \left(\mu_0 - 0.5\sigma_0^2\right) t$$

and normal variance

$$(\sigma_n)^2(t) = \sigma_0^2 t \, ,$$

i.e., the logarithm of the solution $X(t)$ has a general normal distribution, where the lognormal moment formulas are given in the Properties B.20. □

The probability density of $X(t)$ is found using the regular calculus chain rule by differentiating the distribution to yield,

$$\phi_{X(t)}(x) = x^{-1}\phi_n\left(\ln(x); \mu_n(t), (\sigma_n)^2(t)\right) \ . \tag{4.38}$$

Although the differentiation of the $\ln(x)$ distribution argument leads to an algebraic pole in $\phi_{X(t)}(x)$, $\phi_{X(t)}(0^+) \equiv 0$, which is in fact the limit as $x \to 0^+$. The leading part of the exponentially small normal distribution term $\exp(-\ln^2(x)/(2\sigma_0^2 t))$ dominates the simple algebraic pole $1/x = \exp(-\ln(x))$ as $x \to 0^+$ with the larger logarithmic exponent in magnitude.

4.1.5 Functions of General Diffusion States and Time: $F(X(t), t)$

The derivation for the special chain rule for the linear SDE logarithm transformation suggests that a more general chain rule for $F(X(t), t)$ will be needed.

Rule 4.13. ***Chain Rule for Diffusion $F(X(t), t)$:*** *Let $Y(t) = F(X(t), t)$, such that function $F(w, t)$ is twice continuously differentiable in x and once in t. Let the $X(t)$ process satisfy the diffusion SDE:*

$$dX(t) = f(X(t), t)dt + g(X(t), t)dW(t) \, , \tag{4.39}$$

$X(0) = x_0$ with probability one, while $f(X(t),t)$ and $g(X(t),t)$ satisfy the mean square integrability conditions (2.44) with the $W(t)$ argument replaced by the $X(t)$ arguments of f and g. Then

$$\begin{aligned} dY(t) &= dF(X(t),t) \\ &\stackrel{\text{dt}}{=} \left(F_t + fF_x + \frac{1}{2}g^2 F_{xx} \right)(X(t),t)dt + (gF_x)(X(t),t)dW(t)\,, \end{aligned} \quad (4.40)$$

where wholesale arguments have been used for the coefficient functions multiplying dt and $dW(t)$, respectively.

Sketch of Proof. Formally, using the increment form of the differential,

$$\begin{aligned} dY(t) &= Y(t+dt) - Y(t) \\ &= F(X(t+dt), t+dt) - F(X(t),t) \\ &= F(X(t)+dX(t), t+dt) - F(X(t),t)\,. \end{aligned}$$

Next, mean square approximations are used with their implied precision-dt,

$$\begin{aligned} dY(t) &\stackrel{\text{dt}}{=} F_t(X(t),t)dt + F_x(X(t),t)dX(t) + \frac{1}{2}F_{xx}(X(t),t)(dX)^2(t) \\ &\stackrel{\text{dt}}{=} F_t(X(t),t)dt + F_x(X(t),t)(f(X(t),t)dt + g(X(t),t)dW(t)) \\ &\quad + \frac{1}{2}F_{xx}(X(t),t)g^2(X(t),t)dt \\ &\stackrel{\text{dt}}{=} \left(F_t(X(t),t) + (fF_x)(X(t),t) + \frac{1}{2}\left(g^2 F_{xx}\right)(X(t),t) \right) dt \\ &\quad + (gF_x)(X(t),t)dW(t) \end{aligned}$$

where the diffusion SDE (4.39) has been substituted for $dX(t)$ and its square, the latter being truncated by the basic diffusion rule $(dW)^2(t) \stackrel{\text{dt}}{=} dt$ and other rules to neglect terms zero in the mean square, such as $(dW)^3(t)$, $dtdW(t)$ and $(dt)^2$, from the useful Table 2.1. □

4.2 Poisson Jump Process Calculus Rules

The Poisson process is quite different from the continuous diffusion process, primarily because of the discontinuity property of the Poisson process and the property that multiple jumps are highly unlikely during small increments in time, Δt.

4.2.1 Jump Calculus Rule for $h(dP(t))$

Thus, the most basic rule is the *zero-one law (ZOL) for jumps* (1.36), in precision-dt compact differential form,

$$(dP)^m(t) \underset{zol}{\stackrel{dt}{=}} dP(t)\,, \quad (4.41)$$

provided the integer $m \geq 1$, the case $m = 0$ being trivial. An immediate generalization of this law is the following corollary:

Corollary 4.14. ***Zero-One Jump Law for h(dP(t)):***

$$h(dP(t)) \stackrel{dt}{\underset{zol}{=}} h(1)dP(t) + h(0)(1 - dP(t)) \,, \tag{4.42}$$

with probability one, provided the function $h(p)$ is right continuous, such that values $h(0)$ and $h(1)$ exist and are bounded.

Proof. This follows by simple substitution of the *zero-one jump law*,

$$h(dP(t)) \stackrel{dt}{\underset{zol}{=}} \left\{ \begin{array}{ll} h(1), & dP(t) = 1 \\ h(0), & dP(t) = 0 \end{array} \right\} \stackrel{dt}{\underset{zol}{=}} h(1)dP(t) + h(0)(1 - dP(t)) \,, \tag{4.43}$$

$dP(t) = 0$ or $dP(t) = 1$ with probability one to precision-dt. □

Formally, the differential $dP(t)$ can be treated as a condition to test whether or not there has been a jump. This form (4.42) of the zero-one law suggests another extension of the **jump function definitions** (B.180, B.181). For example recall in (B.187) for a jump at t_1

$$[F](X(t_1), t_1) = F(X(t_1^+), t_1^+) - F(X(t_1^-), t_1^-) \,.$$

Definition 4.15. ***Jump Function [h](dP(t)):***

$$[h](dP(t)) \stackrel{dt}{\underset{zol}{=}} h(dP(t)) - h(0) \,, \tag{4.44}$$

to precision-dt, provided $h(p)$ be right continuous, such that values $h(0)$ and $h(dP(t))$ exist and are bounded.

With this definition, version (4.42) of the **zero-one law** can immediately be written,

Corollary 4.16. ***Zero-One Jump Law for h(dP(t)) with Jump function:***

$$h(dP(t)) \stackrel{dt}{\underset{zol}{=}} h(0) + [h](dP(t)) \,, \tag{4.45}$$

in terms of the jump function $[h](dP(t))$. Alternatively, the jump function is written as

$$[h](dP(t)) \stackrel{dt}{\underset{zol}{=}} (h(1) - h(0))dP(t). \tag{4.46}$$

4.2.2 Jump Calculus Rule for $\mathcal{H}(P(t), t)$

Equations (4.45, 4.46) are a primitive differential chain rule for functions of the Poisson differential $dP(t)$ only. However, more complex rules will be needed, for

instance a chain rule for a combination of a simple Poisson jump process in $P(t)$ and a deterministic process with explicit dependence on t:

Rule 4.17. ***Chain Rule for $\mathcal{H}(P(t),t)$:***
Let $\mathcal{H}(p,t)$ be once continuously differentiable in t and right continuous in p.

$$d\mathcal{H}(P(t),t) \stackrel{dt}{=}_{zol} \mathcal{H}_t(P(t),t)dt + [\mathcal{H}](P(t),t)\,, \tag{4.47}$$

where

$$[\mathcal{H}](P(t),t) \stackrel{dt}{=}_{zol} (\mathcal{H}(P(t)+1,t) - \mathcal{H}(P(t),t))\,dP(t)\,, \tag{4.48}$$

is the corresponding jump function definition for functions of $P(t)$ and t.

Sketch of Proof. Proceeding formally with differential precision-dt, the differential definition as an increment yields,

$$\begin{aligned} d\mathcal{H}(P(t),t) &= \mathcal{H}(P(t+dt),t+dt) - \mathcal{H}(P(t),t) \\ &= \mathcal{H}(P(t)+dP(t),t+dt) - \mathcal{H}(P(t),t)\,. \end{aligned}$$

Next using the Zero-One Jump law (4.42) for $h(dP(t))$ on

$$\mathcal{H}(P(t)+dP(t),t+dt)$$

for fixed $(P(t),t)$ to take $dP(t)$ out of its first argument and then expanding the second argument dt to two terms up to $\mathcal{H}_t$,

$$\begin{aligned} d\mathcal{H}(P(t),t) &\stackrel{dt}{=}_{zol} \mathcal{H}(P(t)+1,t+dt)dP(t) \\ &\qquad + \mathcal{H}(P(t)+0,t+dt)(1-dP(t)) - \mathcal{H}(P(t),t) \\ &\stackrel{dt}{=}_{zol} (\mathcal{H}(P(t),t) + \mathcal{H}_t(P(t),t)dt)(1-dP(t)) \\ &\qquad + (\mathcal{H}(P(t)+1,t) + \mathcal{H}_t(P(t)+1,t)dt)dP(t) - \mathcal{H}(P(t),t) \\ &\stackrel{dt}{=}_{zol} \mathcal{H}_t(P(t),t)dt + (\mathcal{H}(P(t)+1,t) - \mathcal{H}(P(t),t))dP(t) \\ &\stackrel{dt}{=}_{zol} \mathcal{H}_t(P(t),t)dt + [\mathcal{H}](P(t),t))\,, \end{aligned}$$

the last line due to using the jump function definition (4.48). Also used was the bilinear differential form

$$dt\;dP(t) \stackrel{dt}{=}_{zol} 0\,,$$

that is mainly responsible for the elimination of combined continuous and jump changes.

The precision-dt jump differential Table 3.2 was used to eliminate terms smaller than precision-dt terms in the mean square sense. The dt factor $\mathcal{H}_t(p,t)$ is the partial derivative of $\mathcal{H}$ with respect to t while p is held fixed. Note that the jump function is defined for all t so that if there is no Poisson jump, then the jump function is identically zero since $dP(t)=0$, the zero jump case. □

Remarks 4.18.

- *The bilinear differential form* $dt\, dP(t) \overset{dt}{\underset{zol}{=}} 0$ *is consistent with the fact that the Poisson process has jump discontinuities and thus jumps must be instantaneous. Consequently, continuous changes and jump changes can be computed independently, since there are zero continuous changes at each jump instant.*
- *This leads to the alternate form of Rule 4.17.*

Rule 4.19. ***Alternate Chain Rule for*** $\mathcal{H}(P(t),t)$***:***
Let $\mathcal{H}(p,t)$ *be once continuously differentiable in* t *and right continuous in* p*.*

$$d\mathcal{H}(P(t),t) \overset{dt}{\underset{zol}{=}} d_{(cont)}\mathcal{H}_(P(t),t) + d_{(jump)}\mathcal{H}(P(t),t)\,, \tag{4.49}$$

where

$$d_{(cont)}\mathcal{H}_(P(t),t) \equiv \mathcal{H}_t(P(t),t)dt \tag{4.50}$$

and

$$d_{(jump)}\mathcal{H}(P(t),t) \equiv [\mathcal{H}](P(t),t)\,. \tag{4.51}$$

Example 4.20. ***Stochastic Jump Power:*** *Let* $a \neq 0$ *and* $b > 0$*. Using the stochastic jump chain rule (4.48) in differential form, then*

$$\begin{aligned} d\left[b^{aP(t)+ct}\right] &\overset{dt}{\underset{zol}{=}} c\ln(b)b^{aP(t)+ct}dt + \left(b^{a(P(t)+1)+ct} - b^{aP(t)+ct}\right)dP(t) \\ &= b^{aP(t)+ct}\left(c\ln(b)dt + (b^a-1)\,dP(t)\right)\,, \end{aligned}$$

where the calculus rule, $d(b^{ct}) = d(e^{c\ln(b)t}) = c\ln(b)b^{ct}$*, for an arbitrary positive power base* b *with an exponential rule has been used.*

The corresponding jump integral derived from this formula is

$$\int_0^t b^{aP(s)+cs}dP(s) \overset{\mathrm{dt}}{=} \frac{1}{b^a-1}\left(\left(b^{aP(t)+ct}-1\right) - c\ln(b)\int_0^t b^{aP(s)+cs}ds\right)\,,$$

provided $b^a \neq 1$*. This integral formula simplifies if* $b = e$ *and* $c = 0$ *to*

$$\int_0^t \exp(aP(s))dP(s) \overset{\mathrm{dt}}{=} (\exp(aP(t)) - 1)/(\exp(a) - 1)\,,$$

but still different from the deterministic version,

$$\int_0^t \exp(as)ds = (\exp(at) - 1)/a.$$

4.2.3 Jump Calculus Rule with General State $Y(t) = F(X(t), t)$:

The chain rule for $F(P(t), t)$ is still too simple, so a chain rule for more general jump processes $X(t)$, such as for $F(X(t), t)$, is needed. First, a definition of a jump function for general transformations is needed.

Definition 4.21. $[Y](t)$ *for General* $Y(t) = F(X(t), t)$*:*
Let the process $Y(t) = F(X(t), t)$ *be a continuous transformation of the process* $X(t)$ *with jump function* $[X](t)$ *at* t*. Then the jump function in* $Y(t)$ *is defined as*

$$[Y](t) = [F](X(t), t) = F(X(t) + [X](t), t) - F(X(t), t)\,. \tag{4.52}$$

Lemma 4.22. $[Y](t)$ *for* $Y(t) = F(X(t), t)$ *with* $[X](t) = h(X(t), t)dP(t)$*:*
Let the process $Y(t) = F(X(t), t)$ *be a continuous transformation of the process* $X(t)$ *with jump function*

$$[X](t)h(X(t), t)dP(t)$$

at t*, then*

$$[Y](t) = [F](X(t), t) = \left(F(X(t) + h(X(t), t), t) - F(X(t), t)\right) dP(t)\,. \tag{4.53}$$

Proof. This follows from the Zero-One-Jump Law (4.42) for $h(dP(t))$ upon substitution of the jump of $[X](t) = h(X(t), t)dP(t)$ into the definition (4.52), so that

$$\begin{aligned}
[Y](t) &\equiv F(X(t) + [X](t), t) - F(X(t), t) \\
&= F(X(t) + h(X(t), t)dP(t), t) - F(X(t), t) \\
&= \left(F(X(t) + h(X(t), t), t) - F(X(t), t)\right) dP(t).
\end{aligned}$$

□

Rule 4.23. ***Chain Rule for Jump in*** $Y(t) = F(X(t), t)$*:*
Let $Y(t) = F(X(t), t)$*, such that the function* $F(x, t)$ *is once continuously differentiable in* x *and once in* t*. Let the* $X(t)$ *process satisfy the jump SDE:*

$$dX(t) = f(X(t), t)dt + h(X(t), t)dP(t)\,, \tag{4.54}$$

$X(0) = x_0$ *with probability one, while* $f(X(t), t)$ *and* $h(X(t), t)$ *satisfy the mean square integrability conditions (2.44) with the* $W(t)$ *argument replaced by the* $X(t)$ *arguments of* f *and* h*. In (4.54), the jump in* $X(t)$ *is* $[X](T_k^-) \equiv X(T_k^+) - X(T_k^-) = h(X(T_k^-), T_k^-)$ *for each* k*th jump-time* T_k *of* $P(t)$*. Then*

$$\begin{aligned}
dY(t) &= dF(X(t), t) \\
&\overset{dt}{\underset{zol}{=}} \left(F_t + fF_x\right)(X(t), t)dt + [F](X(t), t)\,,
\end{aligned} \tag{4.55}$$

where wholesale arguments have been used for the coefficient functions multiplying dt and dP(t), respectively, and where the jump in Y(t) = F(X(t), t) is given in (4.53) of Lemma 4.22.

Sketch of Proof. Formally, using the increment form of the differential,

$$\begin{aligned} dY(t) &= Y(t+dt) - Y(t) \\ &= F(X(t+dt), t+dt) - F(X(t), t) \\ &= F(X(t) + dX(t), t+dt) - F(X(t), t) \, . \end{aligned}$$

Next, as for (4.47,4.49) of the two prior rules, the instantaneous jump changes (terms in $dP(t)$ only, such that $[X](t) = h(X(t), t)dP(t)$) are treated separately from the continuous and smooth deterministic changes (terms in dt only, such that $dX^{((det))}(t) = f(X(t), t)dt$), then the mean square approximations are used with their implied precision-dt,

$$\begin{aligned} dY(t) &\underset{zol}{\overset{dt}{=}} F_t(X(t), t)dt + F_x(X(t), t)f(X(t), t)dt \\ &\qquad + (F(X(t) + [X](t), t) - F(X(t), t)) \\ &\underset{zol}{\overset{dt}{=}} (F_t + fF_x)(X(t), t)dt + (F(X(t) + h(X(t), t)dP(t), t) - F(X(t), t)) \\ &\underset{zol}{\overset{dt}{=}} (F_t + fF_x)(X(t), t)dt + (F(X(t) + h(X(t), t), t) - F(X(t), t))\, dP(t) \, , \end{aligned}$$

where the zero-one jump law (4.46) has been used to take the $dP(t)$ out of the argument of F and let it multiply the jump change in F in the last line of the above equation. Note that the jump change has been defined, so that if there is no Poisson jump, then the jump function is zero. □

4.2.4 Transformations of Linear Jump with Drift SDEs

Consider the jump SDE, linear in the state process $X(t)$, with time-dependent coefficients,

$$dX(t) = X(t)\left(\mu(t)dt + \nu(t)dP(t)\right) \, , \tag{4.56}$$

where here the initial condition is $X(t_0) = x_0 > 0$ with probability one, $\mu(t)$ is called the **drift** or deterministic coefficient and $\nu(t)$ is called the **jump amplitude** coefficient of the Poisson jump term. The jump in state is $[X](T_k) = \nu(T_k^-)$ for each jump of $P(t)$, i.e., $[P](T_k) = 1$ for each k. Assume that the rate coefficients, $\mu(t)$ and $\nu(t)$ are bounded, while $\nu(t) > -1$. In the deterministic and the linear diffusion cases, transforming the state variable to its logarithm makes the right hand side independent of the transformed state variable, so let

$$Y(t) = F(X(t)) \equiv \ln(X(t)) \, .$$

The most recent jump chain rule (4.55, 4.53) is applicable in this case with

$$f(X(t), t) = X(t)\mu(t)$$

and

$$h(X(t),t) = X(t)\nu(t),$$

although the increment form of $dF(X(t))$ can be directly expanded to get the same result. Since only the first partial derivative and the jump function of F are needed, while F does not depend on t, then

$$F_x(X(t)) = 1/X(t)\,, \quad F_t(X(t)) \equiv 0\,,$$

and from (4.53)

$$[F](X(t)) \underset{zol}{\overset{dt}{=}} (\ln(X(t) + X(t)\nu(t)) - \ln(X(t)))\, dP(t) = \ln(1+\nu(t))dP(t)\,, \quad (4.57)$$

where the logarithm subtraction rule $\ln(A) - \ln(B) = \ln(A/B)$, provided $A > 0$ and $B > 0$, has been used to cancel out the linear state dependence in the jump term. Note that the jump amplitude becomes singular as $\nu(t) \to (-1)^+$, approaching a massive disaster to the state. Thus,

$$\begin{aligned} dY(t) &= dF(X(t)) = F_x(X(t))X(t)\mu(t)dt + [F](X(t)) \\ &\underset{zol}{\overset{dt}{=}} \mu(t)dt + \ln(1+\nu(t))dP(t)\,. \end{aligned} \quad (4.58)$$

The infinitesimal mean of $Y(t)$, assuming the jump rate is time dependent $\mathrm{E}[dP(t)] = \lambda(t)dt$ too, is

$$\mathrm{E}[dY(t)] = (\mu(t) + \lambda(t)\ln(1+\nu(t)))\, dt \quad (4.59)$$

and the **infinitesimal variance** is

$$\mathrm{Var}[dY(t)] \overset{\mathrm{dt}}{=} \lambda(t)\ln^2(1+\nu(t))dt\,, \quad (4.60)$$

noting that the jump amplitude has a power effect between the infinitesimal expectation and the variance unlike the Poisson infinitesimal property that $\mathrm{Var}[dP(t)] = \mathrm{E}[dP(t)]$.

Since the final right hand side of (4.58) does not depend on the state $Y(t)$, we can easily integrate for $Y(t)$ explicitly leading to

$$Y(t) = Y(t_0) + \int_{t_0}^{t} \mu(s)ds + \int_{t_0}^{t} \ln(1+\nu(s))dP(s). \quad (4.61)$$

Exponentiation leads to the formal solution for the original state,

$$X(t) = X(t_0)\exp\left(\int_{t_0}^{t} \mu(s)ds + \int_{t_0}^{t} \ln(1+\nu(s))dP(s)\right)\,. \quad (4.62)$$

Linear Jump SDEs with Constant Coefficients:

If the SDE has **constant coefficients**, $\mu(t) = \mu_0$, $\nu(t) = \nu_0$ and $\lambda(t) = \lambda_0$, then the solution is simpler,

$$\begin{aligned} X(t) &\overset{ims}{=} X(t_0)\exp\left(\mu_0(t-t_0) + \ln(1+\nu_0)(P(t) - P(t_0))\right) \\ &= X(t_0)\exp\left(\mu_0(t-t_0)\right)(1+\nu_0)^{(P(t)-P(t_0))}\,, \end{aligned} \quad (4.63)$$

where, in the last line, the exponential-logarithm inverse relation, $\exp(a\ln(b)) = b^a$, has been used to move the Poisson term out of the exponential.

In this pure jump with drift process, the moments are computed using the Poisson distribution (1.21) coupled with the stationary property that the distribution depends only on the increment,

$$\begin{aligned}\text{Prob}[P(t) - P(t_0) = k] &= \text{Prob}[P(t - t_0) = k] = p_k(\lambda_0(t - t_0)) \\ &= e^{-\lambda_0(t-t_0)} \frac{(\lambda_0(t - t_0))^k}{k!} \, .\end{aligned}$$

Thus, the calculation of the mean of the process in (4.63) is

$$\begin{aligned}\text{E}[X(t)] &= x_0 e^{\mu_0(t-t_0)} e^{-\lambda_0(t-t_0)} \sum_{k=0} \frac{(\lambda_0(t - t_0))^k}{k!} (1 + \nu_0)^k \\ &= x_0 e^{\mu_0(t-t_0) - \lambda_0(t-t_0)} e^{\lambda_0(t-t_0)(1+\nu_0)} \\ &= x_0 e^{\mu_0 + \lambda_0\nu_0)(t-t_0)} \, ,\end{aligned}$$

growing in time if $\mu_0 + \lambda_0\nu_0 > 0$, but decaying if $\mu_0 + \lambda_0\nu_0 < 0$. Note that $\lambda_0 > 0$, but both μ_0 and ν_0 can be of any sign. The corresponding calculation of the variance of $X(t)$ is

$$\begin{aligned}\text{Var}[X(t)] &= \text{E}[X^2(t)] - \text{E}^2[X(t)] \\ &= x_0^2 e^{2\mu_0(t-t_0)} e^{-\lambda_0(t-t_0)} \sum_{k=0} \frac{(\lambda_0(t - t_0))^k}{k!} (1 + \nu_0)^{2k} - \text{E}^2[X(t)] \\ &= x_0^2 e^{2\mu_0(t-t_0) - \lambda_0(t-t_0)} e^{\lambda_0(t-t_0)(1+\nu_0)^2} - x_0^2 e^{2(\mu_0 + \lambda_0\nu_0)(t-t_0)} \\ &= x_0^2 e^{2(\mu_0 + \lambda_0\nu_0)(t-t_0)} \left(e^{\lambda_0\nu_0^2)(t-t_0)} - 1 \right) \\ &= \text{E}^2[X(t)] \left(e^{\lambda_0\nu_0^2)(t-t_0)} - 1 \right) \, ,\end{aligned}$$

so the growth or decay is proportional to the mean squared, but amplified asymptotically by the growing term $exp(\lambda_0\nu_0^2)(t - t_0))$. For the distribution, see the Subsection 4.3.3 for the linear jump-diffusion SDE case.

Applications include stochastic population growth where $X(t)$ is the population size, such that the population grows exponentially at intrinsic growth rate $\mu(t)$ in absence of stochastic disasters, but suffers from a random linear disaster if the jump amplitude rate $-1 < \nu(t) < 0$ or random linear bonanza if $\nu(t) > 0$. See also Ryan and Hanson [241] or Chapter 11 on **Biological Applications**.

4.3 Jump-Diffusion Rules and SDEs

Wiener diffusion and simple Poisson jump processes provide an introduction to elementary stochastic differential equation (SDE) in continuous time for the **simple jump-diffusion** state process $X(t)$,

$$dX(t) = f(X(t), t)dt + g(X(t), t)dW(t) + h(X(t), t)dP(t) \, , \qquad (4.64)$$

where $X(0) = x_0$, with a set of continuous coefficient functions $\{f, g, h\}$, possibly nonlinear in the state $X(t)$. However, in the process of introducing the component Markov processes, too many rules have been accumulated and in this section most of these rules will be combined into one rule or a few rules.

4.3.1 Jump-Diffusion Conditional Infinitesimal Moments

The conditional infinitesimal moments for the state process are useful for application modeling and are given by

$$\mathrm{E}[dX(t)|X(t) = x] = (f(x,t) + \lambda(t)h(x,t))dt \,, \tag{4.65}$$

and

$$\mathrm{Var}[dX(t)|X(t) = x] = \left(g^2(x,t) + \lambda(t)h^2(x,t)\right) dt \,, \tag{4.66}$$

using (1.1,1.2,4.64), and assuming that the Poisson process is independent of the Wiener process.

The jump in the state at jumps T_k in the Poisson process, i.e., $[P](T_k) = 1$, is not an infinitesimal moment but serves as a simple property of the SDE and is given by

$$[X](T_k) \equiv X(T_k^+) - X(T_k^-) = h(X(T_k^-), T_k^-) \tag{4.67}$$

or

$$[X](t) = h(X(t), t)dP(t) \,, \tag{4.68}$$

under the assumption that the jumps are instantaneous so there is no time continuous changes and that in the interval $(t, t+dt]$ there is only time for one jump, if any, of the Poisson term by the zero-one jump rule (1.35). Note that no $dP(t)$ appears in (4.67) since a jump is assumed at $t = T_k$. The jump amplitude evaluation (4.67) at the pre-jump time value T_k^- follows from the Itô forward integration approximation and the right continuity of $P(t)$, as discussed in the previous chapter and also means that the jump amplitude depends only at the immediate pre-jump value of h, but not on the post-jump value which in a sense is in the future.

The infinitesimal moment and jump properties are very useful for modeling approximations of real applications, by providing a basis for estimating the coefficient functions f, g, and h, as well as some of the process parameters, at least in the first approximation, through comparison to the empirical values of the basic probability corresponding of the stochastic integral equation.

4.3.2 Stochastic Jump-Diffusion Chain Rule

The corresponding stochastic chain rule for calculating the differential of a composite process $F(X(t), t)$ begins by interpreting the differential as an infinitesimal increment and recognizing that since the Poisson jumps are instantaneous there is no time for continuous changes. Thus, a critical concept in deriving the chain rule is that the continuous changes and jump changes can be calculated independently.

The state process is decomposed into continuous changes,

$$d_{(cont)}X(t) = f(X(t), t)dt + g(X(t), t)dW(t) \tag{4.69}$$

and discontinuous or jump changes,

$$d_{(jump)}X(t) = [X](t) = h(X(t),t)dP(t) \tag{4.70}$$

such that

$$dX(t) = d_{(cont)}X(t) + d_{(jump)}X(t)\,. \tag{4.71}$$

Another critical concept is the transformation of the conditioning for the jump. The differential Poisson $dP(t)$ serves as the conditioning for the existence of a jump. This jump conditioning follows from the probability distribution for the differential Poisson process (1.23) which behaves asymptotically for small λdt as the zero-one jump law,

$$\Phi_{dP(t)}(k;\lambda dt) = \text{Prob}[dP(t)=k] = \left\{ \begin{array}{ll} 1-\lambda dt, & k=0 \\ \lambda dt, & k=1 \\ 0, & k>1 \end{array} \right\} + \text{O}^2(\lambda dt)\,, \tag{4.72}$$

so that $dP(t)$ behaves as an **indicator function of the jump counter** k with neglected error $\text{O}^2(dt) = \text{o}(dt)$, i.e., $dP(t) = 0$ with asymptotic probability $(1-\lambda dt)$ if there is no jump and $dP(t) = 1$ with asymptotic probability (λdt) if there is a jump, while multiple jumps are likely to be negligible.

Thus, the change of a composite function of the state process $X(t)$, $dF(X(t),t)$, can be decomposed into the sum of continuous and discontinuous changes.

The function $F(x,t)$ is assumed to be at least twice continuously differentiable in x and once in t. Due to the non-smoothness (1.6), a two term Taylor approximation from continuous calculus yields, with subscripts denoting partial derivatives, the continuous change,

$$\begin{aligned} d_{(cont)}F(X(t),t) \simeq\; & F_t(X(t),t)dt + F_x(X(t),t)d_{(cont)}X(t) \\ & + \frac{1}{2}F_{xx}(X(t),t)(d_{(cont)}X(t))^2\,, \end{aligned}$$

which would be the chain rule for the compound function $F(X(t),t)$ of a deterministic function $X(t)$ with the non-smooth property in (1.6). The discontinuous change follows from the transformation of the jump in $X(t)$ at time t given in the previous section to the jump in the composite function $Y(t) = F(X(t),t)$,

$$d_{(jump)}F(X(t),t) = \left(F(X(t)+h(X(t),t),t) - F(X(t),t)\right)dP(t)\,,$$

using the jump

$$[X](t) = h(X(t),t),t)dP(t)$$

and the continuity of F in t, such that when there is a jump at time T_k in $dP(t)$, the jump in F is evaluated at the pre-jump time T_k^-, else the discontinuous contribution is zero. Combining the continuous and discontinuous process changes while keeping at terms non-zero in the mean square limit sense yields

$$\begin{aligned} dF(X(t),t) &= F(X(t)+dX(t),t+dt) - F(X(t),t) \\ &= F_t(X(t),t)dt + F_x(X(t),t)\cdot\left(f(X(t),t)dt + g(X(t),t)dW(t)\right) \\ &\quad + \frac{1}{2}F_{xx}(X(t),t)\cdot g^2(X(t),t)dt \\ &\quad + (F(X(t)+h(X(t),t),t) - F(X(t),t))dP(t)\,. \end{aligned} \tag{4.73}$$

Rewriting (4.73) slightly leads to the final statement Itô stochastic chain rule for jump-diffusions with simple Poisson jumps:

Rule 4.24. *Let $F(x,t)$ be twice continuously differentiable in x and once in t.*

$$dF(X(t),t) = = \left(F_t + fF_x + \frac{1}{2}g^2F_{xx}\right)(X(t),t)dt + (gF_x)(X(t),t)dW(t) + (F(X(t)+h(X(t),t),t) - F(X(t),t))\,dP(t)\,. \tag{4.74}$$

Here, to summarize, it is assumed that the Wiener process is **independent** of the Poisson processes and that the quadratic differential Wiener process $(dW)^2(t)$ can be replaced by its mean square limiting value which is dt within **precision-**dt. Thus, the part of the $\mathrm{O}(dt)$ change dF due to the Wiener process requires a second derivative beyond the regular calculus first derivative Taylor approximation and thus the non-smooth Wiener property plays a strong role compared to its stochastic or random property. The second derivative term is a diffusion term and hence the Wiener process is called a diffusion process. However, the motivations for stochastic diffusions and physical diffusions are quite different, but they both lead to diffusion equations. The jump term uses the **zero-one jump indicator property** of $dP(t)$, so

$$\begin{aligned}[F](X(t),t) &= F(X(t)+[X](t),t) - F(X(t),t)\\ &= F(X(t)+h(X(t),t)dP(t),t) - F(X(t),t)\\ &= (F(X(t)+h(X(t),t),t) - F(X(t),t))\,dP(t)\,,\end{aligned}$$

to pass the jump differential $dP(t)$ from the state argument of $F(x,t)$ to a multiplying factor of the potential jump difference $F(x+h(x,t),t) - F(x,t)$. If there is a jump at $t = T_k$ then $dP(t)$ produces a change in the arguments $(X(t),t)$ of both F and h to $(x,t) = (X(T_k^-),T_k^-)$. However, if the F and h are continuous in the explicit t-arguments, then $(x,t) = (X(T_k^-),T_k)$ can be used.

Remark 4.25. *Several authors use artificial arguments like $(X(t^-),t^-)$ when treating Markov jump process SDEs or their corresponding integral equations due to using an incomplete Poisson or related process model.*

4.3.3 Linear Jump-Diffusion SDEs

Let the linear diffusion and jump SDEs be combined into a single SDE:

$$dX(t) = X(t)\,(\mu(t)dt + \sigma(t)dW(t) + \nu(t)dP(t))\,, \tag{4.75}$$

$X(t_0) = x_0 > 0$ with probability one (this is for specificity, but only $x_0 \neq 0$ is sufficient), where the set of coefficients $\{\mu(t),\sigma(t),\nu(t),\lambda(t)\}$ are assumed to be bounded and integrable, with $\nu(t) > -1$ (otherwise, positivity of $X(t)$ cannot be maintained) and $\sigma(t) > 0$ (for consistency with the interpretation as a standard deviation coefficient of the process). The logarithmic transformation of the state

process $Y(t) = \ln(X(t))$ transforms away the state from the right hand side of the SDE using the jump-diffusion chain rule (4.74) and the first two logarithmic derivatives, so

$$dY(t) = (\mu(t) - \sigma^2(t)/2)dt + \sigma(t)dW(t) + \ln(1+\nu(t))dP(t)\,. \tag{4.76}$$

SDE (4.76) is a linear combination of the deterministic, diffusion and jump processes with deterministic time dependent coefficients so can be immediately but formally integrated to yield,

$$Y(t) = y_0 + \int_{t_0}^{t} \left((\mu(s) - \sigma^2(s)/2)ds + \sigma(s)dW(s) + \ln(1+\nu(s))dP(s)\right)\,, \tag{4.77}$$

where $y_0 = \ln(x_0)$, recalling that it has been assumed that $x_0 > 0$. Inverting logarithmic state $Y(t)$ back to the original state

$$X(t) = \exp(Y(t))$$

leads to

$$X(t) = x_0 \exp\left(\int_{t_0}^{t} \left((\mu(s) - \sigma^2(s)/2)ds + \sigma(s)dW(s) + \ln(1+\nu(s))dP(s)\right)\right). \tag{4.78}$$

Linear Jump-Diffusion SDEs with Constant Coefficients:

For the special case of constant rate coefficients, $\mu(t) = \mu_0$, $\sigma(t) = \sigma_0$, $\nu(t) = \nu_0$ and $\lambda(t) = \lambda_0$, also setting $t_0 = 0$, leads to the SDE:

$$dX(t) = X(t)\left(\mu_0 dt + \sigma_0 dW(t) + \nu_0 dP(t)\right)\,, \tag{4.79}$$

$X(t_0) = x_0 > 0$ with probability one with solution:

$$\begin{aligned} X(t) &= x_0 \exp\left((\mu_0 - \sigma_0^2/2)t + \sigma_0 W(t) + \ln(1+\nu_0)P(t)\right) \\ &= x_0(1+\nu_0)^{P(t)} \exp\left((\mu_0 - \sigma_0^2/2)t + \sigma_0 W(t)\right)\,, \end{aligned} \tag{4.80}$$

applying the logarithm-exponential inverse property.

Using the density $\phi_{W(t)}(w)$ for the diffusion $W(t)$ in (1.7) and the discrete distribution $\Phi_{P(t)}(k) = p_k(\lambda_0 t)$ for the jump process $P(t)$, together with the pairwise independence of the two processes, then the state expectation can be found directly as

$$\begin{aligned} E[X(t)] &= x_0 e^{(\mu_0 - \sigma_0^2/2)t} e^{-\lambda_0 t} \sum_{k=0}^{\infty} \frac{(\lambda_0 t)^k}{k!}(1+\nu_0)^k \\ &\qquad \cdot \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} e^{-w^2/(2t)} e^{\sigma_0 w} dw \\ &= x_0 e^{(\mu_0 - \sigma_0^2/2)t} e^{-\lambda_0 t} e^{\lambda_0 t(1+\nu_0)} e^{\sigma_0^2 t/2} \\ &= x_0 e^{(\mu_0 + \lambda_0\nu_0)t}\,, \end{aligned} \tag{4.81}$$

where the exponential series and **completing the square** technique have been used. It is interesting to note that the conditional infinitesimal expectation relative to the $X(t)$ for this constant coefficient case is

$$E[dX(t)|X(t)]/X(t) = (\mu_0 + \lambda_0\nu_0)dt\,,$$

provided that the given condition value $X(t) \neq 0$, which means that if the above infinitesimal expected result is interpreted implying the expected rate then the state expectation in (4.81) is the same result as for the equivalent deterministic process. Note the above equation is equivalent to $E[dX(t)|X(t) = x]/x = \mu_0 + \lambda_0\nu_0)dt$ with $x \neq 0$, but it is unnecessary to introduce the extra realized value x for $X(t)$ and later it will be seen that this extra introduction would be awkward in nested conditional expectations for stochastic dynamic programming in Chapter 6. For more on this **quasi-deterministic equivalence** for linear stochastic processes, see Hanson and Ryan [114].

Using similar applications of the same techniques, the state variance is computed to be

$$\begin{aligned}
\mathrm{Var}[X(t)] &= \mathrm{E}\left[(X(t) - \mathrm{E}[X(t)])^2\right] = \mathrm{E}\left[X^2(t)\right] - \mathrm{E}^2[X(t)] \\
&= x_0^2 e^{2(\mu_0 - \sigma_0^2/2)t}\left(\mathrm{E}\left[e^{2\sigma_0 W(t)}(1+\nu_0)^{2P(t)}\right]\right. \\
&\quad \left. -\mathrm{E}^2\left[e^{\sigma_0 W(t)}(1+\nu_0)^{P(t)}\right]\right) \\
&= x_0^2 e^{2(\mu_0 - \sigma_0^2/2)t}\left(e^{2\sigma_0^2 t}e^{\lambda_0 t((1+\nu_0)^2-1)} - e^{\sigma_0^2 t}e^{2\lambda_0\nu_0 t}\right) \\
&= x_0^2 e^{2(\mu_0+\lambda_0\nu_0)t}\left(e^{(\sigma_0^2+\lambda_0\nu_0^2)t} - 1\right) \\
&= \mathrm{E}^2[X(t)]\left(e^{(\sigma_0^2+\lambda_0\nu_0^2)t} - 1\right)\,. \qquad (4.82)
\end{aligned}$$

The conditional infinitesimal variance relative to the square of the state, in this constant coefficient case, is

$$\mathrm{Var}[dX(t)|X(t)]/X^2(t) = (\sigma_0^2 + \lambda_0\nu_0^2)dt\,,$$

provided $X(t) \neq 0$, which in turn is the time integral of the exponent, $(\sigma_0^2 + \lambda_0\nu_0^2)t$, in the last line of (4.82) and since this exponent must be positive ($\lambda_0 > 0$) ensuring exponential amplification in time relative to the expectation exponential with exponent $((\mu_0 + \lambda_0\nu_0)t)$, which could be of any sign. The usual measure of the relative changes of a random variable is called the **coefficient of variation**, which here is

$$\mathrm{CV}[X(t)] \equiv \frac{\sqrt{\mathrm{Var}[X(t)]}}{\mathrm{E}[X(t)]} = \sqrt{e^{(\sigma_0^2+\lambda_0\nu_0^2)t} - 1}\,, \qquad (4.83)$$

provided $X(t) \neq 0$, which grows exponentially with time t. The $\mathrm{CV}[X(t)]$ is often used in the sciences to represent results, due to its dimensionless form. The dimensionless form makes it easier to pick out general trends or properties, especially if the $\mathrm{CV}[X(t)]$ distills down to something very simple.

The probability density for the solution $X(t)$ in (4.80) in the case of the constant coefficient, linear jump-diffusion SDE can be found by application of the **law of total probability** (B.92) and the probability inversion principle in Lemma B.19. Thus, assuming $x_0 > 0$ and $\sigma_0 > 0$,

$$\begin{aligned}
\Phi_{X(t)}(x) &\equiv \text{Prob}[X(t) \le x] \\
&= \sum_{k=0}^{\infty} \text{Prob}\left[x_0 e^{(\mu_0 - 0.5\sigma_0^2)t + \sigma_0 W(t)} (1+\nu_0)^{P(t)} \le x \middle| P(t) = k \right] \\
&\qquad \cdot \text{Prob}\left[P(t) = k\right] \\
&= \sum_{k=0}^{\infty} p_k\left(\lambda_0 t\right) \text{Prob}\left[x_0 e^{(\mu_0 - 0.5\sigma_0^2)t + \sigma_0 W(t)} (1+\nu_0)^{k} \le x \right] \\
&= \sum_{k=0}^{\infty} p_k\left(\lambda_0 t\right) \text{Prob}\left[W(t) \le \left(\ln(x/x_0) - (\mu_0 - 0.5\sigma_0^2)t - k\ln(1+\nu_0) \right)/\sigma_0 \right] \\
&= \sum_{k=0}^{\infty} p_k\left(\lambda_0 t\right) \Phi_{W(t)}\left(\left(\ln(x/x_0) - (\mu_0 - 0.5\sigma_0^2)t - k\ln(1+\nu_0) \right)/\sigma_0 \right) \\
&= \sum_{k=0}^{\infty} p_k\left(\lambda_0 t\right) \Phi_{n}\left(\left(\ln(x/x_0) - (\mu_0 - 0.5\sigma_0^2)t - k\ln(1+\nu_0) \right)/\sigma_0; 0, t \right) \\
&= \sum_{k=0}^{\infty} p_k\left(\lambda_0 t\right) \Phi_{n}\left(\ln(x); \ln(x_0) + (\mu_0 - 0.5\sigma_0^2)t + k\ln(1+\nu_0), \sigma_0^2 t \right) \ ,
\end{aligned}$$

where $\Phi_{W(t)}$ is the distribution of $W(t)$ (B.22) given in terms of the normal distribution Φ_n (B.18). The last step again follows from the conversion identity from standard to general normal distribution, given in Exercise 9 on Page B72. Thus, we have just proven the following jump-diffusion probability distribution theorem for the linear constant coefficient SDE by elementary probability principles.

Theorem 4.26. ***Jump-Diffusion Probability Distribution for Linear Constant-Coefficient SDE:***
Let $X(t)$ formally satisfy the scalar, linear, constant coefficient SDE (4.79) with initial condition $X(0) = x_0 > 0$, Then for each value of the jump counter k, the distribution is a sequence of distributions,

$$\Phi_{X(t)}(x) = \sum_{k=0}^{\infty} p_k(\lambda_0 t) \Phi_{X(t)}^{(k)}(x) \ ,$$

where each term of the sequence has the form,

$$\Phi_{X(t)}^{(k)}(x) = \Phi_n\left(\ln(x); \mu_n^{(k)}(t), \sigma_0^2 t \right) \ ,$$

i.e., is a lognormal distribution (B.30) with normal mean

$$\mu_n^{(k)}(t) = \ln(x_0) + \left(\mu_0 - 0.5\sigma_0^2\right) t + k\ln(1+\nu_0)$$

and normal variance

$$(\sigma_n)^2(t) = \sigma_0^2 t\,.$$

For each k the logarithm of the solution $X(t)$ has a general normal distribution, where the lognormal moment formulas are given in the Properties B.20. The probability density of $X(t)$ is found by chain rule differentiating the distribution to yield,

$$\phi_{X(t)}(x) = \sum_{k=0}^{\infty} p_k\,(\lambda_0 t)\,x^{-1}\phi_n\left(\ln(x); \mu_n^{(k)}(t), (\sigma_n)^2(t)\right)\,, \tag{4.84}$$

for $x > 0$, such that $\phi_{X(t)}(0) \equiv \phi_{X(t)}(0^+) = 0$.

Remarks 4.27.

- *The fact $\phi_{X(t)}(0) \equiv \phi_{X(t)}(0^+) = 0$ is because for the limit as $x \to 0^+$, the exponentially small normal distribution term dominates the simple algebraic pole from $1/x$.*
- *For each k, the normal mean is shifted by an amount $\ln(1+\nu_0)$ and is weighted by the Poisson jump counting probability $p_k(\lambda_0 t) = exp(-\lambda_0 t)(\lambda_0 t)^k/k!$ so the contributions decay like those of the exponential series.*

Solution Simulations for Linear Jump-Diffusion SDEs with Constant Coefficients:

Upon merging and modifying the simulation algorithms for small time increments in Fig. 1.1 using the cumulative sum of normal random generated Wiener increment approximations together with the cumulative sum of uniform random generated Poisson increment approximations with acceptance-rejection technique [230, 96] to model the zero-one jump law, simulations of the linear jump diffusion process with constant parameters solution (4.80) are shown in Fig. 4.3. The basic simulation is performed on the approximate exponent increment

$$\Delta Y_i \simeq (\mu_0 - \sigma_0^2/2)\Delta t + \sigma_0 \Delta W_i + \ln(1+\nu_0)\Delta P_i, \tag{4.85}$$

corresponding to SDE (4.85), where $\Delta t = 0.001$ for this MATLAB generated figure,

$$\Delta W_i \simeq DW(i)\,, \Delta t = Dt\,, \quad \text{where} \quad DW = \texttt{sqrt}(Dt) * \texttt{randn}(N,1)\,;$$

and

$$\Delta P_i \simeq DP(i) = U(DU(i); ul, ur)\,, \quad \text{where} \quad DU = \texttt{rand}(N,1)\,,$$

for $i = 1:N$ with $X(0) = x_0$, $U(u; ul, ur)$ is the unit step function on the centered interval $[ul = (1-\lambda_0\Delta t)/2, ur = (1+\lambda_0\Delta t)/2]$, approximating the zero-one jump law through the acceptance-rejection method [230, 96]. The state exponent, YS, sample path, starting from a zero initial condition $YS(1) = 0$ rather than $\ln(x_0)$, for $i = 1:N$, is approximated by

$$YS(i+1) = YS(i) + (\mu_0 - \sigma_0^2/2) * Dt + \sigma_0 * DW(i) + \texttt{log}(1+\nu_0) * DP(i)\,,$$

with $t(i+1) = i * Dt$. The desired state, XS, sample path is approximated by

$$X(t(i+1)) \simeq XS(i+1) = x_0 * \exp(YS(i+1)) \, .$$

The mean trajectory, XM,

$$\mathrm{E}[X(t(i+1))] \simeq XM(i+1) = x_0 * \exp((\mu_0 + \lambda_0 * \nu_0)t(i+1))$$

is also displayed in the figure along with upper XT exponential standard deviation estimate

$$\mathrm{E}[X(t(i+1))] * V(i+1) \simeq XT(i+1) = XM(i+1) * V(i+1)$$

and lower XB exponential standard deviation estimate

$$\mathrm{E}[X(t(i+1))]/V(i+1) \simeq XB(i+1) = XM(i+1)/V(i+1) \, ,$$

where the factor

$$V(i+1) = exp(\sqrt{\mathrm{Var}[Y(t(i+1))]}) = \exp\left(\sqrt{(\sigma_0^2 + \lambda_0 * log^2(1+\nu_0))t(i+1)}\right)$$

is the exponential of the standard deviation of the exponent process $Y(t)$ in discrete form. Alternatively, one plus or minus the coefficient of variation formula (4.83) could be used to form a deviation factor, but the factor $V(i+1)$ above is more appropriate since it corresponds better to the finite difference simulation approximation. Although the jump amplitude is only a 10 per cent decrement, the jumps are very noticeable in the figure, while both the jump and diffusion component processes result in excesses beyond the indicated upper and lower standard deviation estimates. The estimates correspond to rough confidence intervals and not bounds. See Program C.14 in the Appendix C for the MATLAB code used to produce this figure.

The same code, in the case of constant parameters, can be used for the pure diffusion model in the example (4.24) by setting $\nu_0 = 0$ for the diffusion as shown in Fig. 4.4 or for the pure jump model in the example (4.56) by setting $\sigma_0 = 0$ for the jump process as shown in Fig. 4.5

Remarks 4.28.

- *Simulation Caution: Note that the constant coefficient closed-form solution (4.80) is not used directly, i.e.,*

$$X_i = X_0(1+\nu_0)^{P_i} \exp\left((\mu_0 - \sigma_0^2/2)t_i + \sigma_0 W_i\right) \, ,$$

for $i = 0 : (n+1)$, where $t_{n+1} = T$ is the final time, by directly simulating the random variables P_i and W_i, since they are not independent of either earlier or later values P_j and W_j for $j \neq i$. So such a simulation would be incorrectly approximated. However, simulating the increment set $\{\Delta P_i, \Delta W_i\}$

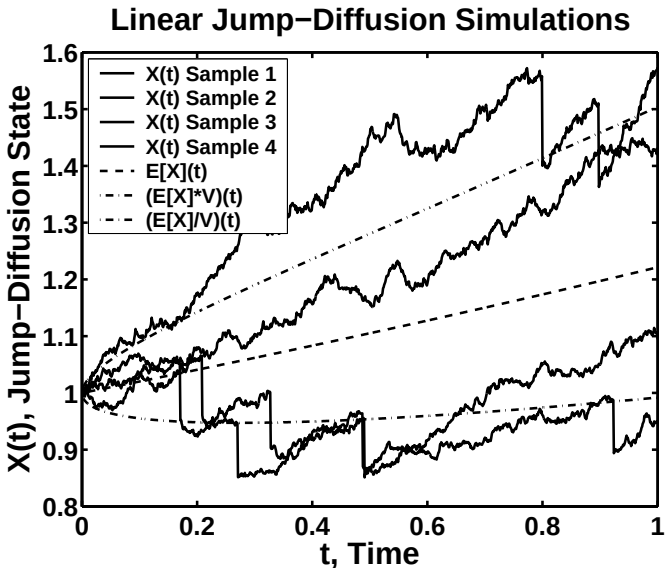


Figure 4.3. *Four linear jump-diffusion sample paths for constant coefficients are simulated using MATLAB [210] with* $N = 1000$ *sample points, maximum time* $T = 1.0$ *and four* `randn` *and four* `rand` *states. Parameter values are* $\mu_0 = 0.5$, $\sigma_0 = 0.10$, $\nu_0 = -0.10$, $\lambda_0 = 3.0$ *and* $x_0 = 1.0$. *In addition to the four simulated states, the expected state* $\mathrm{E}[X(t)]$ *and two deviation measures* $\mathrm{E}[X(t)] * V(t)$ *and* $\mathrm{E}[X(t)]/V(t)$, *where the factor* $V(t)$ *is based on the standard deviation of the state exponent* $Y(t)$.

for $i = 0 : n$ *would be an appropriate use of the approximate independence property of the pseudo-random number generators of* ΔP_i *and* ΔW_i, *i.e.,*

$$X_{i+1} = X_i(1+\nu_0)^{\Delta P_i} \exp\left((\mu_0 - \sigma_0^2/2)\Delta t_i + \sigma_0 \Delta W_i\right) ,$$

for $i = 0 : n$, *noting*

$$\Delta Y_i = (\mu_0 - \sigma_0^2/2)\Delta t_i + \sigma_0 \Delta W_i + \ln(1+\nu_0)\Delta P_i$$

and that $\exp(\ln(1+\nu_0)\Delta P_i) = (1+\nu_0)^{\Delta P_i}$ *using the exponential-logarithm inverse relationship. Considering finite precision arithmetic, this would be similar to using*

$$W_{i+1} = \sum_{j=0}^{i} \Delta W_j \quad \text{and} \quad P_{i+1} = \sum_{j=0}^{i} \Delta P_j \, ,$$

for $(i+1) = 1 : (n+1)$.

It is ***important to build simulations in independent increments****.*

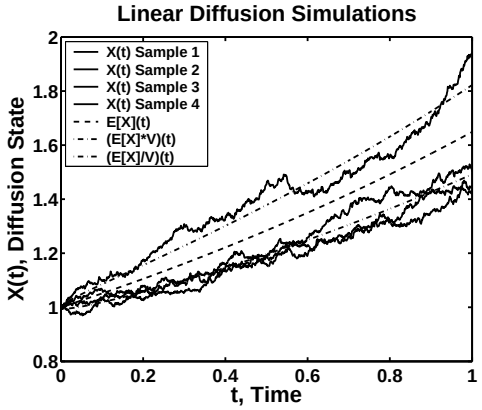


Figure 4.4. *Four linear pure diffusion sample paths for constant coefficients are simulated using MATLAB [210] with* $N = 1000$ *sample points, maximum time* $T = 1.0$ *and four* `randn` *states. Parameter values are* $\mu_0 = 0.5$, $\sigma_0 = 0.10$, $\nu_0 = 0.0$, *and* $x_0 = 1.0$. *In addition to the four simulated states, the expected state* $\mathrm{E}[X(t)]$ *and two deviation measures* $\mathrm{E}[X(t)] * V(t)$ *and* $\mathrm{E}[X(t)]/V(t)$ *are displayed, where the factor* $V(t)$ *is based on the standard deviation of the state exponent* $Y(t)$.

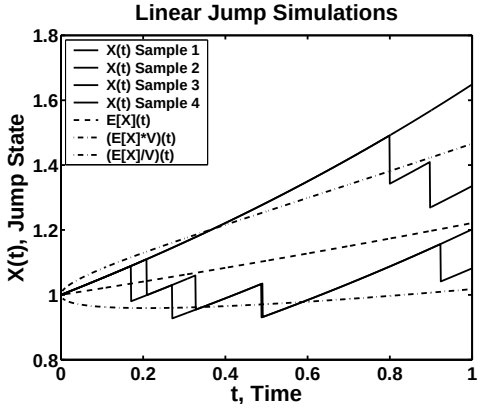


Figure 4.5. *Four linear pure jump with drift sample paths for constant coefficients are simulated using MATLAB [210] with* $N = 1000$ *sample points, maximum time* $T = 1.0$ *and four* `randn` *states. Parameter values are* $\mu_0 = 0.5$, $\sigma_0 = 0.0$, $\nu_0 = -0.10$, *and* $x_0 = 1.0$. *In addition to the four simulated states, the expected state* $\mathrm{E}[X(t)]$ *and two deviation measures* $\mathrm{E}[X(t)] * V(t)$ *and* $\mathrm{E}[X(t)]/V(t)$ *are displayed, where the factor* $V(t)$ *is based on the standard deviation of the state exponent* $Y(t)$.

- *Other SDE codes can be found in the literature. Maple™ codes for jump-diffusions can be found in the paper of Cyganowski, Grüne and Kloeden [64] along with higher approximations. In the numerical SDE tutorial review, D. Higham lists some very readable MATLAB™ codes modeled on techniques from the superb MATLAB™ guide of D. and N. Higham. Both Maple™ and MATLAB™ codes for diffusion SDEs for finance can be found in D. Higham and Kloeden [143] along with higher order approximations. See also the recent probability and SDEs book of Cyganowski, Kloeden and Ombach [66] for more on Maple codes for diffusions. For diffusion SDE codes in Mathematica™ the reader can consult the computational finance oriented book of Stojanovic [259]. Higher order, but older, diffusion SDE codes are found in the computational Kloeden, Platen and Schurz [166] and are also used for the illustrations in the more theoretical treatise of Kloeden and Platen [165]. However these codes are in Turbo-Pascal, not often used now.*

- *More computational SDE methods will be discussed in the compact Section 9.1 of Chapter 9. This section is a good introduction to SDE simulations for readers and instructors interested in exploring the topic further. Since the Itô forward integration uses the Euler's or the tangent line method for purely deterministic processes and Euler's method is perhaps the crudest numerical method for differential equations, higher order numerical methods are important when accuracy is important. Some sample codes are given in Section 9.1 and in Appendix C. See also [64, 143, 66, 139]. However, Euler's method is the most genuine application of Itô's stochastic integration for Markov processes in continuous time, although the simulations sample size should be large for reasonable representation of the stochastic processes.*

Linear Jump-Diffusion SDEs with Time-Dependent Coefficients:

While linear constant coefficient SDEs often occur in applications such as elementary finance, time-dependence of market parameters can play an important role. For this reason, our attention returns to the time-dependent coefficients of the linear SDE solution (4.78) and the expected state trajectory. However, the procedure is more complex than in the simple constant coefficient case, since the expectations of exponentials of integral are required. First the pure diffusion case is considered then the pure jump case in the following two lemmas and related corollaries.

Lemma 4.29. ***Expectation of*** $\exp(\int \sigma dW(s))$***:***
Let $\sigma(t)$ *be square integrable on* $[t_0, t]$*, then*

$$\mathrm{E}\left[\exp\left(\int_{t_0}^{t}\sigma(s)dW(s)\right)\right] \overset{ims}{=} \exp\left(\frac{1}{2}\int_{t_0}^{t}\sigma^2(s)ds\right) . \tag{4.86}$$

Sketch of Proof. To keep the justification reasonably brief and maintain the usefulness as an integration technique, the stochastic diffusion integral will first

be formally decomposed into a forward Itô sum, averaged and then recomposed back into a deterministic integral. The justification of each step will be indicated in short hand on the sign of the relation, but the more rigorous Itô limits will be omitted. Let $t_i = t_0 + i * \Delta t$ for $i = 0 : n+1$ be a proper partition of $[t_0, t]$ with $\Delta t = (t - t_0)/(n+1)$, $\Delta W_i = W(t_{i+1}) - W(t_i)$ and $\sigma_i = \sigma(t_i)$ for $i = 0 : n$.

$$\begin{aligned}
\mathrm{E}\left[\exp\left(\int_{t_0}^{t} \sigma(s)dW(s)\right)\right] &\overset{ims}{\simeq} \mathrm{E}\left[\exp\left(\sum_{i=0}^{n} \sigma_i \Delta W_i\right)\right] \\
&\overset{law}{\underset{exp.}{=}} \mathrm{E}\left[\prod_{i=0}^{n} \exp\left(\sigma_i \Delta W_i\right)\right] \overset{ind}{\underset{inc}{=}} \prod_{i=0}^{n} \mathrm{E}_{\Delta W_i}\left[\exp\left(\sigma_i \Delta W_i\right)\right] \\
&\overset{norm.}{\underset{dist.}{=}} \prod_{i=0}^{n} \int_{-\infty}^{+\infty} \frac{\exp\left(-\frac{w^2}{2\Delta t} + \sigma_i w\right)}{\sqrt{2\pi\Delta t}} dw \overset{comp.}{\underset{sq.}{=}} \prod_{i=0}^{n} \exp\left(\sigma_i^2 \Delta t/2\right) \\
&\overset{ims}{\simeq} \exp\left(\tfrac{1}{2}\int_{t_0}^{t} \sigma^2(s)ds\right) .
\end{aligned}$$

□

Lemma 4.30. ***Expectation of*** $\exp(\int \ln(1+\nu)dP(s))$***:***
Let $\lambda(t)\nu(t)$ be integrable on $[t_0, t]$, then

$$\mathrm{E}\left[\exp\left(\int_{t_0}^{t} \ln(1+\nu(s))dP(s)\right)\right] \overset{ims}{=} \exp\left(\int_{t_0}^{t} \lambda(s)\nu(s)ds\right) . \tag{4.87}$$

Sketch of Proof. Again, to keep the justification reasonably brief and maintain the usefulness as an integration technique, the stochastic diffusion integral will first be formally decomposed into a forward Itô sum, averaged and then recomposed back into a deterministic integral. The justification of each step will be indicated in short hand on the sign of the relation, but the more rigorous Itô limits will be omitted. Again, let $t_i = t_0 + i * \Delta t$ for $i = 0 : n+1$ be a proper partition of $[t_0, t]$ with $\Delta t = (t - t_0)/(n+1)$, $\Delta P_i = P(t_{i+1}) - P(t_i)$, $\lambda_i = \lambda(t_i)$ and $\nu_i = \nu(t_i)$ for $i = 0 : n$.

$$\begin{aligned}
\mathrm{E}\left[\exp\left(\int_{t_0}^{t} \ln(1+\nu(s))dP(s)\right)\right] &\overset{ims}{\simeq} \mathrm{E}\left[\exp\left(\sum_{i=0}^{n} \ln(1+\nu_i)\Delta P_i\right)\right] \\
&\overset{law}{\underset{exp.}{=}} \mathrm{E}\left[\prod_{i=0}^{n} \exp\left(\ln(1+\nu_i)\Delta P_i\right)\right] \overset{ind}{\underset{inc}{=}} \prod_{i=0}^{n} \mathrm{E}_{\Delta P_i}\left[\exp\left(\ln(1+\nu_i)\Delta P_i\right)\right] \\
&\overset{pois.}{\underset{dist.}{=}} \prod_{i=0}^{n} \sum_{k=0}^{\infty} e^{-\lambda_i \Delta t} \frac{\lambda_i \Delta t}{k!} (1+\nu_i)^k \overset{exp.}{\underset{sum.}{=}} \prod_{i=0}^{n} e^{-\lambda_i \Delta t + \lambda_i \Delta t(1+\nu_i)} \\
&\overset{ims}{\simeq} \exp\left(\int_{t_0}^{t} \lambda(s)\nu(s)ds\right) .
\end{aligned}$$

□

Using diffusion and jump Lemmas 4.29-4.30, the expectation of the state trajectory $X(t)$ (4.78) for the linear SDE with time-dependent coefficients (4.75) can be readily calculated:

Theorem 4.31. ***Expectation of*** $X(t)$ ***in the Linear Jump-Diffusion SDE with Time-Dependent Coefficients Case:***

Let $\mu(t)$, $\sigma^2(t)$ and $\lambda(t)\nu(t)$ be integrable on $[t_0, t]$, then

$$\begin{aligned} \mathrm{E}[X(t)] &= \mathrm{E}\left[x_0 \exp\left(\int_{t_0}^{t} ((\mu(s) - \sigma^2(s)/2)ds + \sigma(s)dW(s) + \ln(1+\nu(s))dP(s))\right)\right] \\ &\overset{ims}{=} x_0 \exp\left(\int_{t_0}^{t} (\mu(s) + \lambda(s)\nu(s))\, ds\right). \end{aligned} \tag{4.88}$$

Proof. The proof is left as an algebraic exercise for the reader, using Lemmas 4.29-4.30. □

For the corresponding variance $\mathrm{Var}[X(t)]$, see Exercise 16 on page 128. Note that the expectation and the variance reduce to the linear SDE, constant coefficients results given in (4.81) for the expectation and (4.82) for the variance in the case of constant coefficients.

4.3.4 SDE Models Exactly Transformable To Purely Time-Varying Coefficients

In this section, a catalog of exactly transformable jump-diffusion SDE models are given and first the notational correlations are listed for ease of interpreting the list of models and their transformations, where conditions are applicable:

List of SDE Models and Their Transformations

- **Original SDE (4.64):**
  $$dX(t) = f(X(t), t)dt + g(X(t), t)dW(t) + h(X(t), t)dP(t)).$$
- **Transformed Process:** $Y(t) = F(X(t), t)$.
- **Transformed SDE:** $dY(t) = (F_t + F_x f + \frac{1}{2}F_{xx}g^2)dt + F_x g dW(t) + [F]dP(t)$.
- **Target Explicit SDE:** $dY(t) = C_1(t)dt + C_2(t)dW(t) + C_3(t)dP(t)$.
- **Original-Target Coefficient Equations:**
  $$F_t + F_x f + \tfrac{1}{2}F_{xx}g^2 = C_1(t);$$
  $$F_x g = C_2(t);$$
  $$[F] \equiv F(x + h(x,t), t) - F(x,t) = C_3(t).$$
- **Original Coefficients:**
  $$f(x,t) = (C_1(t) - F_t(x,t) - \tfrac{1}{2}F_{xx}(x,t)C_2^2(t)/F_x^2(x,t))/F_x(x,t);$$
  $$g(x,t) = C_2(t)/F_x(x,t);$$
  $$h(x,t) = -x + F^{-1}(F(x,t) + C_3(t)).$$

See Table 4.1 for examples.

In their theoretical and numerical treatise on stochastic differential equations, Kloeden and Platen [165, Section 4.4] list many exact solutions for diffusion SDEs.

Table 4.1. *Table of Example Transforms Listing Original Coefficients in terms of Target and Transform Coefficients:*

Transform $Y \to$ $F(x,t)$	**Plant Coefficient** $f(x,t)$	**Gaussian Coefficient** $g(x,t)$	**Poisson Coefficient** $h(x,t)$
x	$C_1(t)$	$C_2(t)$	$C_3(t)$
$a(t)x + b(t)$	$\frac{C_1(t) - a'(t)x - b'(t)}{a(t)}$	$\frac{C_2(t)}{a(t)}$	$\frac{C_3(t)}{a(t)}$
$a(t)x^2$	$\frac{C_1(t) - a'(t)x^2 - \frac{C_2^2(t)}{4a(t)x^2}}{2a(t)x}$	$\frac{C_2(t)}{2a(t)x}$	$-x \pm \sqrt{x^2 + \frac{C_3(t)}{a(t)}}$
$\frac{a(t)}{x+b(t)}$	$\frac{C_2^2}{a^2}(x+b)^3 - \frac{C_1}{a}(x+b)^2$ $+\frac{a'}{a}(x+b) - b'$	$-\frac{C_2}{a}(x+b)^2$	$-\frac{C_3(x+b)^2}{C_3(x+b)+a}$
$a(t)\frac{x+c(t)}{x+b(t)}$, $\{b \neq c\}$	$\frac{1}{b-c}\Big(\frac{C_2^2}{a^2(b-c)}(x+b)$ $\frac{C_1}{a}(x+b)^2$ $+\frac{b'}{a}(x+c) - c'(x+b)\Big)$ $-\frac{a'}{a}(x+b)(x+c)$ $+\frac{b'}{a}(x+c) - c'(x+b)\Big)$	$-\frac{C_2}{a(b-c)}(x+b)^2$	$-\frac{C_3(x+b)^2}{C_3(x+b)-a(b-c)}$
$a(t)e^{b(t)x}$	$-(\frac{a'}{ab} + \frac{b'}{b}) + \frac{C_1}{ab}e^{-bx}$ $-\frac{1}{2}\frac{C_2^2}{a^2b}e^{-2bx}$	$\frac{C_2}{ab}e^{-bx}$	$\frac{1}{b}\ln\left(\frac{C_3}{a}e^{-bx} + 1\right)$
$a(t)\ln(x) + b(t)$	$\Big(\frac{C_1}{a} + \frac{C_2^2}{2a^2}$ $-\frac{a'}{a}\ln(x) - \frac{b'}{a}\Big)x$	$\frac{C_2}{a}x$	$\left(e^{C_3/a} - 1\right)x$

They also give a comprehensive treatment of convergence and stability of numerical approximations to solutions of SDEs that are well beyond this text.

4.4 Poisson Noise is White Noise Too!

Noise can be rapid fluctuations or disturbances, so stochastic processes are sometimes called noise as well. Another typical feature of noise is that it contains many frequencies, so such noise can also be called **colored noise**. If the noise contains all frequencies then it is called **white noise**, in analogy with white light containing all colors of the light spectrum.

There are two principal kinds of white noise in stochastic processes, **Gaussian white noise** if the noise is normally distributed or **Poisson white noise** if the noise is Poisson distributed. The white-ness of the noise relies heavily on the independent increment property.

However, many use the term white noise without qualification to refer to Gaussian white noise, perhaps because of the lack of familiarity with the other main Markov process in continuous time that also is white noise, the Poisson process. An exception is Arnold [13], who treats mainly Gaussian white noise, but properly mentions that Poisson noise is also white noise as well.

Consider the Gaussian case first. It is necessary to look at the covariance of the Wiener increments at different times relative to the time increment,

$$\begin{aligned} C_{\Delta W(t)}(h) &\equiv \mathrm{Cov}[\Delta W(t)/\Delta t, \Delta W(t+h)/\Delta t] \\ &= \mathrm{E}[\Delta W(t)\Delta W(t+h)]/(\Delta t)^2 \,, \end{aligned} \tag{4.89}$$

where $\Delta W(t) \equiv W(t+\Delta t) - W(t)$, the time increment $\Delta t > 0$ as usual, but $h \neq 0$. The covariance in (4.89) is also related to the correlation coefficient (B.142) between $\Delta W(t)$ and $\Delta W(t+h)$, noting that the reciprocal $1/\Delta t$ scales out of the correlation coefficient, by

$$\rho_{\Delta W(t)}(h) = \Delta t \cdot C_{\Delta W(t)}(h) \,, \tag{4.90}$$

using $\sqrt{\mathrm{Var}[\Delta W(t)]} = \sqrt{\mathrm{Var}[\Delta W(t+h)]} = \sqrt{\Delta t}$. Since $W(t)$ is not differentiable, the finite difference approximation $\Delta W(t)/\Delta t$ is used in place of its rate or velocity so we can eventually let $\Delta t \to 0^+$. Using the independent increments property and the zero-mean property $\mathrm{E}[\Delta W(t)] = 0$, separating $\Delta W(t)$ and $\Delta W(t+h)$ into independent and common increments,

$$\begin{aligned} C_{\Delta W(t)}(h) &= \frac{1}{(\Delta t)^2} \left\{ \begin{array}{ll} 0, & h \leq -\Delta t \\ \Delta t + h, & -\Delta t \leq h \leq 0 \\ \Delta t - h, & 0 \leq h \leq \Delta t \\ 0, & h \geq +\Delta t \end{array} \right\} \\ &= \frac{(\Delta t - |h|)}{(\Delta t)^2} U(h; -\Delta t, +\Delta t) \,, \end{aligned} \tag{4.91}$$

where $U(x; a, b)$ is the unit step function on $[a, b]$ and is used to give $C_{\Delta W(t)}(h)$ a more compact form.

Next, we seek the limiting generalized behavior of $C_{\Delta W(t)}(h)$ as $\Delta t \to 0^+$ by considering the integral of a sufficiently well-behaved "test" function, $F(h)$ and by

using the step function representation in (4.91),

$$\begin{aligned}\int_{-\infty}^{+\infty} C_{\Delta W(t)}(h)F(h)dh &= \int_{-\Delta t}^{+\Delta t} \frac{(\Delta t - |h|)}{(\Delta t)^2} F(h)dh \\ &= \int_{-1}^{+1} (1-|u|)F(u\Delta t)du \\ &\to F(0)\int_{-1}^{+1}(1-|u|)du = F(0)\,, \end{aligned} \tag{4.92}$$

where the change of variables $h = u\Delta t$ moved all Δt's into the argument of f and subsequently an expansion retained the leading term and neglected errors of order Δt. Thus, we have the generalized form corresponding to the covariance of differential $dW(t)$,

$$C_{dW(t)}(h) = \delta(h)\,, \tag{4.93}$$

where $\delta(h)$ is the Dirac delta function. Thus, Gaussian noise is also called **delta-correlated** and delta-correlation is closely connected with the notion of white noise, but note that the actual correlation coefficient (4.90) goes to zero as $\Delta t \to 0^+$.

Finally, to examine the frequency spectrum of $C_{dW(t)}(h)$, consider the power density spectrum using the Fourier transform,

$$\begin{aligned}\mathcal{F}[C_{dW(t)}](k) &= \frac{1}{\pi}\int_{-\infty}^{+\infty} e^{-ikh} C_{dW(t)}(h)dh \\ &\stackrel{\text{gen}}{=} \frac{1}{\pi}\int_{-\infty}^{+\infty} e^{-ikh}\delta(h)dh = \frac{1}{\pi}\,, \end{aligned} \tag{4.94}$$

which is certainly constant, so $C_{dW(t)}(h)$ has a flat frequency spectrum and thus represents an approximation to Gaussian white noise, $dW(t)$ being normally distributed.

Similarly, for the simple Poisson process, which we consider in the zero-mean form,

$$\Delta\widehat{P}(t) \equiv \Delta P(t) - \lambda_0 \Delta t$$

where $\lambda_0 > 0$ is a constant jump rate, then the covariance of the time-separated finite difference velocities is

$$\begin{aligned} C_{\Delta P(t)}(h) &\equiv \text{Cov}[\Delta\widehat{P}(t)/\Delta t, \Delta\widehat{P}(t+h)/\Delta t] \\ &= \text{E}[\Delta\widehat{P}(t)\Delta\widehat{P}(t+h)]/(\Delta t)^2 \\ &= \frac{\lambda_0(\Delta t - |h|)}{(\Delta t)^2} U(h; -\Delta t, +\Delta t) \\ &\stackrel{\text{gen}}{\longrightarrow} \lambda_0\delta(h) \stackrel{\text{gen}}{=} C_{dP(t)}(h)\,, \end{aligned} \tag{4.95}$$

taking a similar limit as $\Delta t \to 0^+$ as with $C_{\Delta W(t)}(h)$ above. Hence, Poisson noise is also **delta-correlated**. For the Poisson increment process, recalling $\text{Var}[\Delta P(t)] = \text{Var}[\Delta\widehat{P}(t)] = \lambda_0\Delta t$, so the corresponding correlation coefficient is

$$\rho_{\Delta P(t)}(h) = \Delta t \cdot C_{\Delta P(t)}(h)/\lambda_0\,. \tag{4.96}$$

Finally, taking the Fourier transform of $C_{dP(t)}(h)$,

$$\begin{aligned}\mathcal{F}[C_{dP(t)}](k) &= \frac{1}{\pi}\int_{-\infty}^{+\infty} e^{-ikh} C_{dP(t)}(h)dh \\ &\stackrel{\text{gen}}{=} \lambda_0 \frac{1}{\pi}\int_{-\infty}^{+\infty} e^{-ikh}\delta(h)dh = \lambda_0 \frac{1}{\pi}\,, \end{aligned} \tag{4.97}$$

which is also a constant so that **Poisson noise is also white noise**.

4.5 Exercises

1. Derive the Itô stochastic integral formulas for

$$\int_0^t \cos(aW(s))dW(s) \quad \text{and} \quad \int_0^t \sin(aW(s))dW(s)\,,$$

where $a \neq 0$ and is a real constant. Also, derive the results when $a = 0$.

2. Find $X(t)$ if

$$\int_0^t X(s)dP(s) = b^{P(t)}\ln(aP(t)+c)\,,$$

where $a > 0$, $b > 0$ and $c > 0$ are real constants.

3. Derive the following using stochastic calculus:

   (a) $\int_0^t \sin(\pi P(s))dP(s) \stackrel{\text{dt}}{=} -\frac{1}{2}\sin(\pi P(t))$.

   (b) $\int_0^t \cos(\pi P(s))dP(s) \stackrel{\text{dt}}{=} \frac{1}{2}(1-\cos(\pi P(t)))$.

   {*Hint: You may need some elementary trigonometric identities.*}

4. Consider the simple linear jump-diffusion SDE,

$$dX(t) = (\mu_d dt + \sigma_d dW(t) + JdP(t))X(t)\,,$$

where the $\{\mu_d, \sigma_d, \mu_j, \sigma_j, \lambda_0\}$ are constants and λ_0 is the Poisson jump rate, while μ_j is the mean of the jump amplitude J and σ_j^2 is the jump amplitude variance. The diffusion process $W(t)$ is independent of the jump process $P(t)$ and the jump amplitude J independent of $P(t)$ conditioned on a jump of $P(t)$. Show that the conditional infinitesimal mean is given by

$$\mathrm{E}[dX(t)|X(t)=x] = (\mu_d + \lambda_0\mu_j)xdt$$

and the conditional infinitesimal variance is given by

$$\mathrm{Var}[dX(t)|X(t)=x] \stackrel{\text{dt}}{=} (\sigma_d^2 + \lambda_0(\sigma_j^2+\mu_j^2))x^2dt,$$

explaining why equality in dt-precision (see Chapter 1 for a definition) is required in the latter but not in the former conditional moment.

5. Show that the **power rules** for (Itô) stochastic integration for Wiener noise can be written as the recursions,

$$\int_0^t W^m(s)dW(s) = \frac{1}{m+1}W^{m+1}(t) - \frac{m}{2}\int_0^t W^{m-1}(s)ds, \qquad (4.98)$$

   (a) Illustrate the application of the formulas to find the results for the cases $m = 2$ and $m = 3$.

   (b) Alternatively, using the (Itô) stochastic chain rule and mathematical induction, show the general result.

6. Solve the following (Itô) diffusion SDE for $X(t)$, $\mathrm{E}[X(t)]$, and $\mathrm{Var}[X(t)]$,

$$dX(t) = \left(a\sqrt{X(t)} + b^2/4\right)dt + b\sqrt{X(t)}dW(t)\,,$$

   where a and b are real constants, and $X(0) = x_0 > 0$, with probability one. {*Hint: Seek a transformation* $Y(t) = f(X(t))$ *for some* f *such that* $Y(t)$ *satisfies a constant coefficient SDE.*}

7. Solve the following (Itô) diffusion SDE for $X(t)$, $\mathrm{E}[X(t)]$, and $\mathrm{Var}[X(t)]$,

$$dX(t) = \left(aX^2(t) + b^2X^3(t)\right)dt + bX^2(t)dW(t)\,,$$

   where a and b are real constants, and $X(0) = x_0 > 0$, with probability one. {*Hint: Seek a transformation* $Y(t) = f(X(t))$ *for some* f *such that* $Y(t)$ *satisfies a constant coefficient SDE.*}

8. Solve the following diffusion SDE for $X(t)$ and $\mathrm{E}[X(t)]$,

$$dX(t) = \left(aX^{3/4}(t) + \frac{3}{8}b^2X^{1/2}(t)\right)dt + bX^{3/4}(t)dW(t)\,,$$

   where a and b are real constants, and $X(0) = x_0 > 0$, with probability one. {*Hint: Find a power transformation to convert the SDE to a constant coefficient SDE.*}

9. Solve the following (Itô) jump SDE for $X(t)$, $\mathrm{E}[X(t)]$, and $\mathrm{Var}[X(t)]$,

$$dX(t) = -aX^2(t)dt - \frac{cX^2(t)}{1+cX(t)}dP(t)\,,$$

   where $\mathrm{E}[P(t)] = \lambda_0 t$ and $X(0) = x_0 > 0$, with probability one, while $a > 0$, $b > 0$ and $\lambda_0 > 0$ are constants. The answer may be left as a sum over the Poisson distribution.
   {*Hint: Seek a transformation* $Y(t) = f(X(t))$ *for some* f *such that* $Y(t)$ *satisfies a constant coefficient SDE.* }

10. Solve the following Poisson jump SDE for $X(t)$ and $\mathrm{E}[X(t)]$,

$$dX(t) = a\sqrt{X(t)}dt + b\left(b + 2\sqrt{X(t)}\right)dP(t)\ ,$$

where $\mathrm{E}[P(t)] = \lambda_0 t$ and $X(0) = x_0 > 0$, with probability one, while λ_0, a and b are real constants. {*Hint: Find a power transformation to convert the SDE to a constant coefficient SDE.*}

11. Show that the (Itô) jump-diffusion SDE for $X(t)$,

$$dX(t) = f(X(t))dt + bX^r(t)dW(t) + h(X(t))dP(t)\ ,$$

can be transformed by $Y(t) = F(X(t))$ to a **constant coefficient SDE**,where $\mathrm{E}[P(t)] = \lambda_0 t$ and $X(0) = x_0 > 0$, with probability one, while λ_0, b and $r \neq 0$ are real constants. In a proper answer, the power forms of $f(X(t))$ and $h(X(t))$ must be derived from the constant coefficient SDE condition.

12. **Martingales:** A **martingale** in continuous time satisfies the essential property that

$$\mathrm{E}[M(t)|M(s)] = M(s), \tag{4.99}$$

for all $0 \leq s < t]$ provided its absolute value has finite expectation, $\mathrm{E}[|M(t)|] < \infty$ for all $t \geq 0$ plus some other technical properties (see Mikosch [209], for instance).

Driftless Log-Linear Process $\Longrightarrow$ Martingale?
Show directly that

$$M_1(t) = \ln(X(t)) - \mathrm{E}[\ln(X(t))] \tag{4.100}$$

is a martingale using that $Y(t) = \ln(X(t))$ symbolically satisfies the solution to general linear SDE transformed to state-lindependent, time-dependent form (4.76).

Remark 4.32. *This type of problem is applicable to many financial problem where the return on a linear financial asset $X(t)$ is transformed to a log-return form $\ln(X(t))$, forming an SDE with state-independent coefficients, so the driftless deviation $M_1(t)$ form in (4.100) is a martingale, a log-martingale. However, readers must be aware of all the assumptions involved. See the next exercise.*

13. **Driftless $\Longrightarrow$ Martingale?**
Prove the following theorem, explicitly justifying every step where an underlined theorem assumption or expectation property is needed.

Theorem 4.33. *If the **Markov** process $X(t)$ is **driftless** (i.e., $\mathrm{E}[X(t)] = 0$) and has **independent increments** (along with the boundedness and technical condition cited with (4.99), then $X(t)$ is a martingale.*

Remark 4.34. *Readers must aware of the direction of the implication. For example, Hull [147, p. 507] states "A martingale is a zero-drift stochastic process.", while Baxter and Rennie [22, p. 79] state "X is a martingale $\Longleftrightarrow$ X is driftless ($\mu_t \equiv 0$).", yet all of the assumptions are not apparent. For example, the state-independence of the SDE for the log-return is in the background.*

14. **Exponential-Martingale Counterexample to Driftless Martingale Requirement?**

    (a) Derive the nonrandom function $\beta(t)$ that makes

    $$M_2(t) = \beta(t)X(t)$$

    a martingale if $X(t)$ symbolically satisfies the linear SDE (4.75).

    (b) Show that $M_2(t)$ is not driftless, i.e.,

    $$\mathrm{E}[M_2(t)] \neq 0,$$

    in absence of trivial initial conditions, i.e., $x_0 \neq 0$.

Remark 4.35. *This is a counterexample showing that if $M(t)$ is a martingale then it in not necessarily a driftless process.*

15. **General Exponential-Expectation Interchange Formula for Linear Jump-Diffusions:**
    Formally show that

    $$\mathrm{E}\left[\exp\left(\int_0^t d\ln(X)(s)\right)\right] = \exp\left(\mathrm{E}\left[\int_0^t \left(\frac{dX}{X}\right)(s)\right]\right), \qquad (4.101)$$

    if $X(t)$ is a linear jump-diffusion process (4.75), verifying that both sides of (4.101) are equivalent. Assume all integrals of process coefficients are bounded.

16. For the solution $X(t)$ (4.78) of the linear SDE (4.75) with time-dependent coefficients, assuming all integrals of process coefficients are bounded,

    (a) Calculate the expectation of the quadratic of the exponential of the diffusion integral in Lemma 4.29 transforming the results of the lemma, or using the same techniques as in the lemma.

    (b) Calculate the expectation of the quadratic of the exponential of the jump integral in Lemma 4.30 transforming the results of the lemma.

(c) Using the result of the first two parts of this exericise and the expectation Theorem 4.31 to show

Theorem 4.36. ***Variance of $X(t)$ in the Linear SDE with Time-Dependent Coefficients Case:***

Let $\mu(t)$, $\sigma^2(t)$ and $\lambda(t)\nu^j(t)$ for $j = 1:2$ be integrable on $[t_0, t]$, then

$$\begin{aligned} \mathrm{Var}[X(t)] \overset{ims}{=} & \; x_0^2 \exp\left(2\int_{t_0}^{t} (\mu(s) + \lambda(s)\nu(s))\, ds\right) \\ & \cdot \left(\exp\left(\int_{t_0}^{t} \left(\sigma^2(s) + \lambda(s)\nu^2(s)\right) ds\right) - 1\right), \end{aligned} \tag{4.102}$$

for the state trajectory $X(t)$ given in (4.78).

Suggested References for Further Reading

- Cyganowski, Grüne and Kloeden, 2002 [64].
- Cyganowski, Kloeden and Ombach, 2002 [66].
- Gard, 1988 [91].
- Glasserman, 2003 [96].
- D. Higham and Kloeden, 2002 [143] and 2005 [144].
- Jazwinski, 1970 [154].
- Karlin and Taylor, 1981 [162].
- Klebaner, 1998 [164].
- Kloeden and Platen, 1992 [165].
- Kloeden, Platen and Schurz, 1994 [166].
- Mikosch, 1998 [209].
- Øksendal, 1998 [222].
- Schuss, 1980 [244].
- Shreve, 2004 [248].
- Snyder and Miller, 1991 [252].
- Taylor and Karlin, 1998 [265].
- Tuckwell, 1995 [270].
- Wonham, 1970 [285].

Chapter 5

Stochastic Calculus for General Markov SDEs: Space-Time Poisson, State-Dependent Noise and Multi-Dimensions

Not everything that counts can be counted,
and not everything that can be counted counts.
—Albert Einstein (1879-1955).

The only reason time is so that everything does not happen at once
—Albert Einstein (1879-1955),
`http://en.wikiquote.org/wiki/Time` .

Time is what prevents everything from happening at once.
Space is what prevents everything from happening to me.
—attributed to John Archibald Wheeler (1911–),
`http://en.wikiquote.org/wiki/Time` .

What about stochastic effects?
—Don Ludwig, University of British Columbia, printed on his tee-shirt to save having to ask it at each seminar.

We are born by accident into a purely random universe. Our lives are determined by entirely fortuitous combinations of genes. Whatever happens happens by chance. The concepts of cause and effect are fallacies. There is only seeming *causes leading to* apparent *effects. Since nothing truly follows from anything else, we swim each day through seas of chaos, and nothing is predictable, not even the events of the very next instant.*

Do you believe that?

If you do, I pity you, because yours must be a bleak and terrifying and comfortless life.
—Robert Silverberg in *The Stochastic Man*, 1975.

This chapter completes the generalization of Markov noise in continuous time for this book, by including space-time Poisson noise, state-dependent SDEs and multi-dimensional SDEs.

5.1 Space-Time Poisson Process

Space-time Poisson processes are also called general compound Poisson processes, marked Poisson point processes and Poisson noise with randomly distributed jump-amplitudes conditioned on a Poisson jump in time. The marks denote the underlying stochastic process for the Poisson jump-amplitude or the space component of the space-time Poisson process, whereas the jump-amplitudes of the simple Poisson process are deterministic or fixed with unit magnitude. The space-time Poisson process is a generalization of the Poisson process. The space-time Poisson process formulation helps in understanding the mechanism for applying it to more general jump applications and generalization of the chain rules of stochastic calculus.

Properties 5.1.

- ***Space-Time Poisson Differential Process:*** *The basic space-time or mark-time Poisson differential process denoted as*

$$d\Pi(t) = \int_{\mathcal{Q}} h(t,q)\mathcal{P}(\mathbf{dt},\mathbf{dq}) \tag{5.1}$$

  on the ***Poisson mark space*** $\mathcal{Q}$ *can be defined using the* ***Poisson random measure*** $\mathcal{P}(\mathbf{dt},\mathbf{dq})$*, which is shorthand measure notation for the measure-set equivalence* $\mathcal{P}(\mathbf{dt},\mathbf{dq}) = \mathcal{P}((t,t+dt],(q,q+dq])$*. The jump-amplitude* $h(t,q)$ *is assumed to be continuous and bounded in its arguments.*

- ***Poisson mark*** Q***:*** *The space Poisson mark* Q *is the underlying IID random variable for the mark-dependent jump-amplitude coefficient denoted by* $h(t,Q)$*=1, i.e., the space part of the space-time Poisson process. The realized variable* $Q = q$ *is used in expectations or conditional expectations, as well as in definition of the type (5.1).*

- ***Time-integrated, Space-Time Poisson Process:***

$$\Pi(t) = \int_0^t \int_{\mathcal{Q}} h(t,q)\mathcal{P}(\mathbf{dt},\mathbf{dq})dt\ . \tag{5.2}$$

- ***Unit Jumps:*** *However, if the jumps have unit amplitudes,* $h(t,Q) \equiv 1$*, then the space time process in (5.1) must be the same result as the simple differential Poisson process* $dP(t;Q)$ *modified with a mark parameter argument to allow for generating mark realizations, and we must have the equivalence,*

$$\int_{\mathcal{Q}} \mathcal{P}(\mathbf{dt},\mathbf{dq}) \equiv dP(t;Q)\ , \tag{5.3}$$

giving the jump number count on $(t, t+dt]$. Integrating both sides of (5.3) on $[0,t]$, gives the jump-count up to time t,

$$\int_0^t \int_{\mathcal{Q}} \mathcal{P}(\mathbf{dt}, \mathbf{dq}) = \int_0^t dP(s;Q) = P(t;Q) \; . \tag{5.4}$$

Further, in terms of Poisson random measure $\mathcal{P}(\mathbf{dt}, \{1\})$ on the fixed set $\mathcal{Q} = \{1\}$, purely the number of jumps in $(t, t+dt]$ is obtained,

$$\int_{\mathcal{Q}} \mathcal{P}(\mathbf{dt}, \mathbf{dq}) = \mathcal{P}(\mathbf{dt}, \{1\}) = P(\mathbf{dt}) = dP(t;1) \equiv dP(t)$$

and the marks are irrelevant.

- ***Purely, Time-Dependent Jumps:*** *If $h(t,Q) = h_1(t)$, then*

$$\int_{\mathcal{Q}} h_1(t) \mathcal{P}(\mathbf{dt}, \mathbf{dq}) \equiv h_1(t) dP(t;Q) \; . \tag{5.5}$$

- ***Compound Poisson Process Form:*** *An alternate form of the space-time Poisson process (5.2), that many may find more comprehensible, is the marked generalization of the* ***simple Poisson process*** *$P(t;Q)$,* ***with IID random mark generation****, that is the counting sum called the* ***compound Poisson process*** *or* ***marked point process****,*

$$\Pi(t) = \sum_{k=1}^{P(t;Q)} h(T_k^-, Q_k) \; , \tag{5.6}$$

where $h(T_k^-, Q_k)$ is the kth jump-amplitude, T_k^- is the pre-jump value of the kth random jump-time, Q_k is the corresponding random jump-amplitude mark realization and for the special case that $P(t;Q)$ is zero the following reverse-sum convention is used,

$$\sum_{k=1}^{0} h(T_k^-, Q_k) \equiv 0 \; , \tag{5.7}$$

for any h. The corresponding differential process has the expectation,

$$\mathrm{E}[dP(t;Q)] = \lambda(t) dt,$$

although it is possible that the jump-rate is mark-dependent (see [223], for example) so that

$$\mathrm{E}[dP(t;Q)] = \mathrm{E}_Q[\lambda(t;Q)] dt.$$

However, it will be assumed here that the jump-rate is mark-independent to avoid complexities with iterated expectations later.

- ***Zero-One Law Compound Poisson Differential Process Form:*** *Given the Poisson compound process form in (5.6), then the corresponding* ***zero-one jump law*** *for the compound Poisson differential process is*

$$d\Pi(t) = h(t, Q)dP(t; Q) \ , \tag{5.8}$$

such that the jump in $\Pi(t)$ *at* $t = T_k$ *is given by*

$$[\Pi](T_k) \equiv \Pi(T_k^+) - \Pi(T_k^-) = h(T_k^-, Q_k) \ . \tag{5.9}$$

For consistency with the Poisson random measure and compound Poisson process forms, it is necessary that

$$\int_0^t h(s, Q)dP(s; Q) = \int_0^t \int_{\mathcal{Q}} h(s, q)\mathcal{P}(\mathbf{ds}, \mathbf{dq}) = \sum_{k=1}^{P(t;Q)} h(T_k^-, Q_k) \ ,$$

so

$$\int_0^t dP(s; Q) = \int_0^t \int_{\mathcal{Q}} \mathcal{P}(\mathbf{ds}, \mathbf{dq}) = P(t; Q)$$

and

$$dP(t; Q) = \int_{\mathcal{Q}} \mathcal{P}(\mathbf{dt}, \mathbf{dq}).$$

Note that the selection of the random marks depends on the existence of the Poisson jumps and the mechanism is embedded in $dP(t; Q)$*, in the formulation of this book.*

- *In the* ***Poisson random measure notation*** $\mathcal{P}(\mathbf{dt}, \mathbf{dq})$*, the arguments* $\mathbf{dt}$ *and* $\mathbf{dq}$ *are semi-closed subintervals when these arguments are expanded*

$$\mathcal{P}(\mathbf{dt}, \mathbf{dq}) = \mathcal{P}((t, t + dt], (q, q + dq])$$

and these subintervals are closed on the left and open on the right due to the definition of the increment, leaving no overlap between differential increments and correspondings to the simple Poisson right continuity property that

$$\Delta P(t; Q) \to P(t^+; Q) - P(t; Q) \ \text{ as } \ \Delta t \to 0^+$$

so we can write $\Delta P(t; Q) = P((t, t + \Delta t]; Q)$ *and* $dP(t; Q) = P((t, t + dt]; Q)$*. When* $t_n = t$ *and* $t_{i+1} = t_i + \Delta t_i$*, the covering set of intervals is* $\{[t_i, t_i + \Delta t_i) \text{ for } i = 0 : n\}$ *plus* t*. If the marks* Q *are continuously distributed then closed subintervals can also be used in* q *argument. For the one-dimensional mark space* $\mathcal{Q}$*,* $\mathcal{Q}$ *can be a finite interval such as* $\mathcal{Q} = [a, b]$ *or an infinite interval such as* $\mathcal{Q} = (-\infty, +\infty)$*. Also, these subintervals are convenient in partitioning continuous intervals since they avoid overlap at the nodes.*

- $\mathcal{P}$ *has* ***independent increments*** *on non-overlapping intervals in time* t *and marks* q*, i.e.,* $\mathcal{P}_{i,k} = \mathcal{P}((t_i, t_i + \Delta t_i], (q_k, q_k + \Delta q_k])$ *is independent of* $\mathcal{P}_{j,\ell} = \mathcal{P}((t_j, t_j + \Delta t_j], (q_\ell, q_\ell + \Delta q_\ell])$*, provided that the time interval* $(t_j, t_j + \Delta t_j]$

has no overlap with $(t_i, t_i + \Delta t_i]$ *and the mark interval* $(q_k, q_k + \Delta q_k]$ *has no overlap with* $(q_\ell, q_\ell + \Delta q_\ell]$. *Recall that* $\Delta P(t_i; Q) \equiv P(t_i + \Delta t_i; Q) - P(t_i; Q)$ *is associated with the time interval* $(t_i, t_i + \Delta t_j]$, *open on the left since the process* $P(t_i; Q)$ *has been subtracted to form the increment.*

- *The* ***expectation of*** $\mathcal{P}(\mathbf{dt}, \mathbf{dq})$ *is*

$$\mathrm{E}[\mathcal{P}(\mathbf{dt}, \mathbf{dq})] = \Phi_Q(\mathbf{dq})\lambda(t)dt \stackrel{\text{gen}}{=} \phi_Q(q)dq\lambda(t)dt \ , \tag{5.10}$$

where, in detail,

$$\begin{aligned} \Phi_Q(\mathbf{dq}) &= \Phi_Q((q, q + dq]) = \Phi_Q(q + dq) - \Phi_Q(q) \\ &= \mathrm{Prob}[Q \le q + dq] - \mathrm{Prob}[Q \le q] = \mathrm{Prob}[q < Q \le q + dq] \\ &\stackrel{\text{gen}}{=} \phi_Q(q)dq \end{aligned}$$

is the probability distribution measure of the Poisson amplitude mark in measure-theoretic notation corresponding to the mark distribution function $\Phi_Q(q)$ *and where* dq *is short hand for the arguments* $(q, q + dq]$, *just as the* dt *in* $\mathcal{P}(\mathbf{dt}, \mathbf{dq})$ *is short hand for* $(t, t + dt]$. *The corresponding mark density will be equal to* $\phi_Q(q)$ *if* Q *is continuously distributed with continuously distributed distribution function and also if equal in the generalized sense (symbol* $\stackrel{\text{gen}}{=}$*), for instance, if* Q *is discretely distributed. Generalized densities will be assumed for almost all distributions encountered in applications. It is also assumed that* Φ_Q *is a proper distribution,*

$$\int_{\mathcal{Q}} \Phi_Q(\mathbf{dq}) = \int_{\mathcal{Q}} \phi_Q(q)dq = 1 \ .$$

- ***Poisson random measure*** $\mathcal{P}(\mathbf{\Delta t}_i, \mathbf{\Delta q}_j)$ *is* ***Poisson distributed****, i.e.,*

$$\mathrm{Prob}[\mathcal{P}(\mathbf{\Delta t}_i, \mathbf{\Delta q}_j) = k] = e^{-\overline{\mathcal{P}}_{i,j}} \left(\overline{\mathcal{P}}_{i,j}\right)^k / k! \ , \tag{5.11}$$

where

$$\overline{\mathcal{P}}_{i,j} = \mathrm{E}[\mathcal{P}(\mathbf{\Delta t}_i, \mathbf{\Delta q}_j)] = \Phi_Q(\mathbf{\Delta q}_j) \int_{\mathbf{\Delta t}_i} \lambda(t)dt = \Phi_Q(\mathbf{\Delta q}_j)\Lambda(\mathbf{\Delta t}_i) \ ,$$

for sets $\mathbf{\Delta t}_i \equiv [t_i, t_i + \Delta t_i)$ *in time and* $\mathbf{\Delta q}_j \equiv [q_j, q_j + \Delta q_j)$ *in marks.*

Thus, as Δt_i *and* Δq_j *approach* 0^+, *then they can be replaced by* dt *and* dq, *respectively, so*

$$\mathrm{Prob}[\mathcal{P}(\mathbf{dt}, \mathbf{dq}) = k] = e^{-\overline{\mathcal{P}}} \left(\overline{\mathcal{P}}\right)^k / k! \ , \tag{5.12}$$

where

$$\overline{\mathcal{P}} = \mathrm{E}[\mathcal{P}(\mathbf{dt}, \mathbf{dq})] = \phi_Q(q)dq\lambda(t)dt \ ,$$

so by the zero-one jump law,

$$\mathrm{Prob}[\mathcal{P}(\mathbf{dt}, \mathbf{dq}) = k] \stackrel{dt}{\underset{zol}{=}} (1 - \overline{\mathcal{P}})\delta_{k,0} + \mathcal{P}\delta_{k,1} \ .$$

- *The **expectation** of* $dP(t;Q) = \int_{\mathcal{Q}} \mathcal{P}(\mathbf{dt}, \mathbf{dq})$ *is*

$$\mathrm{E}\left[\int_{\mathcal{Q}} \mathcal{P}(\mathbf{dt}, \mathbf{dq})\right] = \lambda(t)dt \int_{\mathcal{Q}} \phi_Q(q)dq = \lambda(t)dt = \mathrm{E}[dP(t;Q)]\ , \quad (5.13)$$

corresponding to the earlier Poisson equivalence (5.3) and using the above proper distribution property. Similarly,

$$\mathrm{E}\left[\int_0^t \int_{\mathcal{Q}} \mathcal{P}(\mathbf{ds}, \mathbf{dq})\right] = \mathrm{E}[P(t;Q)] = \int_0^t \lambda(s)ds = \Lambda(t).$$

- *The **variance** of* $\int_{\mathcal{Q}} \mathcal{P}(\mathbf{dt}, \mathbf{dq}) \equiv dP(t;Q)$ *is by definition*

$$\mathrm{Var}\left[\int_{\mathcal{Q}} \mathcal{P}(\mathbf{dt}, \mathbf{dq})\right] = \mathrm{Var}[dP(t;Q)] = \lambda(t)dt. \quad (5.14)$$

Since

$$\mathrm{Var}\left[\int_{\mathcal{Q}} \mathcal{P}(\mathbf{dt}, \mathbf{dq})\right] = \int_{\mathcal{Q}} \int_{\mathcal{Q}} \mathrm{Cov}[\mathcal{P}(\mathbf{dt}, \mathbf{dq_1}), \mathcal{P}(\mathbf{dt}, \mathbf{dq_2})]\ ,$$

then

$$\mathrm{Cov}[\mathcal{P}(\mathbf{dt}, \mathbf{dq_1}), \mathcal{P}(\mathbf{dt}, \mathbf{dq_2})] \stackrel{\text{gen}}{=} \lambda(t)dt\phi_Q(q_1)\delta(q_1 - q_2)dq_1dq_2\ , \quad (5.15)$$

analogous to (1.48) for $\mathrm{Cov}[dP(s_1), dP(s_2)]$*. Similarly, since*

$$\mathrm{Var}\left[\int_t^{t+\Delta t} \int_{\mathcal{Q}} \mathcal{P}(\mathbf{ds}, \mathbf{dq})\right] = \mathrm{Var}[\Delta P(t;Q)] = \Delta\Lambda(t)$$

and

$$\mathrm{Var}\left[\int_t^{t+\Delta t} \int_{\mathcal{Q}} \mathcal{P}(\mathbf{ds}, \mathbf{dq})\right] = \int_t^{t+\Delta t} \int_t^{t+\Delta t} \int_{\mathcal{Q}} \int_{\mathcal{Q}} \mathrm{Cov}[\mathcal{P}(\mathbf{ds_1}, \mathbf{dq_1}), \mathcal{P}(\mathbf{ds_2}, \mathbf{dq_2})]\ ,$$

then

$$\mathrm{Cov}[\mathcal{P}(\mathbf{ds_1}, \mathbf{dq_1}), \mathcal{P}(\mathbf{ds_2}, \mathbf{dq_2})] \stackrel{\text{gen}}{=} \lambda(s_1)\delta(s_2 - s_1)ds_1ds_2 \cdot\phi_Q(q_1)\delta(q_1 - q_2)dq_1dq_2\ , \quad (5.16)$$

embodying the independent increment properties in both time and mark arguments of the space-time or mark-time Poisson process in differential form.

- *It is assumed that jump-amplitude function* h *has **finite second order moments**, i.e.,*

$$\int_{\mathcal{Q}} |h(t,q)|^2 \phi_Q(q)dq < \infty\ , \quad (5.17)$$

for all $t \geq 0$*, and in particular,*

$$\int_0^t \int_{\mathcal{Q}} |h(s,q)|^2 \phi_Q(q)dq\lambda(s)ds < \infty\ . \quad (5.18)$$

- *From Theorem 3.12 (p. 72) and Eq. (3.12), a generalization of the standard compound Poisson process is obtained,*

$$\int_0^t \int_{\mathcal{Q}} h(s,q)\mathcal{P}(\mathbf{ds},\mathbf{dq}) = \sum_{k=1}^{P(t;Q)} h(T_k^-, Q_k) \;, \tag{5.19}$$

i.e., the jump-amplitude counting version of the space-time integral, where T_k is the kth jump-time of a Poisson process $P(t;Q)$ and provided comparable assumptions are satisfied. This is also consistent for the infinitesimal counting sum form in (5.6) and the convention (5.7) applies for (5.19). This form is a special case of the filtered compound Poisson process considered in Snyder and Miller [252, Chapter 5]. The form (5.19) is somewhat awkward due to the presence of three random variables, $P(t;Q)$, T_k and Q_k, requiring multiple iterated expectations.

- *For **compound Poisson process with time-independent jump-amplitude**, $h(t,q) = h_2(q)$ (the simplest case being $h(t,q) = q$), then*

$$\Pi_2(t) = \int_0^t \int_{\mathcal{Q}} h_2(q)\mathcal{P}(\mathbf{ds},\mathbf{dq}) = \int_{\mathcal{Q}} h_2(q)\mathcal{P}([0,t),\mathbf{dq}) = \sum_{k=1}^{P(t;Q)} h_2(Q_k), \tag{5.20}$$

where the sum is zero when $P(t;Q) = 0$, the jump-amplitudes $h_2(Q_k)$ form a set of IID random variables independent of the jump-times of the Poisson process $P(t;Q)$, see [55] and Snyder and Miller [252, Chapter 4]. The mean can be computed by double iterated expectations, since the jump-rate is mark-independent,

$$\begin{aligned} \mathrm{E}[\Pi_2(t)] &= \mathrm{E}_{P(t;Q)}\left[\sum_{k=1}^{P(t;Q)} \mathrm{E}_Q[h_2(Q_k)|P(t;Q)]\right] \\ &= \mathrm{E}_{P(t;Q)}\left[P(t;Q)\mathrm{E}_Q[h_2(Q)]\right] = \mathrm{E}_Q[h_2(Q)]\Lambda(t) \;, \end{aligned}$$

where the IID property and more have been used, e.g., $\Lambda(t) = \int_0^t \lambda(s)ds$.

Similarly, the variance is calculated, letting $\overline{h}_2 \equiv \mathrm{E}_Q[h_2(Q)]$,

$$\begin{aligned}
\mathrm{Var}[\Pi_2(t)] &= \mathrm{E}\left[\left(\sum_{k=1}^{P(t;Q)} h_2(Q_k) - \overline{h}_2\Lambda(t)\right)^2\right] \\
&= \mathrm{E}\left[\left(\sum_{k=1}^{P(t;Q)} \left(h_2(Q_k) - \overline{h}_2\right) + \overline{h}_2(P(t;Q) - \Lambda(t))\right)^2\right] \\
&= \mathrm{E}_{P(t;Q)}\Bigg[\sum_{k_1=1}^{P(t;Q)}\sum_{k_2=1}^{P(t;Q)} \mathrm{E}_Q\left[\left(h_2(Q_{k_1}) - \overline{h}_2\right)\left(h_2(Q_{k_2}) - \overline{h}_2\right)\right] \\
&\quad + 2\overline{h}_2(P(t;Q) - \Lambda(t)) \sum_{k=1}^{P(t;Q)} \mathrm{E}_Q\left[h_2(Q_k) - \overline{h}_2\right] + \overline{h}_2^2(P(t;Q) - \Lambda(t))^2\Bigg] \\
&= \mathrm{E}_{P(t;Q)}\big[P(t;Q)\mathrm{Var}_Q[h_2(Q)] + 2\overline{h}_2(P(t;Q) - \Lambda(t))P(t;Q)\cdot 0 \\
&\quad + \overline{h}_2^2(P(t;Q) - \Lambda(t))^2\big] \\
&= \left(\mathrm{Var}_Q[h_2(Q)] + \overline{h}_2^2\right)\Lambda(t) = \mathrm{E}_Q\left[h_2^2(Q)\right]\Lambda(t)\,,
\end{aligned}$$

using the IID property, separation into mean-zero forms and the variance-expectation identity (B.188).

- *For* ***compound Poisson process with both time- and mark-dependence****,* $h(t,q)$ *and* $\lambda(t;q)$*, then*

$$\Pi(t) = \int_0^t \int_{\mathcal{Q}} h(s,q)\mathcal{P}(\mathbf{ds},\mathbf{dq}) = \sum_{k=1}^{P(t;Q)} h(T_k^-, Q_k), \tag{5.21}$$

however, the iterated expectations technique is not too useful for the compound Poisson form, due to the additional dependence introduced by the jump-time, T_k *and the jump-rate* $\lambda(t;q)$*, but the Poisson random measure form is more flexible,*

$$\begin{aligned}
\mathrm{E}[\Pi(t)] &= \mathrm{E}\left[\int_0^t\int_{\mathcal{Q}} h(s,q)\mathcal{P}(\mathbf{ds},\mathbf{dq})\right] = \int_0^t\int_{\mathcal{Q}} \lambda(s,q)h(s,q)\phi_Q(q)dq\,ds \\
&= \int_0^t \mathrm{E}_Q[\lambda(s,Q)h(s,Q)]ds.
\end{aligned}$$

- *Consider the generalization of* ***mean square limits*** *to include mark space integrals. For ease of integration in mean square limits, let the* ***mean-zero Poisson*** *random measure be denoted by*

$$\widetilde{\mathcal{P}}(\mathbf{dt},\mathbf{dq}) \equiv \mathcal{P}(\mathbf{dt},\mathbf{dq}) - \mathrm{E}[\mathcal{P}(\mathbf{dt},\mathbf{dq})] = \mathcal{P}(\mathbf{dt},\mathbf{dq}) - \phi_Q(q)dq\lambda(t)dt \tag{5.22}$$

and corresponding space-time integral be

$$\widetilde{I} \equiv \int_{\mathcal{Q}} h(t,q)\widetilde{\mathcal{P}}(\mathbf{dt},\mathbf{dq})\,. \tag{5.23}$$

Let $\mathcal{T}_n = \{t_i | t_{i+1} = t_i + \Delta t_i \text{ for } i = 0:n, t_0 = 0, t_{n+1} = t, \max_i[\Delta t_i] \to 0 \text{ as } n \to +\infty\}$ *be a proper partition of* $[0,t)$. *Let* $\mathcal{Q}_m = \{\Delta\mathcal{Q}_j \text{ for } j = 1:m | \cup_{j=1}^m \Delta\mathcal{Q}_j = \mathcal{Q}\}$ *be a proper partition of the mark space* $\mathcal{Q}$, *noting that the subsets* $\Delta\mathcal{Q}_j$ *are disjoint is implicit. Let* $h(t,q)$ *be a continuous function in time and marks. Let the corresponding partially discrete approximation*

$$\widetilde{I}_{m,n} \equiv \sum_{i=0}^{n}\sum_{j=1}^{m} h(t_i, q_j^*) \int_{\mathcal{Q}_j} \widetilde{\mathcal{P}}([t_i, t_i + \Delta T), dq_j) \,, \tag{5.24}$$

for some $q_j^* \in \Delta\mathcal{Q}_j$. *Note that if* $\mathcal{Q}$ *is a finite interval* $[a,b]$, *then* $\mathcal{Q}_j = [q_j, q_j + \Delta q]$ *using even spacing with* $q_1 = a$, $q_{m+1} = b$ *and* $\Delta q = (b-a)/m$. *Then* $\widetilde{I}_{m,n}$ *converges in the mean square limit to* $\widetilde{I}$ *if*

$$\mathrm{E}[(\widetilde{I} - \widetilde{I}_{m,n})^2] \to 0 \,, \tag{5.25}$$

as m *and* $n \to +\infty$.

For more advanced and abstract treatments of Poisson random measure, see Gihman and Skorohod [94, Part 2, Chapter 2], Snyder and Miller [252, Chapter 4 and 5], Cont and Tankov [59], Øksendal and Sulem [223] or the applied to abstract bridge Chapter 12.

Theorem 5.2. ***Basic infinitesimal moments of the space-time Poisson process:***

$$\mathrm{E}[d\Pi(t)] = \lambda(t)dt \int_{\mathcal{Q}} h(t,q)\phi_Q(q)dq \equiv \lambda(t)dt\mathrm{E}_Q[h(t,Q)] \equiv \lambda(t)dt\overline{h}(t) \tag{5.26}$$

and

$$\mathrm{Var}[d\Pi(t)] = \lambda(t)dt \int_{\mathcal{Q}} h^2(t,q)\phi_Q(q)dq = \lambda(t)dt\mathrm{E}_Q[h^2(t;Q)] \equiv \lambda(t)dt\overline{h^2}(t). \tag{5.27}$$

Proof. The jump-amplitude function $h(t,Q)$ is independently distributed, through the mark process Q, from the underlying Poisson counting process here, except that this jump in space is conditional on the occurrence of the jump-time or count of the underlying Poisson process. However, the function $h(t,q)$ is deterministic since it depends on the realization q in the space-time Poisson definition, rather than the random variable Q. The infinitesimal mean (5.26) is straightforward,

$$\begin{aligned}\mathrm{E}[d\Pi(t)] &= \mathrm{E}\left[\int_{\mathcal{Q}} h(t,q)\mathcal{P}(\mathbf{dt},\mathbf{dq})\right] = \int_{\mathcal{Q}} h(t,q)\mathrm{E}[\mathcal{P}(\mathbf{dt},\mathbf{dq})] \\ &= \lambda(t)dt \int_{\mathcal{Q}} h(t,q)\phi_Q(q)dq = \lambda(t)dt\mathrm{E}_Q[h(t,Q)] \equiv \lambda(t)dt\overline{h}(t)\end{aligned}$$

noting that the expectation operator applied to the mark integral can be moved to apply just to the Poisson random measure $\mathcal{P}(\mathbf{dt},\mathbf{dq})$.

However, the result for the variance in (5.27) is not so obvious, but the covariance formula for two Poisson random measures with differing mark variables $\mathrm{Cov}[\mathcal{P}(\mathbf{dt},\mathbf{dq}_1),\mathcal{P}(\mathbf{dt},\mathbf{dq}_2)]$ (5.15) will be useful by converting to the mean-zero Poisson random measure $\widetilde{\mathcal{P}}(\mathbf{dt},\mathbf{dq})$ in (5.22),

$$\begin{aligned}
\mathrm{Var}[d\Pi(t)] &= \mathrm{E}\left[\left(\int_{\mathcal{Q}} h(t,q)\mathcal{P}(\mathbf{dt},\mathbf{dq}) - \overline{h}(t)\lambda(t)dt\right)^2\right] \\
&= \mathrm{E}\left[\left(\int_{\mathcal{Q}} (h(t,q)\mathcal{P}(\mathbf{dt},\mathbf{dq}) - h(t,q)\phi_Q(q)\lambda(t)dt)\right)^2\right] \\
&= \mathrm{E}\left[\left(\int_{\mathcal{Q}} h(t,q)\widetilde{\mathcal{P}}(\mathbf{dt},\mathbf{dq})\right)^2\right] \\
&= \mathrm{E}\left[\int_{\mathcal{Q}} h(t,q_1)\int_{\mathcal{Q}} h(t,q_2)\widetilde{\mathcal{P}}(\mathbf{dt},\mathbf{dq}_1)\widetilde{\mathcal{P}}(\mathbf{dt},\mathbf{dq}_1)\right] \\
&= \int_{\mathcal{Q}} h(t,q_1)\int_{\mathcal{Q}} h(t,q_2)\mathrm{Cov}\left[\widetilde{\mathcal{P}}(\mathbf{dt},\mathbf{dq}_1),\widetilde{\mathcal{P}}(\mathbf{dt},\mathbf{dq}_1)\right] \\
&= \lambda(t)dt\int_{\mathcal{Q}} h^2(t,q_1)\phi_Q(q_1)dq_1 = \lambda(t)dt\mathrm{E}_Q\left[h^2(t,Q)\right] \equiv \lambda(t)dt\overline{h^2}(t)\ .
\end{aligned}$$

□

Examples 5.3.

- ***Uniformly Distributed Jump Amplitudes:***
  As an example of a continuous distribution, consider the uniform density for the jump-amplitude mark Q be given by

$$\phi_Q(q) = \frac{1}{b-a}U(q;a,b),\ a<b\ , \tag{5.28}$$

  where $U(q;a,b) = \mathbf{1}_{q\in[a,b]}$ is the step or indicator function for the interval $[a,b]$, i.e., $U(q;a,b)$ is one when $a \le q \le b$ and zero otherwise. The first few moments are

$$\begin{aligned}
\mathrm{E}_Q[1] &= \frac{1}{b-a}\int_a^b dq = 1\ , \\
\mathrm{E}_Q[Q] &= \frac{1}{b-a}\int_a^b qdq = \frac{b+a}{2}\ , \\
\mathrm{Var}_Q[Q] &= \frac{1}{b-a}\int_a^b (q-(b+a)/2)^2dq = \frac{(b-a)^2}{12}\ .
\end{aligned}$$

  In the case of the log-uniform amplitude letting $Q = \ln(1+h(Q))$ be the mark-amplitude relationship using the log-transformation form from the linear SDE problem (4.76), then

$$h(Q) = e^Q - 1$$

and the expected jump-amplitude is

$$\mathrm{E}_Q[h(Q)] = \frac{1}{b-a}\int_a^b (e^q - 1)dq = \frac{e^b - e^a}{b-a} - 1 \ .$$

- ***Poisson Distributed Jump Amplitudes:***
  As an example of a discrete distribution of jump-amplitudes, consider

$$\Phi_Q(k) = p_k(u) = e^{-u}\frac{u^k}{k!} \ ,$$

for $k = 0 : \infty$. Thus, the jump process is a Poisson-Poisson process or a Poisson-mark Poisson process. The mean and variance are

$$\mathrm{E}_Q[Q] = u \ ,$$
$$\mathrm{Var}_Q[Q] = u \ .$$

Remark 5.4. *For the general discrete distribution,*

$$\Phi_Q(k) = p_k \ , \quad \sum_{k=0}^{\infty} p_k = 1 \ ,$$

the comparable continuized form is

$$\Phi_Q(q) \stackrel{\text{gen}}{=} \sum_{k=0}^{\infty} H_R(q-k)p_k = \sum_{k=0}^{\lfloor q \rfloor} p_k \ ,$$

where $H_R(q)$ is again the right-continuous Heaviside step function and $\lfloor q \rfloor$ is the maximum integer not exceeding q. The corresponding generalized density is

$$\phi_Q(q) \stackrel{\text{gen}}{=} \sum_{k=0}^{\infty} \delta_R(q-k)p_k \ .$$

The reader should verify that this density yields the correct expectation and variance forms.

5.2 State-Dependent Generalization of Jump-Diffusion SDEs

5.2.1 State-Dependent Generalization for Space-Time Poisson Processes

The space-time Poisson process is generalized to include state-dependence with $X(t)$ in both the jump-amplitude and the Poisson measure, such that

$$d\Pi(t; X(t), t) = \int_{\mathcal{Q}} h(X(t), t, q)\mathcal{P}(\mathbf{dt}, \mathbf{dq}; X(t), t) \tag{5.29}$$

on the Poisson mark space $\mathcal{Q}$ with Poisson random measure $\mathcal{P}(\mathbf{dt}, \mathbf{dq}; X(t), t)$, which helps to describe the space-time Poisson mechanism and related calculus. The space-time state-dependent Poisson mark, $Q = q$ is again the underlying random variable for the state-dependent and mark-dependent jump-amplitude coefficient $h(x, t, q)$. The double time t arguments of $d\Pi$, dP and $\mathcal{P}$ are not considered redundant for applications, since the first time t or time set dt is the usual Poisson jump process implicit time dependence, while the second to the right of the semi-colon denotes explicit or parametric time dependence paired with explicit state dependence that is known in advance and is appropriate for the application model.

Alternatively, the Poisson zero-one law form may be used, i.e.,

$$d\Pi(t; X(t), t) \stackrel{dt}{=}_{zol} h(X(t), t, Q)dP(t; Q, X(t), t) \, , \tag{5.30}$$

with the jump of $\Pi(t; X(t), t)$ being

$$[\Pi](T_k) = h(X(T_k^-), T_k^-, Q_k)$$

at jump-time T_k and jump-mark Q_k. The multitude of random variables in this sum means that expectations or other Poisson integrals will be very difficult to calculate even by conditional expectation iterations.

Definition 5.5. *The **conditional expectation of** $\mathcal{P}(\mathbf{dt}, \mathbf{dq}; X(t), t)$ is*

$$\mathrm{E}[\mathcal{P}(\mathbf{dt}, \mathbf{dq}; X(t), t)|X(t) = x] = \phi_Q(q; x, t)dq\lambda(t; x, t)dt \, , \tag{5.31}$$

where $\phi_Q(q; x, t)dq$ is the probability density of the now state-dependent Poisson amplitude mark and the jump rate $\lambda(t; x, t)$ now has state-time dependence. In this notation, the relationship to the simple counting process is given by

$$\int_{\mathcal{Q}} \mathcal{P}(\mathbf{dt}, \mathbf{dq}; X(t), t) = dP(t; Q, X(t), t) \, .$$

Hence, when $h(x, t, q) = \widetilde{h}(x, t)$, i.e., independent of the mark q, the space-time Poisson is the simple jump process with mark-independent amplitude,

$$d\Pi(t; X(t), t) = \widetilde{h}(X(t), t)dP(t; Q, X(t), t) \, ,$$

but with non-unit jumps in general. Effectively the same form is obtained when there is a single discrete mark, e.g., $\phi_Q(q) = \delta(q-1)$, so $h(x, t, q) = h(x, t, 1)$ always.

Theorem 5.6. ***Basic conditional infinitesimal moments of the state-dependent Poisson process:***

$$\begin{aligned} \mathrm{E}[d\Pi(t; X(t), t)|X(t) = x] &= \int_{\mathcal{Q}} h(x, t, q)\phi_Q(q; x, t)dq\lambda(t; x, t)dt \\ &\equiv \mathrm{E}_Q[h(x, t, Q)]\lambda(t; x, t)dt \end{aligned} \tag{5.32}$$

and

$$\begin{aligned}\mathrm{Var}[d\Pi(t;X(t),t)|X(t)=x] &= \int_{\mathcal{Q}} h^2(x,t,q)\phi_Q(q;x,t)dq\lambda(t;x,t)dt \\ &\equiv \mathrm{E}_Q[h^2(x,t;Q)]\lambda(t;x,t)dt\ . \end{aligned} \tag{5.33}$$

Proof. The justification is the same justification as for Eqs. (5.27-5.27). It is assumed that the jump-amplitude $h(x,t,Q)$ is independently distributed due to Q from the underlying Poisson counting process here, except that this jump in space is conditional on the occurrence of the jump-time of the underlying Poisson process. □

5.2.2 State-Dependent Jump-Diffusion SDEs

The general, scalar stochastic differential equation (SDE) takes the form

$$\begin{aligned} dX(t) = &\ f(X(t),t)dt + g(X(t),t)dW(t) + \int_{\mathcal{Q}} h(X(t),t,q)\mathcal{P}(\mathbf{dt},\mathbf{dq};X(t),t) \\ \overset{\mathrm{dt}}{=} &\ f(X(t),t)dt + g(X(t),t)dW(t) + h(X(t),t,Q)dP(t;Q,X(t),t)\ , \end{aligned} \tag{5.34}$$

for the state process $X(t)$ with a set of continuous coefficient functions $\{f,g,h\}$. However, the SDE model is just a useful symbolic model for many applied situations, but the more basic model relies on specifying the method of integration. So

$$\begin{aligned} X(t) = &\ X(t_0) + \int_{t_0}^{t} (f(X(s),s)ds + g(X(s),s)dW(s) \\ &\ + h(X(t),s,Q)dP(s;Q,X(s),s)) \\ \overset{ims}{=} &\ X(t_0) + \operatorname*{ms\,lim}_{n\to\infty}\left[\sum_{i=0}^{n}\left(f_i\Delta t_i + g_i\Delta W_i + \sum_{k=0}^{\Delta P_i} h_{i,k}\right)\right]\ , \end{aligned} \tag{5.35}$$

where $f_i = f(X_i,t_i)$, $g_i = g(X_i,t_i)$, $h_{i,k} = h(X_i,T_k,Q_k)$, $\Delta t_i = t_{i+1} - t_i$, $\Delta P_{i,k} = \Delta P(t_i;Q,X_i,t_i)$ and $\Delta W_i = \Delta W(t_i)$. Here, T_k is the kth jump-time and $\{Q,Q_k\}$ are the corresponding random marks.

The **conditional infinitesimal moments for the state process** are

$$\mathrm{E}[dX(t)|X(t)=x] = f(x,t)dt + \overline{h}(x,t)\lambda(t;x,t)dt\ , \tag{5.36}$$

$$\overline{h}(x,t)\lambda(t;x,t)dt \equiv \mathrm{E}_Q[h(x,t,Q)]\lambda(t;x,t)dt\ , \tag{5.37}$$

and

$$\mathrm{Var}[dX(t)|X(t)=x] = g^2(x,t)dt + \overline{h^2}(x,t)\lambda(t;x,t)dt\ , \tag{5.38}$$

$$\overline{h^2}(x,t)\lambda(t;x,t)dt \equiv \mathrm{E}_Q[h^2(x,t,Q)]\lambda(t;x,t)dt\ , \tag{5.39}$$

using (1.1,5.32,5.33,5.34) and assuming that the Poisson process is independent of the Wiener process. The jump in the state at jump time T_k in the underlying Poisson process is

$$[X](T_k) \equiv X(T_k^+) - X(T_k^-) = h(X(T_k^-), T_k^-, Q_k) \ , \tag{5.40}$$

for $k = 1, 2, \ldots$, now depending on the kth mark Q_k at the pre-jump-time T_k^- at the kth jump.

Rule 5.7. ***Stochastic Chain Rule for State-Dependent SDEs:***
The ***stochastic chain rule*** *for a sufficiently differentiable function $Y(t) = F(X(t), t)$ has the form*

$$\begin{aligned} dY(t) &= dF(X(t),t) \stackrel{\text{sym}}{=} F(X(t) + dX(t), t + dt) - F(X(t), t) \\ &= d_{(cont)}F(X(t),t) + d_{(jump)}F(X(t),t) \\ &\stackrel{\text{dt}}{=} F_t(X(t),t)dt + F_x(X(t),t)(f(X(t),t)dt + g(X(t),t)dW(t)) \\ &\quad + \frac{1}{2}F_{xx}(X(t),t)g^2(X(t),t)dt \\ &\quad + \int_{\mathcal{Q}} (F(X(t) + h(X(t),t,q),t) - F(X(t),t))\mathcal{P}(\mathbf{dt}, \mathbf{dq}; X(t), t) \ , \end{aligned} \tag{5.41}$$

to precision-dt. It is sufficient that F be twice continuously differentiable in x and once in t.

5.2.3 Linear State-Dependent SDEs

Let the state-dependent jump-diffusion process satisfy a SDE linear in the state process $X(t)$ with time-dependent rate coefficients,

$$dX(t) \stackrel{\text{dt}}{=} X(t)\left(\mu_d(t)dt + \sigma_d(t)dW(t) + \nu(t,Q)dP(t;Q)\right) \ , \tag{5.42}$$

for $t > t_0$, with $X(t_0) = X_0$ and $\mathrm{E}[dP(t;Q)] = \lambda(t)dt$, where $\mu_d(t)$ denotes the mean and $\sigma_d^2(t)$ denotes the variance of the diffusion process, while Q_k denotes the kth mark and T_k denotes the kth time of the jump process.

Again, using the log-transformation $Y(t) = \ln(X(t))$ and the stochastic chain rule (5.41),

$$dY(t) \stackrel{\text{dt}}{=} (\mu_d(t) - \sigma_d^2(t)/2)dt + \sigma_d(t)dW(t) + \ln\left(1 + \nu(t,Q)\, dP(t;Q)\right), \tag{5.43}$$

with immediate integrals

$$Y(t) = \ln(x_0) + \int_{t_0}^{t} dY(s) \tag{5.44}$$

and

$$X(t) = x_0 \exp\left(\int_{t_0}^{t} dY(s)\right) \ , \tag{5.45}$$

or in recursive form,

$$X(t+\Delta t) = X(t)\exp\left(\int_t^{t+\Delta t} dY(s)\right) . \tag{5.46}$$

Linear Mark-Jump-Diffusion Simulation Forms

For simulations, a small time-step, $\Delta t_i \ll 1$, approximation of the recursive form (5.46) would be more useful, with $X_i = X(t_i)$, $\mu_i = \mu_d(t_i)$, $\sigma_i = \sigma_d(t_i)$, $\Delta W_i = \Delta W(t_i)$, $\Delta P_i = \Delta P(t_i; Q)$ and the convenient jump-amplitude coefficient approximaton, $\nu(t,Q) \simeq \nu_0(Q) \equiv \exp(Q) - 1$, i.e.,

$$X_{i+1} \simeq X_i \exp\left((\mu_i - \sigma_i^2/2)\Delta t_i + \sigma_i \Delta W_i\right)(1+\nu_0(Q))^{\Delta P_i} , \tag{5.47}$$

for $i = 1 : N$ time-steps, where a zero-one jump law approximation has been used.

For the diffusion part, it has been shown that

$$\mathrm{E}\left[e^{\sigma_i \Delta W_i}\right] = e^{\sigma_i^2 \Delta t_i/2},$$

using the completing the square technique. In addition, there is the following lemma for the jump part of (5.47):

Lemma 5.8. ***Jump Term Expectation***

$$\mathrm{E}\Big[(1+\nu_0(Q))^{\Delta P_i}\Big] = e^{\lambda_i \Delta t_i \mathrm{E}[\nu_0(Q)]}, \tag{5.48}$$

where $\mathrm{E}[\Delta P_i] = \lambda_i \Delta t_i$ *and* $\nu_0(Q) = \exp(Q) - 1$.

Proof. Using given forms, iterated expectations, the Poisson distribution and the IID property of the marks Q_k, then

$$\begin{aligned}
\mathrm{E}\left[(1+\nu_0(Q))^{\Delta P_i}\right] &= \mathrm{E}\left[e^{Q\Delta P_i}\right] \\
&= e^{-\lambda_i \Delta t_i} \sum_{k=0}^{\infty} (\lambda_i \Delta t_i)^k \mathrm{E}_Q\left[e^{kQ}\right] \\
&= e^{-\lambda_i \Delta t_i} \sum_{k=0}^{\infty} (\lambda_i \Delta t_i)^k \left(\mathrm{E}_Q\left[e^{Q}\right]\right)^k \\
&= e^{-\lambda_i \Delta t_i} e^{\lambda_i \Delta t_i \mathrm{E}_Q\left[e^{Q}\right]} \\
&= e^{\lambda_i \Delta t_i \mathrm{E}_Q[\nu_0(Q)]}.
\end{aligned}$$

□ An immediate consequence of this result is the following corollary:

Corollary 5.9. ***Discrete State Expectations:***

$$\mathrm{E}[X_{i+1}|X_i] \simeq X_i \exp((\mu_i + \lambda_i \mathrm{E}_Q[\nu_0(Q)])\Delta t_i) \tag{5.49}$$

and

$$\mathrm{E}[X_{i+1}] \simeq x_0 \exp\left(\sum_{j=0}^{i}(\mu_j + \lambda_j \mathrm{E}_Q[\nu_0(Q)])\Delta t_j\right). \tag{5.50}$$

Further, as Δt_i and $\delta t_n \to 0^+$, the continuous form of the expectation follows and is given later in Corollary 5.13 on page 148 using other justification.

Example 5.10. ***Linear, Time-Independent, Constant-Rate Coefficient Case:***
In the linear, time-independent rate-coefficient case with $\mu_d(t) = \mu_0$, $\sigma_d(t) = \sigma_0$, $\lambda(t) = \lambda_0$ and $\nu(t, Q) = \nu_0(Q) = e^Q - 1$,

$$X(t) = x_0 \exp\left((\mu_0 - \sigma_0^2/2)(t - t_0) + \sigma_0(W(t) - W(t_0)) + \sum_{k=1}^{P(t;Q)-P(t_0;Q)} \nu_0 Q_k\right), \tag{5.51}$$

where the Poisson counting sum form is now more manageable since the marks do not depend on the pre-jump-times T_k^-.

Using the independence of the three underlying stochastic processes, $(W(t) - W(t_0))$, $(P(t;Q) - P(t_0;Q))$ and Q_i, as well as the stationarity of the first two and the law of exponential to separate exponentials, leads to partial reduction of the expected state process:

$$\begin{aligned}
\mathrm{E}[X(t)] &= x_0 e^{(\mu_0 - \sigma_0^2/2)(t-t_0)} \cdot \mathrm{E}_W\left[e^{\sigma_0 W(t-t_0)}\right] \cdot \sum_{k=0}^{\infty} \mathrm{E}[P(t - t_0; Q) = k] \mathrm{E}\left[e^{\sum_{\ell=1}^{k} Q_\ell}\right] \\
&= x_0 e^{(\mu_0 - \sigma_0^2/2)(t-t_0)} \int_{-\infty}^{+\infty} \frac{e^{-w^2/(2(t-t_0))}}{\sqrt{2\pi(t-t_0)}} e^{\sigma_0 w} dw \\
&\quad \cdot e^{-\lambda_0(t-t_0)} \sum_{k=0}^{\infty} \frac{(\lambda_0(t-t_0))^k}{k!} \prod_{i=1}^{k} \mathrm{E}_Q\left[e^Q\right] \\
&= x_0 e^{\mu_0(t-t_0)} e^{-\lambda_0(t-t_0)} \sum_{k=0}^{\infty} \frac{(\lambda_0(t-t_0))^k}{k!} \mathrm{E}_Q^k\left[e^Q\right] \\
&= x_0 e^{\left(\mu_0 + \lambda_0\left(\mathrm{E}_Q\left[e^Q\right] - 1\right)\right)(t-t_0)}
\end{aligned} \tag{5.52}$$

where $\lambda_0(t - t_0)$ is the Poisson parameter and $\mathcal{Q} = (-\infty, +\infty)$ is taken as the mark space for specificity with

$$\mathrm{E}_Q\left[e^Q\right] = \int_{\mathcal{Q}} e^q \phi_Q(q) dq \,.$$

Little more useful simplification can be obtained analytically, except for infinite expansions or equivalent special functions, when the mark density $\phi_Q(q)$ is specified. Numerical procedures may be more useful for practical purposes. The state expectation in this distributed mark case (5.52) should be compared to pure constant linear coefficient case (4.81) of Chapter 4.

Exponential Expectations:

Sometimes it is necessary to get the expectation of an exponential of the integral of a jump-diffusion process. The procedure is much more complicated for distributed amplitude Poisson jump processes than for diffusions since the mark-time process is a product process, the product of the mark process and the Poisson process. For the time-independent coefficient case, as in a prior example, the exponential processes are easily separable by the law of exponents. However, for the time-dependent case, it is necessary to return to use the space-time process $\mathcal{P}$ and the decomposition approximation used in the mean square limit. The h in the following theorem might be the amplitude coefficient in (5.43) or $h(s,q) = q = \ln(1+\nu(s,q))$, for instance.

Theorem 5.11. ***Expectation for the Exponential of Space-Time Counting Integrals:***
Assuming finite second order moments for $h(t,q)$ and convergence in the mean square limit,

$$\mathrm{E}\left[\exp\left(\int_{t_0}^{t}\int_{\mathcal{Q}} h(s,q)\mathcal{P}(\mathbf{ds},\mathbf{dq})\right)\right] = \exp\left(\int_{t_0}^{t}\int_{\mathcal{Q}}\left(e^{h(s,q)}-1\right)\phi_Q(q,s)dq\lambda(s)ds\right)$$
$$\equiv \exp\left(\int_{t_0}^{t}\overline{(e^h-1)}(s)\lambda(s)ds\right) . \qquad (5.53)$$

Proof. Let the proper partition of the mark space over disjoint subsets be

$$\mathcal{Q}_m = \{\Delta\mathcal{Q}_j \text{ for } j = 1{:}m | \cup_{j=1}^{m}\Delta\mathcal{Q}_j = \mathcal{Q}\}$$

Since Poisson measure is Poisson distributed,

$$\Phi_{\mathcal{P}_j}(k) = \mathrm{Prob}[\mathcal{P}(\mathbf{dt},\Delta\mathcal{Q}_j) = k] = e^{-\overline{\mathcal{P}_j}}\frac{(\overline{\mathcal{P}_j})^k}{k!}$$

with Poisson parameter

$$\overline{\mathcal{P}_j} \equiv \mathrm{E}[\mathcal{P}(\mathbf{dt},\Delta\mathcal{Q}_j)] = \lambda(t)dt\Phi_Q(\Delta\mathcal{Q}_j,t_i) .$$

for each subset $\{\Delta\mathcal{Q}_j\}$.

Similarly, let the proper partition over the time interval be

$$\mathcal{T}_n = \{t_i | t_{i+1} = t_i + \Delta t_i \text{ for } i = 0{:}n, t_0 = 0, t_{n+1} = t, \max_i[\Delta t_i] \to 0 \text{ as } n \to +\infty\} .$$

The disjoint property over subsets and time intervals means $\mathcal{P}([t_i,t_i+\Delta t_i),\Delta\mathcal{Q}_j)$ and $\mathcal{P}([t_i,t_i+\Delta t_i),\Delta\mathcal{Q}'_j)$ will be pairwise independent provided $j' \neq j$ for fixed i corresponding to the Property (5.15) for infinitesimals, while the $\mathcal{P}([t_i,t_i+\Delta t_i),\Delta\mathcal{Q}_j)$ and $\mathcal{P}([t_i,t_i+\Delta t'_i),\Delta\mathcal{Q}'_j)$ will be pairwise independent provided $i' \neq i$ and $j' \neq j$, corresponding to the Property (5.16) for infinitesimals.

For brevity, let $h_{i,j} \equiv h(t_i, q_j^*)$ where $q_j^* \in \Delta\mathcal{Q}_j$, $\mathcal{P}_{i,j} \equiv \mathcal{P}_i([t_i, t_i + \Delta t_i), \Delta\mathcal{Q}_j)$ and $\overline{\mathcal{P}}_{i,j} \equiv \lambda_i \Delta t_i \Phi_Q(\Delta\mathcal{Q}_j)$.

Using mean square limits, $\mathcal{P}_{i,j}$ playing the dual roles of the two increments $(\Delta t_i, \Delta\mathcal{Q}_j)$, the law of exponents and independence (symbol $\overset{ind}{=}_{inc}$),

$$\begin{aligned}
\mathrm{E}\left[\exp\left(\int_{t_0}^{t}\int_{\mathcal{Q}} h\mathcal{P}\right)\right] &\overset{ims}{=} \operatorname*{ms\,lim}_{m,n\to\infty} \mathrm{E}\left[\exp\left(\sum_{i=0}^{n}\sum_{j=1}^{m} h_{i,j}\mathcal{P}_{i,j}\right)\right] \\
&\overset{ind}{=}_{inc} \operatorname*{ms\,lim}_{m,n\to\infty} \Pi_{i=0}^{n}\Pi_{j=1}^{m}\mathrm{E}\left[\exp\left(h_{i,j}\mathcal{P}_{i,j}\right)\right] \\
&= \operatorname*{ms\,lim}_{m,n\to\infty} \Pi_{i=0}^{n}\Pi_{j=1}^{m}\exp\left(-\overline{\mathcal{P}_{i,j}}\right)\sum_{k_{i,j}=0}^{\infty}\frac{\overline{\mathcal{P}_{i,j}}^{k_{i,j}}}{k_{i,j}!}\exp\left(h_{i,j}k_{i,j}\right) \\
&= \operatorname*{ms\,lim}_{m,n\to\infty} \Pi_{i=0}^{n}\Pi_{j=1}^{m}\exp\left(\overline{\mathcal{P}_{i,j}}\left(\exp(h_{i,j})-1\right)\right) \\
&= \operatorname*{ms\,lim}_{m,n\to\infty} \exp\left(\sum_{i=0}^{n}\sum_{j=1}^{m}\left(\exp(h_{i,j})-1\right)\lambda_i\Delta t_i\Phi_Q(\Delta Q_i, t_i)\right) \\
&\overset{ims}{=} \exp\left(\int_{t_0}^{t}\int_{\mathcal{Q}}\left(\exp(h(s,q))-1\right)\phi_Q(q,s)dq\lambda(s)ds\right) \\
&\equiv \exp\left(\int_{t_0}^{t}\overline{\left(\exp(h(s,Q))-1\right)}\lambda(s)ds\right) .
\end{aligned}$$

Thus, the main technique is to unassemble the mean square limit discrete approximation to get at the independent random part, take its expectation and then reassemble the mean square limit back again, justifying the interchange of expectation and exponentiation-integration. □

Remarks 5.12.

- *Note that the mark space subset $\Delta\mathcal{Q}_j$ is never used directly as a discrete element of integration, since the subset would be infinite if the mark space were infinite. The mark space element is only used through the distribution which would be bounded. This is quite unlike the time domain where we can select t to be finite. If the mark space were finite, say $\mathcal{Q} = [a, b]$, then a concrete partition of $[a, b]$ similar to the time-partition can be used.*

- *Also note that the dependence on $(X(t), t)$ was not used, but could be considered suppressed but absorbed into the existing t dependence of h and $\mathcal{P}$.*

Corollary 5.13. ***Expectation of X(t) for Linear SDE:***

Let $X(t)$ be the solution (5.45) with $\overline{\nu}(t) \equiv \mathrm{E}[\nu(t,Q)]$ of (5.42), then

$$\mathrm{E}[X(t)] = x_0 \exp\left(\int_{t_0}^{t} (\mu_d(s) + \lambda(s)\overline{\nu}(s))\, ds\right) \tag{5.54}$$

$$= x_0 \exp\left(\int_{t_0}^{t} \mathrm{E}[dX(s)/X(s)] ds\right) . \tag{5.55}$$

Proof. The jump part, the main part, follows from exponential Theorem 5.11 Eq. (5.53) and the lesser part for the diffusion is left as an exercise for the reader.

However, note that the exponent is the time integral of $\mathrm{E}[dX(t)/X(t)]$, the relative conditional infinitesimal mean, which is independent of $X(s)$ and is valid only for the linear mark-jump-diffusion SDE. □

Remark 5.14. *The relationship in (5.55) is a **quasi-deterministic equivalence** for linear mark-jump-diffusion SDEs and was shown by Hanson and Ryan [114] in 1989. They also produced a nonlinear jump counter example that has a formal closed form solution in terms of the gamma function, for which the result does not hold and a very similar example is given in Exercise 9 in Chapter 4.*

Moments of Log-Jump-Diffusion Process:

For the log-jump-diffusion process $dY(t)$ in (5.43), suppose that the jump-amplitude is time-independent and that the mark variable was conveniently chosen as

$$Q = \ln(1 + \nu(t,Q))$$

so that the SDE has the form

$$dY(t) \stackrel{\mathrm{dt}}{=} \mu_{ld}(t)dt + \sigma_d(t)dW(t) + QdP(t;Q) , \tag{5.56}$$

or in the case of applications for which the time step Δt is an increment that is not infinitesimal like dt there is some probability of more than one jump,

$$\Delta Y(t) = \mu_{ld}(t)\Delta t + \sigma_d(t)\Delta W(t) + \sum_{k=1}^{\Delta P(t;Q)} Q_k . \tag{5.57}$$

The results for the infinitesimal case (5.56) are contained in the incremental case (5.57).

The first few moments can found in general for (5.57), and if up to the fourth moment, then the skew and kurtosis coefficients can be calculated. These calculations can be expedited by the following lemma concerning sums of zero-mean IID random variables:

Lemma 5.15. ***Zero-mean IID random variable sums:***
Let $\{X_i | i = 1{:}n\}$ be a set of zero-mean IID random variables, i.e., $\mathrm{E}[X_i] = 0$. Let

$M^{(m)} \equiv E[X_i^m]$ *be the* m*th moment and*

$$S_n^{(m)} \equiv \sum_{i=1}^{n} X_i^m,$$

with $S_n^{(1)} = S_n$ *being the usual partial sum over the set and*

$$\mathrm{E}[S_n^{(m)}] = nM^{(m)} \,, \tag{5.58}$$

then the expectation of powers of S_n *for* $m = 1{:}4$ *are*

$$\mathrm{E}\left[(S_n)^m\right] = \left\{ \begin{array}{ll} 0, & m = 1 \\ nM^{(2)}, & m = 2 \\ nM^{(3)}, & m = 3 \\ nM^{(4)} + 3n(n-1)\left(M^{(2)}\right)^2, & m = 4 \end{array} \right\}. \tag{5.59}$$

Proof. First by the linear property of the expectation and the IID properties of the X_i,

$$\mathrm{E}\left[S_n^{(m)}\right] = \sum_{i=1}^{n} \mathrm{E}[X_i^m] = \sum_{i=1}^{n} M^{(m)} = nM^{(m)} \,. \tag{5.60}$$

The $m = 1$ case is trivial due to the zero mean property of the X_i's and the linearity of the expectation operator, $\mathrm{E}[S_n] = \sum_{i=1}^{n} \mathrm{E}[X_i] = 0$.

For $m = 2$, the induction hypothesis from (5.59) is

$$\mathrm{E}\left[S_n^2\right] \equiv \mathrm{E}\left[\left(\sum_{i=1}^{n} X_i^2\right)\right] = nM^{(2)}$$

with initial condition at $n = 1$ is $\mathrm{E}[S_1^2] = \mathrm{E}[X_1^2] = M^{(2)}$ by definition. The hypothesis can be easily proved by partial sum recursion $S_{n+1} = S_n + X_{n+1}$, application of the binomial theorem, expectation linearity and the zero-mean IID property:

$$\begin{aligned} \mathrm{E}\left[S_{n+1}^2\right] &= \mathrm{E}\left[(S_n + X_{n+1})^2\right] = \mathrm{E}\left[S_n^2 + 2X_{n+1}S_n + X_{n+1}^2\right] \\ &= nM^{(2)} + 2 \cdot 0 \cdot 0 + M^{(2)} = (n+1)M^{(2)} \,. \end{aligned} \tag{5.61}$$

QED for $m = 2$.

Similarly for the power $m = 3$, again beginning with the induction hypothesis

$$\mathrm{E}\left[S_n^3\right] \equiv \mathrm{E}\left[\left(\sum_{i=1}^{n} X_i\right)^3\right] = nM^{(3)}$$

with initial condition at $n = 1$ is $\mathrm{E}[S_1^3] = \mathrm{E}[X_1^3] = M^{(3)}$ by definition. Using the same techniques as in (5.61),

$$\begin{aligned} \mathrm{E}\left[S_{n+1}^3\right] &= \mathrm{E}\left[(S_n + X_{n+1})^3\right] = \mathrm{E}\left[S_n^3 + 3X_{n+1}S_n^2 + 3X_{n+1}^2S_n^2 + X_{n+1}^3\right] \\ &= nM^{(3)} + 3 \cdot 0 \cdot nM^{(2)} + 3 \cdot M^{(2)} \cdot 0 + M^{(3)} = (n+1)M^{(3)} \,. \end{aligned} \tag{5.62}$$

QED for $m = 3$.

Finally, the case for the power $m = 4$ is a little different since an additional nontrivial term arises from the product of the squares of two independent variables. The induction hypothesis is

$$\mathrm{E}\left[S_n^4\right] \equiv \mathrm{E}\left[\left(\sum_{i=1}^{n} X_i\right)^4\right] = nM^{(4)} + 3n(n-1)(M^{(2)})^2$$

with initial condition at $n = 1$ is $\mathrm{E}[S_1^4] = \mathrm{E}[X_1^4] = M^{(4)}$ by definition. Using the same techniques as in (5.61),

$$\begin{aligned}
\mathrm{E}\left[S_{n+1}^4\right] &= \mathrm{E}\left[(S_n + X_{n+1})^4\right] = \mathrm{E}\left[S_n^4 + 4X_{n+1}S_n^3 + 6X_{n+1}^2S_n^2 + 4X_{n+1}^3S_n^1 + X_{n+1}^4\right] \\
&= nM^{(4)} + 3n(n-1)(M^{(2)})^2 + 4 \cdot 0 \cdot nM^{(3)} + 6 \cdot M^{(2)} \cdot nM(2) \\
&\quad +4 \cdot M^{(3)} \cdot 0 + M^{(4)} \\
&= (n+1)M^{(4)} + 3(n+1)((n+1)-1)(M^{(2)})^2 \, . \qquad (5.63)
\end{aligned}$$

QED for $m = 4$. □

Remark 5.16. *The results here depend on the IID and zero-mean properties, but do not otherwise depend on the particular distribution of the random variables. The results are used in the following theorem:*

Theorem 5.17. ***Some Moments of the Log-jump-diffusion (ljd) Process*** $\Delta Y(t)$***:***
Let $\Delta Y(t)$ satisfy the stochastic difference equation (5.57), the marks Q_k be IID with mean $\mu_j \equiv \mathrm{E}_Q[Q_k]$ and variance $\sigma_j^2 \equiv \mathrm{Var}_Q[Q_k]$, then the first four moments, $m = 1{:}4$, are

$$\mu_{ljd}(t) \equiv \mathrm{E}[\Delta Y(t)] = (\mu_{ld}(t) + \lambda(t)\mu_j)\Delta t \, ; \qquad (5.64)$$

$$\sigma_{ljd}(t) \equiv \mathrm{Var}[\Delta Y(t)] = \left(\sigma_d^2(t) + \left(\sigma_j^2 + \mu_j^2\right)\lambda(t)\right)\Delta t \, ; \qquad (5.65)$$

$$M_{ljd}^{(3)}(t) \equiv \mathrm{E}\left[(\Delta Y(t) - \mathrm{E}[\Delta Y(t)])^3\right] = \left(M_j^{(3)} + \mu_j\left(3\sigma_j^2 + \mu_j^2\right)\right)\lambda(t)\Delta t \, , \qquad (5.66)$$

where $M_j^{(3)} \equiv \mathrm{E}_Q[(Q_i - \mu_j)^3]$;

$$\begin{aligned}
M_{ljd}^{(4)}(t) &\equiv \mathrm{E}\left[(\Delta Y(t) - \mathrm{E}[\Delta Y(t)])^4\right] \\
&= \left(M_j^{(4)} + 4\mu_j M_j^{(3)} + 6\mu_j^2\sigma_j^2 + \mu_j^4\right)\lambda(t)\Delta t \\
&\quad +3\left(\sigma_d^2(t) + \left(\sigma_j^2 + \mu_j^2\right)\lambda(t)\right)^2(\Delta t)^2 \, , \qquad (5.67)
\end{aligned}$$

where $M_j^{(4)} \equiv \mathrm{E}_Q[(Q_i - \mu_j)^4]$.

Proof. One general technique for calculating moments of the log-jump-diffusion process is **iterated expectations**, so

$$\begin{aligned}\mu_{ljd}(t) = \mathrm{E}[\Delta Y(t)] &= \mu_{ld}(t)\Delta t + \sigma_d(t)\cdot 0 + \mathrm{E}_{\Delta P(t;Q)}\left[\mathrm{E}_Q\left[\sum_{i=1}^{\Delta P(t;Q)} Q_i \,\middle|\, \Delta P(t;Q)\right]\right] \\ &= \mu_{ld}(t)\Delta t + \mathrm{E}_{\Delta P(t;Q)}\left[\sum_{i=1}^{\Delta P(t;Q)} \mathrm{E}_Q[Q_i]\right] \\ &= \mu_{ld}(t)\Delta t + \mathrm{E}_{\Delta P(t;Q)}[\Delta P(t;Q)\mathrm{E}_Q[Q_i]] = \left(\mu_{ld}(t) + \mu_j\lambda(t)\right)\Delta t\ ,\end{aligned}$$

proving the first moment formula.

For the higher moments, the main key technique for efficient calculation of the moments is decomposing the log-jump-diffusion process deviation into zero-mean deviation factors, i.e.,

$$\Delta Y(t) - \mu_{ljd}(t) = \sigma_d(t)\Delta W(t) + \sum_{i=1}^{\Delta P(t;Q)} (Q_i - \mu_j) + \mu_j(\Delta P(t;Q) - \lambda(t)\Delta t).$$

In addition, the multiple applications of the binomial theorem and the convenient increment power Tables 1.1 for $\Delta W(t)$ and 1.2 for $\Delta P(t;Q)$ are used.

The incremental process variance is found by

$$\begin{aligned}\sigma_{ljd}(t) \equiv\ & \mathrm{Var}[\Delta Y(t)] \\ =\ & \mathrm{E}\left[\left(\sigma_d(t)\Delta W(t) + \textstyle\sum_{i=1}^{\Delta P(t;Q)}(Q_i-\mu_j) + \mu_j(\Delta P(t;Q) - \lambda(t)\Delta t)\right)^2\right] \\ =\ & \sigma_d^2(t)\mathrm{E}_{\Delta W(t)}[(\Delta W)^2(t)] + 2\sigma_d\cdot 0 \\ & + \mathrm{E}\left[\left(\textstyle\sum_{i=1}^{\Delta P(t;Q)}(Q_i-\mu_j) + \mu_j(\Delta P(t;Q) - \lambda(t)\Delta t)\right)^2\right] \\ =\ & \sigma_d^2(t)\Delta t + \mathrm{E}_{\Delta P(t;Q)}\left[\textstyle\sum_{i=1}^{\Delta P(t;Q)}\sum_{k=1}^{\Delta P(t;Q)} \mathrm{E}_Q[(Q_i-\mu_j)(Q_k-\mu_j)]\right. \\ & + 2\mu_j(\Delta P(t;Q) - \lambda(t)\Delta t)\textstyle\sum_{i=1}^{\Delta P(t;Q)}\mathrm{E}_Q[(Q_i-\mu_j)] \\ & \left. + \mu_j^2(\Delta P(t;Q) - \lambda(t)\Delta t)^2\right] \\ =\ & \sigma_d^2(t)\Delta t + \mathrm{E}_{\Delta P(t;Q)}\left[\Delta P(t;Q)\sigma_j^2 + 0 + \mu_j^2(\Delta P(t;Q) - \lambda(t)\Delta t)^2\right] \\ =\ & \left(\sigma_d^2(t) + \left(\sigma_j^2 + \mu_j^2\right)\lambda(t)\right)\Delta t\ .\end{aligned}$$

The case of the third central moment is similarly calculated,

$$
\begin{aligned}
M_{ljd}^{(3)}(t) \equiv & \quad \mathrm{E}\left[(\Delta Y(t) - \mu_{ljd}(t))^3\right] \\
= & \quad \mathrm{E}\left[\left(\sigma_d(t)\Delta W(t) + \textstyle\sum_{i=1}^{\Delta P(t;Q)}(Q_i - \mu_j) + \mu_j(\Delta P(t;Q) - \lambda(t)\Delta t)\right)^3\right] \\
= & \quad \sigma_d^3(t)\mathrm{E}_{\Delta W(t)}\left[(\Delta W)^3(t)\right] \\
& +3\sigma_d^2\mathrm{E}_{\Delta W(t)}\left[(\Delta W)^2(t)\right]\mathrm{E}\left[\textstyle\sum_{i=1}^{\Delta P(t;Q)}(Q_i - \mu_j) + \mu_j(\Delta P(t;Q) - \lambda(t)\Delta t)\right] \\
& +3\sigma_d \cdot 0 + \mathrm{E}\left[\left(\textstyle\sum_{i=1}^{\Delta P(t;Q)}(Q_i - \mu_j) + \mu_j(\Delta P(t;Q) - \lambda(t)\Delta t)\right)^3\right] \\
= & \quad \sigma_d^3(t)\cdot 0 + 3\sigma_d^2(t)\Delta t \cdot 0 \\
& +\mathrm{E}_{\Delta P(t;Q)}\left[\textstyle\sum_{i=1}^{\Delta P(t;Q)}\sum_{k=1}^{\Delta P(t;Q)}\sum_{\ell=1}^{\Delta P(t;Q)}\mathrm{E}_Q[(Q_i - \mu_j)(Q_k - \mu_j)(Q_\ell - \mu_j)]\right. \\
& +3\mu_j(\Delta P(t;Q) - \lambda(t)\Delta t)\textstyle\sum_{i=1}^{\Delta P(t;Q)}\sum_{k=1}^{\Delta P(t;Q)}\mathrm{E}_Q[(Q_i - \mu_j)(Q_k - \mu_j)] \\
& \left.+3\mu_j^2(\Delta P(t;Q) - \lambda(t)\Delta t)^2 \cdot 0 + \mu_j^3(\Delta P(t;Q) - \lambda(t)\Delta t)^3\right] \\
= & \quad \mathrm{E}_{\Delta P(t;Q)}\left[\Delta P(t;Q)M_j^{(3)} + 3\mu_j(\Delta P(t;Q) - \lambda(t)\Delta t)\Delta P(t;Q)\sigma_j^2\right. \\
& \left.+\mu_j^3(\Delta P(t;Q) - \lambda(t)\Delta t)^3\right] \\
= & \quad \left(M_j^{(3)} + \mu_j\left(3\sigma_j^2 + \mu_j^2\right)\right)\lambda(t)\Delta t\ .,
\end{aligned}
$$

depending only on the jump component of the jump-diffusion.

The case of the fourth central moment is similarly calculated,

$$
\begin{aligned}
M^{(4)}_{ljd}(t) \equiv\ & \mathrm{E}\left[(\Delta Y(t)-\mu_{ljd}(t))^4\right] \\
=\ & \mathrm{E}\left[\left(\sigma_d(t)\Delta W(t)+\textstyle\sum_{i=1}^{\Delta P(t;Q)}(Q_i-\mu_j)+\mu_j(\Delta P(t;Q)-\lambda(t)\Delta t)\right)^4\right] \\
=\ & \sigma_d^4(t)\mathrm{E}_{\Delta W(t)}\left[(\Delta W)^4(t)\right]+4\sigma_d^3\cdot 0+6\sigma_d^2\mathrm{E}_{\Delta W(t)}\left[(\Delta W)^2(t)\right] \\
& \mathrm{E}\left[\left(\textstyle\sum_{i=1}^{\Delta P(t;Q)}(Q_i-\mu_j)+\mu_j(\Delta P(t;Q)-\lambda(t)\Delta t)\right)^2\right] \\
& +4\sigma_d\cdot 0+\mathrm{E}\left[\left(\textstyle\sum_{i=1}^{\Delta P(t;Q)}(Q_i-\mu_j)+\mu_j(\Delta P(t;Q)-\lambda(t)\Delta t)\right)^4\right] \\
=\ & 3\sigma_d^4(t)(\Delta t)^2+6\sigma_d^2(t)\Delta t\mathrm{E}_{\Delta P(t;Q)}\left[\textstyle\sum_{i=1}^{\Delta P(t;Q)}\sum_{k=1}^{\Delta P(t;Q)}\right. \\
& \mathrm{E}_Q[(Q_i-\mu_j)(Q_k-\mu_j)] \\
& +2\mu_j(\Delta P(t;Q)-\lambda(t)\Delta t)\cdot 0+\mu_j^2(\Delta P(t;Q)-\lambda(t)\Delta t)^2] \\
& +\mathrm{E}_{\Delta P(t;Q)}\left[\textstyle\sum_{i=1}^{\Delta P(t;Q)}\sum_{k=1}^{\Delta P(t;Q)}\sum_{\ell=1}^{\Delta P(t;Q)}\sum_{m=1}^{\Delta P(t;Q)}\right. \\
& \mathrm{E}_Q[(Q_i-\mu_j)(Q_k-\mu_j)(Q_\ell-\mu_j)(Q_m-\mu_j)] \\
& +4\mu_j(\Delta P(t;Q)-\lambda(t)\Delta t)\textstyle\sum_{i=1}^{\Delta P(t;Q)}\sum_{k=1}^{\Delta P(t;Q)}\sum_{\ell=1}^{\Delta P(t;Q)} \\
& \mathrm{E}_Q[(Q_i-\mu_j)(Q_k-\mu_j)(Q_\ell-\mu_j)] \\
& +6\mu_j^2(\Delta P(t;Q)-\lambda(t)\Delta t)^2\textstyle\sum_{i=1}^{\Delta P(t;Q)}\sum_{k=1}^{\Delta P(t;Q)}\mathrm{E}_Q[(Q_i-\mu_j)(Q_k-\mu_j)] \\
& +4\mu_j^3(\Delta P(t;Q)-\lambda(t)\Delta t)^3\cdot 0+\mu_j^4(\Delta P(t;Q)-\lambda(t)\Delta t)^4] \\
=\ & 3\sigma_d^4(t)(\Delta t)^2+6\sigma_d^2(t)\Delta t\mathrm{E}_{\Delta P(t;Q)}\left[\Delta P(t;Q)\sigma_j^2+\mu_j^2(\Delta P(t;Q)-\lambda(t)\Delta t)^2\right] \\
& +\mathrm{E}_{\Delta P(t;Q)}\left[\Delta P(t;Q)M_j^{(4)}+3\Delta P(t;Q)(\Delta P(t;Q)-1)\sigma_j^4\right. \\
& +4\mu_j(\Delta P(t;Q)-\lambda(t)\Delta t)\Delta P(t;Q)M_j^{(3)} \\
& +6\mu_j^2(\Delta P(t;Q)-\lambda(t)\Delta t)^2\Delta P(t;Q)\sigma_j^2+\mu_j^4(\Delta P(t;Q)-\lambda(t)\Delta t)^4] \\
=\ & \left(M_j^{(4)}+4\mu_jM_j^{(3)}+6\mu_j^2\sigma_j^2+\mu_j^4\right)\lambda(t)\Delta t \\
& +3\left(\sigma_d^2(t)+\left(\sigma_j^2+\mu_j^2\right)\lambda(t)\right)^2(\Delta t)^2\,,
\end{aligned}
$$

completing the proofs for moments $m=1{:}4$.

Also, used throughout, the expectations of odd powers of $\Delta W(t)$, the single powers of $(Q_i-\mu_j)$ and the single powers of $(\Delta P(t;Q)-\lambda(t)\Delta t)$ were immediately set to zero. In addition, the evaluation of the mark deviation sums of the form $\mathrm{E}[(\sum_{i=1}^k(Q_i-\mu_j)^m]$ for $m=1:4$ is based upon general formulas of Lemma 5.15. □

Remarks 5.18.

- *Recall that the third and fourth moments are measures of skewness and peakedness (kurtosis), respectively. The normalized representations in the current notation are the coefficient of skewness,*

$$\eta_3[\Delta Y(t)]\equiv M^{(3)}_{ljd}(t)/\sigma^3_{ljd}(t), \tag{5.68}$$

from (B.11), and the coefficient of kurtosis,

$$\eta_4[\Delta Y(t)]\equiv M^{(4)}_{ljd}(t)/\sigma^4_{ljd}(t)\,. \tag{5.69}$$

from (B.12).

- *For example, if the marks are normally or uniformly distributed, then*

$$M_j^{(3)} = 0,$$

since the normal and uniform distributions are both symmetric about the mean, so they lack skew and

$$\eta_3[\Delta Y(t)] = \frac{\mu_j \left(3\sigma_j^2 + \mu_j^2\right) \lambda(t)\Delta t}{\sigma_{ljd}^3(t)} = \frac{\mu_j \left(3\sigma_j^2 + \mu_j^2\right) \lambda(t)}{\left(\sigma_d^2(t) + \left(\sigma_j^2 + \mu_j^2\right) \lambda(t)\right)^3 (\Delta t)^2} ,$$

using $\sigma_{ljd}(t)$ given by (5.65). For the uniform distribution, the mean μ_j is given explicitly in terms of the uniform interval $[a, b]$ by (B.15) and the variance σ_j^2 by (B.16), while for the normal distribution, μ_j and σ_j^2 are the normal model parameters. In general, the normal and unform distribution versions of the log-jump-diffusion process will have skew although the component incremental diffusion and mark processes are skew-less.

In the normal and uniform mark cases, the fourth moment of the jump marks are

$$M_j^{(4)}/\sigma_j^4 = \left\{ \begin{array}{ll} 3, & \text{normal } Q_i \\ 1.8, & \text{uniform } Q_i \end{array} \right\} ,$$

which are in fact the coefficients of kurtosis for the normal and uniform distributions, respectively, so

$$\eta_4[\Delta Y(t)] = \left(\left\{ \begin{array}{ll} 3, & \text{normal } Q_i \\ 1.8, & \text{uniform } Q_i \end{array} \right\} \sigma_j^4 + 6\mu_j^2\sigma_j^2 + \mu_j^4 \right) \lambda(t)\Delta t/\sigma_{ljd}^4(t) \\ +3 \left(\sigma_d^2(t) + \left(\sigma_j^2 + \mu_j^2\right) \lambda(t)\right)^2 (\Delta t)^2/\sigma_{ljd}^4(t) .$$

- *The moment formulas for the differential log-jump-diffusion process $dY(t)$ follow immediately from Theorem 5.17 by dropping terms $\mathrm{O}((\Delta t)^2)$ and replacing Δt by dt.*

Distribution of Increment Log-Process:

Theorem 5.19. ***Distribution of the State Increment Logarithm Process for Linear Marked Jump-Diffusion SDE:***
Let logarithm-transform jump-amplitude be $\ln(1 + \nu(t, q)) = q$, then the increment of the logarithm process $Y(t) = \ln(X(t))$, assuming $X(t_0) = x_0 > 0$, approximately satisfies,

$$\Delta Y(t) \simeq \mu_{ld}(t)\Delta t + \sigma_d(t)\Delta W(t) + \sum_j^{\Delta P(t;Q)} \widehat{Q}_j , \tag{5.70}$$

where $\mu_{ld}(t) \equiv \mu_d(t) - \sigma_d^2(t)/2$ is the log-diffusion drift, $\sigma_d > 0$ and the $\widehat{Q}_j$ are pairwise IID jump marks for $P(s;Q)$ for $s \in (t, t+\Delta t]$, counting only jumps associated with $\Delta P(t;Q)$ given $P(t;Q)$, with common density $\phi_Q(q)$. The $\widehat{Q}_j$ are independent of both $\Delta P(t;Q)$ and $\Delta W(t)$.

Then the distribution of the log-process $Y(t)$ is the Poisson sum of nested convolutions

$$\Phi_{\Delta Y(t)}(x) \simeq \sum_{k=1}^{\infty} p_k(\lambda(t)\Delta t) \left(\Phi_{\Delta G(t)} \left(*\phi_Q\right)^k\right)(x) \,, \tag{5.71}$$

where $\Delta G(t) \equiv \mu_{ld}(t)\Delta t + \sigma_d(t)\Delta W(t)$ is the infinitesimal Gaussian process and $(\Phi_{\Delta G(t)}(*\phi_Q)^k)(x)$ denotes a convolution of one distribution with k identical densities ϕ_Q. The corresponding log-process density is

$$\phi_{\Delta Y(t)}(x) \simeq \sum_{k=1}^{\infty} p_k(\lambda(t)\Delta t) \left(\phi_{\Delta G(t)} \left(*\phi_Q\right)^k\right)(x) \,, \tag{5.72}$$

Proof. By the law of total probability (B.92), the distribution of the log-jump-diffusion $\Delta Y(t) \simeq \Delta G(t) + \sum_j^{\Delta P(t;Q)} \widehat{Q}_j$ is

$$\begin{aligned}
\Phi_{\Delta Y(t)}(x) &= \text{Prob}[\Delta Y(t) \le x] = \text{Prob}\left[\Delta G(t) + \sum_{j=1}^{\Delta P(t;Q)} \widehat{Q}_j \le x\right] \\
&= \sum_{k=0}^{\infty} \text{Prob}\left[\Delta G(t) + \sum_{j=1}^{\Delta P(t;Q)} \widehat{Q}_j \le x | \Delta P(t;Q) = k\right] \text{Prob}[\Delta P(t;Q) = k] \\
&= \sum_{k=0}^{\infty} p_k(\lambda(t)\Delta t)\Phi^{(k)}(x) \,,
\end{aligned} \tag{5.73}$$

where $p_k(\lambda(t)\Delta t)$ is the Poisson distribution with parameter $\lambda(t)\Delta t$ and letting

$$\Phi^{(k)}(x) \equiv \text{Prob}\left[\Delta G(t) + \sum_{j=1}^{k} \widehat{Q}_j \le x\right] .$$

For each discrete condition $\Delta P(t;Q) = k$, $\Delta Y(t)$ is the sum of $k+1$ terms, the normally distributed Gaussian diffusion part $\Delta G(t) = \mu_{ld}(t)\Delta t + \sigma_d(t)\Delta W(t)$ and the Poisson counting sum $\sum_{j=1}^{k} \widehat{Q}_j$ where the marks $\widehat{Q}_j$ are assumed to be IID but otherwise distributed with density $\phi_Q(q)$, while independent of the diffusion and the Poisson counting differential process $\Delta P(t;Q)$. Using the fact that $\Delta W(t)$ is normally distributed with zero-mean and Δt-variance,

$$\begin{aligned}
\Phi_{\Delta G(t)}(x) &= \text{Prob}[\Delta G(t) \le x] = \text{Prob}[\mu_{ld}(t)\Delta t + \sigma_d(t)\Delta W(t) \le x] \\
&= \text{Prob}[\Delta W(t) \le (x - \mu_{ld}(t)\Delta t)/\sigma_d(t)] = \Phi_{\Delta W(t)}((x - \mu_{ld}(t)\Delta t)/\sigma_d(t)) \\
&= \Phi_n((x - \mu_{ld}(t)\Delta t)/\sigma_d(t); 0, \Delta t) = \Phi_n(x; \mu_{ld}(t)\Delta t, \sigma_d^2(t)\Delta t) \,,
\end{aligned}$$

provided $\sigma_d(t) > 0$, while also using identities for normal distributions, where $\Phi_n(x;\mu,\sigma^2)$ denotes the normal distribution with mean μ and variance σ^2.

Since $\Phi^{(k)}$ is the distribution for the sum of $k+1$ independent random variables, one normally distibuted and k IID jump marks $\widehat{Q}_j$ for each k, $\Phi^{(k)}$ will be the nested convolutions as given in (B.100). Upon expanding in convolutions starting from the distribution for the random variable $\Delta G(t)$ and the kth Poisson counting sum

$$J_k \equiv \sum_{j=1}^{k} \widehat{Q}_j,$$

$$\Phi^{(k)}(x) = \left(\Phi_{\Delta G(t)} * \phi_{J_k}\right)(x) = \left(\Phi_{\Delta G(t)} \prod_{i=1}^{k} (*\phi_{Q_i})\right)(x) = \left(\Phi_{\Delta G(t)} (*\phi_Q)^k\right)(x) \, ,$$

using the identically distributed property of the Q_i's and the compact convolution operator notation

$$\left(\Phi_{\Delta G(t)} \prod_{i=1}^{k} (*\phi_{Q_i})\right)(x) = ((\cdots((\Phi_{\Delta G(t)} * \phi_{Q_1}) * \phi_{Q_2}) \cdots * \phi_{Q_{k-1}}) * \phi_{Q_k})(x) \, ,$$

which collapses to the operator power form for IID marks since $\prod_{i=1}^{k} c = c^k$ for some constant c. Substituting the distribution into the law total probability form (5.73), the desired result (5.71), which when differentiated with respect to x yields the kth density $\phi_{\Delta Y(t)}(x)$ in (5.72). □

Remark 5.20. *Several specialized variations of this theorem are found in Hanson and Westman [123, 125], but corrections to these papers are made here.*

Corollary 5.21. ***Density of Linear Jump-Diffusion with Log-Normally Distributed Jump Amplitudes:***
Let $X(t)$ be a linear jump-diffusion satisfying the SDE (5.70) and let the jump-amplitude mark Q be normally distributed such that

$$\phi_Q(x;t) = \phi_n(x;\mu_j(t),\sigma_j^2(t)) \tag{5.74}$$

with jump mean $\mu_j(t) = \mathrm{E}[Q]$ and jump variance $\sigma_j^2(t) = \mathrm{Var}[Q]$. Then the jump-diffusion density of the log-process $Y(t)$ is

$$\phi_{\Delta Y(t)}(x) = \sum_{k=1}^{\infty} p_k(\lambda(t)\Delta t)\phi_n(x;\mu_{ld}(t)\Delta t + k\mu_j(t), \sigma_d^2(t)\Delta t + k\sigma_j^2(t)) \, . \tag{5.75}$$

Proof. By (B.101) the convolution of two normal densities is a normal distribution with a mean that is the sum of means and a variance that is the sum of the variances.

Similarly, by the induction exercise result in (B.198) the pairwise convolution of one normally distributed diffusion process $\Delta G(t) = \mu_{ld}(t)\Delta t + \sigma_d(t)\Delta W(t)$ density and k random mark Q_i density ϕ_Q for $i = 1{:}k$ will be a normal density whose mean is the sum of the $k+1$ means and whose variance is the sum of the $k+1$ variances. Thus starting with the result (5.73) and then applying (B.198),

$$\begin{aligned}\phi_{\Delta Y(t)}(x) &= \sum_{k=1}^{\infty} p_k(\lambda(t)\Delta t)\left(\phi_{\Delta G(t)}\,(*\phi_Q)^k\right)(x) \\ &= \sum_{k=1}^{\infty} p_k(\lambda(t)\Delta t)\phi_n\left(x; \mu_{ld}(t)\Delta t + \sum_{i=1}^{k}\mu_j(t), \sigma_d^2(t)\Delta t + \sum_{i=1}^{k}\sigma_j^2(t)\right) \\ &= \sum_{k=1}^{\infty} p_k(\lambda(t)\Delta t)\phi_n(x; \mu_{ld}(t)\Delta t + k\mu_j(t), \sigma_d^2(t)\Delta t + k\sigma_j^2(t))\ .\end{aligned}$$

□

Remark 5.22. *The normal jump-amplitude jump-diffusion distribution has been used in financial applications, initially by Merton [202] and then by others such as Düvelmeyer [75], Andersen et al. [6] and Hanson and Westman [123].*

Corollary 5.23. ***Density of Linear Jump-Diffusion with Log-Uniformly Distributed Jump Amplitudes:***
Let $X(t)$ be a linear jump-diffusion satisfying the SDE (5.70), and let the jump-amplitude mark Q be uniformly distributed (5.28). i.e.,

$$\phi_Q(q) = \frac{1}{b-a}U(q; a, b)\ ,$$

where $U(q; a, b)$ is the unit step function on $[a, b]$ with $a < b$. The jump mean is $\mu_j(t) = (b+a)/2$ and jump variance is $\sigma_j^2(t) = (b-a)^2/12$.

Then the jump-diffusion density of the increment log-process $\Delta Y(t)$ satisfies the general convolution form (5.72), i.e.,

$$\phi_{\Delta Y(t)}(x) = \sum_{k=1}^{\infty} p_k(\lambda(t)\Delta t)\left(\phi_{\Delta G(t)}\,(*\phi_Q)^k\right)(x) = \sum_{k=1}^{\infty} p_k(\lambda(t)\Delta t)\phi_{\text{ujd}}^{(k)}(x)\ , \quad (5.76)$$

where $p_k(\lambda(t)\Delta t)$ is the Poisson distribution with parameter $\lambda(t)$. The $\Delta G(t) = \mu_{ld}(t)\Delta t + \sigma_d(t)\Delta W(t)$ is the diffusion term and Q is the uniformly distributed jump-amplitude mark. The first few coefficients of $p_k(\lambda(t)\Delta t)$ for the uniform jump-distribution (ujd) are

$$\phi_{\text{ujd}}^{(0)}(x) = \phi_{\Delta G(t)}(x) = \phi_n(x; \mu_{ld}(t)\Delta t, \sigma_d^2(t)\Delta t)\ , \quad (5.77)$$

where $\phi_n(x; \mu_{ld}(t)\Delta t, \sigma_d^2(t)\Delta t)$ denotes the normal density with mean $\mu_{ld}(t)$ and variance $\sigma_d(t)\Delta t$,

$$\phi_{\text{ujd}}^{(1)}(x) = \left(\phi_{\Delta G(t)} * \phi_Q\right)(x) = \phi_{sn}(x-b, x-a; \mu_{ld}(t)\Delta t, \sigma_d^2(t)\Delta t)\ , \quad (5.78)$$

where ϕ_{sn} *is the* ***secant-normal density***

$$\begin{aligned}\phi_{sn}(x_1, x_2; \mu, \sigma^2) &\equiv \frac{1}{(x_2 - x_1)} \Phi_n(x_1, x_2; \mu, \sigma^2) \\ &\equiv \frac{\Phi_n(x_2; \mu, \sigma^2) - \Phi_n(x_1; \mu, \sigma^2)}{x_2 - x_1}\end{aligned} \tag{5.79}$$

with normal distribution $\Phi_n(x_1, x_2; \mu, \sigma^2)$ *such that*

$$\Phi_n(x_i; \mu, \sigma^2) \equiv \Phi_n(-\infty, x_i; \mu, \sigma^2)$$

for $i = 1:2$, *and*

$$\begin{aligned}\phi^{(2)}_{\text{ujd}}(x) &= \left(\phi_{\Delta G(t)}(*\phi_Q)^2\right)(x) \\ &= \frac{2b - x + \mu_{ld}(t)\Delta t}{b - a} \phi_{sn}(x - 2b, x - a - b; \mu_{ld}(t)\Delta t, \sigma^2(t)\Delta t) \\ &\quad + \frac{x - 2a - \mu_{ld}(t)\Delta t}{b - a} \phi_{sn}(x - a - b, x - 2a; \mu_{ld}(t)\Delta t, \sigma^2_d(t)\Delta t) \\ &\quad + \frac{\sigma^2_d(t)\Delta t}{(b - a)^2} \left(\phi_n(x - 2b; \mu_{ld}(t)\Delta t, \sigma^2_d(t)\Delta t)\right. \\ &\quad \left. -2\phi_n(x - a - b; \mu_{ld}(t)\Delta t, \sigma^2_d(t)\Delta t) + \phi_n(x - 2a; \mu_{ld}(t)\Delta t, \sigma^2_d(t)\Delta t)\right) .\end{aligned} \tag{5.80}$$

Proof. First the finite range of the jump-amplitude uniform density is used to truncate the convolution integrals for each k using existing results for the mark convolutions like $\phi^{(2)}_{\text{uq}}(x) = (\phi_Q * \phi_Q)(x) = \phi_{Q_1+Q_2}(x)$ for IID marks when $k = 2$.

The case for $k = 0$ is trivial since it is given in the theorem equations (5.77).

For $k = 1$ jump,

$$\begin{aligned}\phi^{(1)}_{\text{ujd}}(x) = (\phi_{\Delta G(t)} * \phi_Q)(x) &= \int_{-\infty}^{+\infty} \phi_{\Delta G(t)}(x - y)\phi_Q(y)dy \\ &= \frac{1}{b - a} \int_a^b \phi_n(x - y; \mu_{ld}(t)\Delta t, \sigma^2_d(t)\Delta t)dy \\ &= \frac{1}{b - a} \int_{x-b}^{x-a} \phi_n(z; \mu_{ld}(t)\Delta t, \sigma^2_d(t)\Delta t)dz \\ &= \frac{1}{b - a} \Phi_n(x - b, x - a; \mu_{ld}(t)\Delta t, \sigma^2_d(t)\Delta t) \\ &= \phi_{sn}(x - b, x - a; \mu_{ld}(t)\Delta t, \sigma^2_d(t)\Delta t) ,\end{aligned}$$

$-\infty < x < +\infty$, upon change of variables and use of identities.

For $k = 2$ jumps, the triangular distribution exercise result (B.199) is

$$\phi^{(2)}_{\text{uq}}(x) = (\phi_Q * \phi_Q)(x) = \frac{1}{(b - a)^2} \left\{ \begin{array}{ll} x - 2a, & 2a \le x < a + b \\ 2b - x, & a + b \le x \le 2b \\ 0, & \text{otherwise} \end{array} \right\} . \tag{5.81}$$

Hence,

$$\begin{aligned}
\phi^{(2)}_{\text{ujd}}(x) &= (\phi_{\Delta G(t)} * (\phi_Q * \phi_Q))(x) = \int_{-\infty}^{+\infty} \phi_{\Delta G(t)}(x-y)(\phi_Q * \phi_Q)(y)dy \\
&= \frac{1}{(b-a)^2}\left(\int_{2a}^{a+b}(y-2a)\phi_{\Delta G(t)}(x-y)dy + \int_{a+b}^{2b}(2b-y)\phi_{\Delta G(t)}(x-y)dy\right) \\
&= \frac{1}{(b-a)^2}\left(\int_{x-a-b}^{x-2a}(x-z-2a)\phi_{\Delta G(t)}(z)dz\right. \\
&\quad \left. + \int_{x-2b}^{x-a-b}(2b-x+z)\phi_{\Delta G(t)}(z)dz\right) \\
&= \frac{2b-x+\mu_{ld}(t)\Delta t}{b-a}\phi_{sn}(x-2b, x-a-b; \mu_{ld}(t)\Delta t, \sigma_d^2(t)\Delta t) \\
&\quad + \frac{x-2a-\mu_{ld}(t)\Delta t}{b-a}\phi_{sn}(x-a-b, x-2a; \mu_{ld}(t)\Delta t, \sigma_d^2(t)\Delta t) \\
&\quad + \frac{\sigma_d^2(t)\Delta t}{(b-a)^2}\left(\phi_n(x-2b; \mu_{ld}(t)\Delta t, \sigma_d^2(t)\Delta t)\right. \\
&\quad \left. -2\phi_n(x-a-b; \mu_{ld}(t)\Delta t, \sigma_d^2(t)\Delta t) + \phi_n(x-2a; \mu_{ld}(t)\Delta t, \sigma_d^2(t)\Delta t)\right) ,
\end{aligned}$$

where the exact integral for the normal density has been used . □

Remarks 5.24.

- *This density form ϕ_{sn} in (5.79) is called a* ***secant-normal density*** *since the numerator is an increment of the normal distribution and the denominator is the corresponding increment in its state arguments, i.e., a secant approximation, which here has the form $\Delta\Phi_n/\Delta x$.*
- *The uniform jump-amplitude jump-diffusion distribution has been used in financial applications, initially by the authors in [125] as a simple, but appropriate, representation of jump component of market distributions, and some errors have been corrected here.*

Example 5.25. ***Linear SDE Simulator for Log-Uniformly Distributed Jump Amplitudes:*** *The linear SDE jump-diffusion simulator in Appendix MATLAB code C.14 can be converted from the simple discrete jump process to the distributed jump process here. The primary change is the generation of the another set of random numbers for the mark process Q, e.g.,*

$$Q = a + (b-a) * \texttt{rand}(1, n+1)$$

for a set of $n+1$ uniformly distributed marks on (a, b) so that the jump amplitudes of $X(t)$ are log-uniformly distributed.

An example is demonstrated in Fig. 5.1 for uniformly distributed marks Q on $(a, b) = (-2, +1)$ and time-dependent coefficients $\{\mu_d(t), \sigma_d(t), \lambda(t)\}$. The MATLAB linear mark-jump-diffusion code C.15 is a modification of the linear jump-diffusion SDE simulator code C.14 illustrated in Fig. 4.3 for constant coefficients and discrete mark-independent jumps. The state exponent $Y(t)$ is simulated as

$$YS(i+1) = YS(i) + (\mu_d(i) - \sigma_d^2(i)/2) * \Delta t + \sigma_d(i) * DW(i) + Q(i) * DP(i) \ ,$$

with $t(i+1) = t0 + i * \Delta t$ for $i = 0 : n$ with $n = 1,000$, $t0 = 0$, $0 \leq t(i) \leq 2$. The incremental Poisson jump term $\Delta P(i) = P(t_i + \Delta t) - P(t_i)$ is simulated by a uniform random number generator on $(0, 1)$ using the acceptance-rejection technique [230, 96] to implement the zero-one jump law to obtain the probability of $\lambda(i)\Delta t$ that there a jump is accepted. The same random state is used to obtain the simulations of uniformly distributed Q on (a, b) conditional on a jump event.

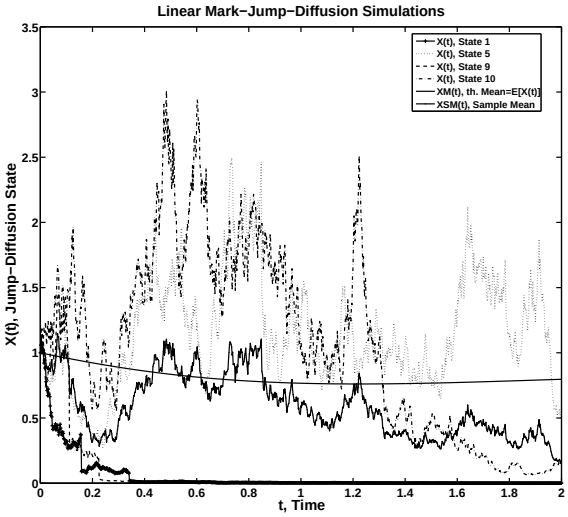


Figure 5.1. *Four linear mark-jump-diffusion sample paths for time-dependent coefficients are simulated using MATLAB [210] with $N = 1,000$ time-steps, maximum time $T = 2.0$ and four* `randn` *and four* `rand` *states. Initially, $x_0 = 1.0$. Parameter values are given in vectorized functions using vector functions and dot-element operations, $\mu_d(t) = 0.1 * sin(t)$, $\sigma_d(t) = 1.5 * exp(-0.01 * t)$ and $\lambda = 3.0 * exp(-t. * t)$. The marks are uniformly distributed on $[-2.0, +1.0]$. In addition to the four simulated states, the expected state $\mathrm{E}[X(t)]$ is presented using quasi-deterministic equivalence (5.55) of Hanson and Ryan [114], but also the sample mean of the four sample paths are presented.*

5.3 Multi-Dimensional Markov SDE

The general, multi-dimensional Markov SDE is presented here, along with the corresponding chain rule, establishing proper matrix-vector notation, or extensions where the standard linear algebra is inadequate, for the sequel. In the case of the vector[1]state process $\mathbf{X}(t) = [X_i(t)]_{n_x \times 1}$ on some n_x-dimensional state space $\mathcal{D}_x$, the multi-dimensional SDE can be of the form,

$$d\mathbf{X}(t) \stackrel{\text{sym}}{=} \mathbf{f}(\mathbf{X}(t), t)dt + g(\mathbf{X}(t), t)d\mathbf{W}(t) + h(\mathbf{X}(t), t, \mathbf{Q})d\mathbf{P}(t; Q, \mathbf{X}(t), t) \ , \quad (5.82)$$

where

$$\begin{aligned} h(\mathbf{X}(t), t, \mathbf{Q})d\mathbf{P}(t; Q, \mathbf{X}(t), t) &\stackrel{\text{dt}}{=} d\Pi(t; \mathbf{X}(t), t) \\ &= \int_{\mathcal{Q}} h(\mathbf{X}(t), t, \mathbf{q})\boldsymbol{\mathcal{P}}(\mathbf{dt}, \mathbf{dq}; \mathbf{X}(t), t) \end{aligned} \quad (5.83)$$

is compact symbolic notation for the space-time Poisson terms, $\mathbf{W}(t) = [W_i(t)]_{n_w \times 1}$ is an n_w-dimensional vector Wiener process, $\mathbf{P}(t; Q, \mathbf{X}(t), t) = [P_i(t; \mathbf{X}(t), t)]_{n_p \times 1}$ is an n_p-dimensional vector state-dependent Poisson process, the coefficient $\mathbf{f}$ has the same dimension at $\mathbf{X}$, and the coefficients in the set $\{g, h\}$ have dimensions commensurate in multiplication with the set of vectors $\{\mathbf{W}, \mathbf{P}\}$, respectively. Here, $\boldsymbol{\mathcal{P}} = [\mathcal{P}_i]_{n_p \times 1}$ is a vector form of the Poisson random measure with mark random vector $\mathbf{Q} = [Q_i]_{n_p \times 1}$ and $\mathbf{dq} = [[q_i, q_i + dq_i)]_{n_p \times 1}$ is the symbolic vector version of the mark measure notation. The $d\mathbf{P}(t; \mathbf{X}(t), t)$ jump-amplitude coefficient has the component form

$$h(\mathbf{X}(t), t; \mathbf{Q}) = [h_{i,j}(\mathbf{X}(t), t; Q_j)]_{n_x \times n_p},$$

such that the jth Poisson component only depends on the jth mark Q_j since simultaneous jumps are unlikely.

In component and jump counter form, the SDE is

$$\begin{aligned} dX_i(t) \stackrel{\text{dt}}{=} f_i(\mathbf{X}(t), t)dt &+ \sum_{j=1}^{n_w} g_{i,j}(\mathbf{X}(t), t)dW_j(t) \\ &+ \sum_{j=1}^{n_p} h_{i,j}(\mathbf{X}(t), t, \mathbf{Q})dP_j(t; Q, \mathbf{X}(t), t) \ , \end{aligned} \quad (5.84)$$

for $i = 1 : n_x$ state components. The jump of the ith state due to the jth Poisson process

$$[X_i](T_{j,k}) = h_{i,j}(\mathbf{X}(T_{j,k}^-), T_{j,k}^-, Q_{j,k}),$$

where $T_{j,k}^-$ is the pre-jump-time and its k realization with jump-amplitude mark $Q_{j,k}$. The diffusion noise components have zero mean,

$$\mathrm{E}[dW_i(t)] = 0 \quad (5.85)$$

[1]Boldface variables or processes denote column vector variables or processes, respectively. The subscript i usually denotes a row index in this notation, while j denotes a column index. For example, $\mathbf{X}(t) = [X_i(t)]_{n_x \times 1}$ denotes that X_i is the ith component for $i = 1 : n_x$ of the single-column vector $\mathbf{X}(t)$.

for $i = 1:n_w$, while **correlations** are allowed between components,

$$\text{Cov}[dW_i(t), dW_j(t)] = \rho_{i,j}dt = [\delta_{i,j} + \rho_{i,j}(1 - \delta_{i,j})]dt, \tag{5.86}$$

for $i, j = 1:n_x$, where $\rho_{i,j}$ is the correlation coefficient between i and j components.

The jump noise components, conditioned on $\mathbf{X}(t) = \mathbf{x}$, are Poisson distributed with $\mathcal{P}$ mean assumed to be of the form

$$\text{E}[\mathcal{P}_j(\mathbf{dt}, \mathbf{dq}_j; \mathbf{X}(t), t)|\mathbf{X}(t) = \mathbf{x}] = \phi_{\mathcal{Q}_j}^{(j)}(q_j; \mathbf{x}, t)dq_j\lambda_j(t; \mathbf{x}, t)dt \ , \tag{5.87}$$

for each jump component $j = 1:n_p$ with jth density $\phi_Q^{(j)}(q_j; \mathbf{x}, t)$ depending only on q_j assuming independence of the marks for different Poisson components but IID for the same component, so that the Poisson mark integral is

$$\begin{aligned}
\text{E}[dP_j(t; Q, \mathbf{X}(t), t)|\mathbf{X}(t) = \mathbf{x}] &= \text{E}\left[\int_{\mathcal{Q}_j} \mathcal{P}_j(\mathbf{dt}, \mathbf{dq}_j; \mathbf{x}(t), t)\right] \\
&= \int_{\mathcal{Q}_j} \text{E}\left[\mathcal{P}_j(\mathbf{dt}, \mathbf{dq}_j; \mathbf{x}(t), t)\right] \\
&= \int_{\mathcal{Q}_j} \phi_Q^{(j)}(q_j; \mathbf{x}, t)dq_i\lambda_j(t; \mathbf{x}, t)dt \\
&= \lambda_j(t; \mathbf{x}, t)dt
\end{aligned} \tag{5.88}$$

for $i = 1:n_p$, while the components are assumed to be uncorrelated, with conditioning $\mathbf{X}(t) = \mathbf{x}$ pre-assumed for brevity,

$$\text{Cov}[\mathcal{P}_j(\mathbf{dt}, \mathbf{dq}_j; \mathbf{x}, t)\mathcal{P}_k(\mathbf{dt}, \mathbf{dq_k}; \mathbf{x}, t)] = \phi_Q^{(j)}(q_j; \mathbf{x}, t)\delta(q_k - q_j)dq_kdq_j\lambda_j(t; \mathbf{x}, t)dt \ , \tag{5.89}$$

generalizing the scalar form (5.15) to vector form, and

$$\begin{aligned}
\text{Cov}[dP_j(t; Q_j, \mathbf{x}, t), dP_k(t; Q_k, \mathbf{x}, t)] &= \int_{\mathcal{Q}_j}\int_{\mathcal{Q}_k} \text{Cov}[\mathcal{P}_j(\mathbf{dt}, \mathbf{dq}_j; \mathbf{x}, t)\mathcal{P}_k(\mathbf{dt}, \mathbf{dq_k}; \mathbf{x}, t)] \\
&= \lambda_j(t; \mathbf{x}, t)dt\ \delta_{j,k}
\end{aligned} \tag{5.90}$$

for $j, k = 1:n_p$, there being enough complexity for most applications. In addition, it is assumed that, as vectors, the diffusion noise $d\mathbf{W}$, Poisson noise $d\mathbf{P}$ and mark random variable $\mathbf{Q}$ are pairwise independent, but the mark random variable depends on the existence of a jump.

This Poisson formulation is somewhat different from others, such as [94, Part 2, Chapter 2]. The linear combination form has been found to be convenient for both jumps and diffusion when there several sources of noise in the application.

5.3.1 Conditional Infinitesimal Moments in Multi-Dimensions

The conditional infinitesimal moments for the vector state process $\mathbf{X}(t)$ are more easily calculated by component first, using the noise infinitesimal moments (5.85-

5.90). The conditional infinitesimal mean is

$$\begin{aligned}
\mathrm{E}[dX_i(t)|\mathbf{X}(t)=\mathbf{x}] &= f_i(\mathbf{x},t)dt + \sum_{j=1}^{n_w} g_{i,j}(\mathbf{x},t)\mathrm{E}[dW_j(t)] \\
&\quad + \sum_{j=1}^{n_p} \int_{\mathcal{Q}_j} h_{i,j}(\mathbf{x},t,q_j) E[\mathcal{P}_j(\mathbf{dt},\mathbf{dq}_j;\mathbf{x},t)] \\
&= f_i(\mathbf{x},t)dt + \sum_{j=1}^{n_p} \int_{\mathcal{Q}_j} h_{i,j}(\mathbf{x},t,q_j)\phi_Q^{(j)}(q_j;\mathbf{x},t)dq_j\lambda_j(t;\mathbf{x},t)dt \\
&= \left[f_i(\mathbf{x},t) + \sum_{j=1}^{n_p} \overline{h}_{i,j}(\mathbf{x},t)\lambda_j(t;\mathbf{x},t) \right] dt \qquad (5.91)
\end{aligned}$$

where $\overline{h}_{i,j}(\mathbf{x},t) \equiv E_Q[h_{i,j}(\mathbf{x},t,Q_j)]$. Thus, in vector form

$$\mathrm{E}[d\mathbf{X}(t)|\mathbf{X}(t)=\mathbf{x}] = \left[\mathbf{f}(\mathbf{x},t)dt + \overline{\boldsymbol{h}}(\mathbf{x},t)\boldsymbol{\lambda}(t;\mathbf{x},t)\right] dt \ , \qquad (5.92)$$

where $\boldsymbol{\lambda}(t;\mathbf{x},t) = [\lambda_i(t;\mathbf{x},t)]_{n_p\times 1}$.

For the conditional infinitesimal covariance, again with pre-assuming conditioning on $\mathbf{X}(t)=\mathbf{x}$ for brevity,

$$\begin{aligned}
\mathrm{Cov}[dX_i(t),dX_j(t)] &= \sum_{k=1}^{n_w}\sum_{\ell=1}^{n_w} g_{i,k}(\mathbf{x},t)g_{j,\ell}(\mathbf{x},t)\mathrm{Cov}[dW_k(t),dW_\ell(t)] \\
&\quad + \sum_{k=1}^{n_p}\sum_{\ell=1}^{n_p} \int_{\mathcal{Q}_k}\int_{\mathcal{Q}_\ell} h_{i,k}(\mathbf{x},t;q_k)h_{j,\ell}(\mathbf{x},t;q_\ell) \\
&\quad \mathrm{Cov}[\mathcal{P}_k(\mathbf{dt},\mathbf{dq}_k;\mathbf{x},t),\mathcal{P}_\ell(\mathbf{dt},\mathbf{dq}_\ell;\mathbf{x},t)] \\
&= \sum_{k=1}^{n_w}\left(g_{i,k}(\mathbf{x},t)g_{j,k}(\mathbf{x},t) + \sum_{\ell\neq k}\rho_{k,\ell}g_{i,k}(\mathbf{x},t)g_{j,\ell}(\mathbf{x},t)\right) dt \\
&\quad + \sum_{k=1}^{n_p}(h_{i,k}h_{j,k})(\mathbf{x},t)\phi_Q^{(k)}(q_k;\mathbf{x},t)\lambda_k(t;\mathbf{x},t)dt \\
&= \sum_{k=1}^{n_w}\left(g_{i,k}(\mathbf{x},t)g_{j,k}(\mathbf{x},t) + \sum_{\ell\neq k}\rho_{k,\ell}g_{i,k}(\mathbf{x},t)g_{j,\ell}(\mathbf{x},t)\right) dt \\
&\quad + \sum_{k=1}^{n_p}\overline{(h_{i,k}h_{j,k})}(\mathbf{x},t)\lambda_k(t;\mathbf{x},t)dt \ , \qquad (5.93)
\end{aligned}$$

for $i=1:n_x$ and $j=1:n_x$, in precision-dt, where the infinitesimal jump-diffusion covariance formulas (5.86) and (5.89) have been used. Hence, the matrix-vector form of this covariance is

$$\begin{aligned}
\mathrm{Cov}[d\mathbf{X}(t),d\mathbf{X}^\top(t)|\mathbf{X}(t)=\mathbf{x}] \overset{\mathrm{dt}}{=} &\left[g(\mathbf{x},t)R' g^\top(\mathbf{x},t)\right. \\
&\left. + \overline{\mathbf{h}\Lambda\mathbf{h}^\top}(\mathbf{x},t)\right] dt \ , \qquad (5.94)
\end{aligned}$$

where

$$R' \equiv [\rho_{i,j}]_{n_w \times n_w} = [\delta_{i,j} + \rho_{i,j}(1-\delta_{i,j})]_{n_w \times n_w} \ , \tag{5.95}$$

$$\Lambda = \Lambda(t;\mathbf{x},t)) = [\lambda_i(t;\mathbf{x},t)\delta_{i,j}]_{n_p \times n_p} \ . \tag{5.96}$$

The jump in the ith component of the state at jump-time $T_{j,k}$ in the underlying jth component of the vector Poisson process is

$$[X_i](T_{j,k}) \equiv X_i(T_{j,k}^+) - X_i(T_{j,k}^-) = h_{i,j}(X(T_{j,k}^-), T_{j,k}^-; Q_{j,k}) \ , \tag{5.97}$$

for $k = 1:\infty$ jumps and $i = 1:n_x$ state components, now depending on the jth mark's kth realization $Q_{j,k}$ at the pre-jump-time $T_{j,k}^-$ at the kth jump of the jth component Poisson process.

5.3.2 Stochastic Chain Rule in Multi-Dimensions

The stochastic chain rule for a scalar function $\mathbf{Y}(t) = \mathbf{F}(\mathbf{X}(t),t)$, twice continuously differentiable in $\mathbf{x}$ and once in t, comes from the expansion,

$$\begin{aligned}
d\mathbf{Y}(t) = d\mathbf{F}(\mathbf{X}(t),t) &= \mathbf{F}(\mathbf{X}(t)+d\mathbf{X}(t), t+dt) - \mathbf{F}(\mathbf{X}(t),t) \\
&= \mathbf{F}_t(\mathbf{X}(t),t) + \sum_{i=1}^{n_x} \frac{\partial \mathbf{F}}{\partial x_i}(\mathbf{X}(t),t) \left(f_i(\mathbf{X}(t),t)dt + \sum_{k=1}^{n_w} g_{i,k}(\mathbf{X}(t),t)dW_k(t) \right) \\
&\quad + \frac{1}{2} \sum_{i=1}^{n_x}\sum_{j=1}^{n_x}\sum_{k=1}^{n_w}\sum_{\ell=1}^{n_w} \left(\frac{\partial^2 \mathbf{F}}{\partial x_i \partial x_j} g_{i,k} g_{j,\ell} \right) (\mathbf{X}(t),t) dW_k(t) dW_\ell(t) \\
&\quad + \sum_{j=1}^{n_p} \int_{\mathcal{Q}} \Big(\mathbf{F}\left(\mathbf{X}(t) + \widehat{\boldsymbol{h}}_j(\mathbf{X}(t),t,q_j), t \right) - \mathbf{F}(\mathbf{X}(t),t) \Big) \\
&\quad \cdot \mathcal{P}_j(\mathbf{dt}, \mathbf{dq}_j; \mathbf{X}(t), t) \ , \\
&\overset{\mathrm{dt}}{=} \left(\mathbf{F}_t(\mathbf{X}(t),t) + \mathbf{f}^\top(\mathbf{X}(t),t) \nabla_x[\mathbf{F}](\mathbf{X}(t),t) \right) dt \\
&\quad + \frac{1}{2} \sum_{i=1}^{n_x}\sum_{j=1}^{n_x} \frac{\partial^2 \mathbf{F}}{\partial x_i \partial x_j} \sum_{k=1}^{n_w} \left(g_{i,k} g_{j,k} + \sum_{\ell \neq k}^{n_w} \rho_{k,\ell} g_{i,k} g_{j,\ell} \right) (\mathbf{X}(t),t) dt \\
&\quad + \sum_{j=1}^{n_p} \int_{\mathcal{Q}_j} \Delta_j[\mathbf{F}] \mathcal{P}_j \\
&= \left[\mathbf{F}_t + \mathbf{f}^\top \nabla_x[\mathbf{F}] + \frac{1}{2} \left(g R' g^\top \right) : \nabla_x \left[\nabla_x^\top [\mathbf{F}] \right] \right] (\mathbf{X}(t),t) dt \\
&\quad + \int_{\mathcal{Q}} \boldsymbol{\Delta}^\top[\mathbf{F}] \boldsymbol{\mathcal{P}}
\end{aligned} \tag{5.98}$$

to precision-dt. Here, the

$$\nabla_x[\mathbf{F}] \equiv \left[\frac{\partial \mathbf{F}}{\partial x_i}(\mathbf{x},t) \right]_{n_x \times 1}$$

is the state space gradient (a column n_x-vector),

$$\nabla_x^\top[\mathbf{F}] \equiv \left[\frac{\partial \mathbf{F}}{\partial x_j}(\mathbf{x},t)\right]_{1\times n_x}$$

is the transpose of the state space gradient (a row n_x-vector),

$$\nabla_x\left[\nabla_x^\top[\mathbf{F}]\right] \equiv \left[\frac{\partial^2 \mathbf{F}}{\partial x_i \partial x_j}(\mathbf{x},t)\right]_{n_x\times n_x}$$

is the Hessian matrix for $\mathbf{F}$, R' is a correlation matrix defined in (5.95),

$$A : B \equiv \sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}B_{i,j} = \text{Trace}[AB^\top] \tag{5.99}$$

is the **double-dot product** of two $n \times n$ matrices, related to the trace,

$$\widehat{\boldsymbol{h}}_j(\mathbf{x},t,q_j) \equiv [h_{i,j}(\mathbf{x},t,q_j)]_{n_x\times 1} \tag{5.100}$$

is the jth jump-amplitude vector corresponding to the jth Poisson process,

$$\begin{aligned}\boldsymbol{\Delta}^\top[\mathbf{F}] =&\ [\Delta_j[\mathbf{F}](\mathbf{X}(t),t,q_j)]_{1\times n_p}\\ \equiv&\ \left[\mathbf{F}(\mathbf{X}(t)+\widehat{\boldsymbol{h}}_j(\mathbf{X}(t),t,q_j),t)-\mathbf{F}(\mathbf{X}(t),t)\right]_{1\times n_p}\end{aligned} \tag{5.101}$$

is the general jump-amplitude change vector for any t and

$$\boldsymbol{\mathcal{P}} = [\mathcal{P}_i(\mathbf{dt},\mathbf{dq}_i;\mathbf{X}(t),t)]_{n_p\times 1}$$

is the Poisson random measure vector condition. The corresponding jump in $\mathbf{Y}(t)$ due to the jth Poisson component and its kth realization is

$$[\mathbf{Y}]\left(T_{j,k}^-\right) = \mathbf{F}\left(\mathbf{X}\left(T_{j,k}^-\right)+\widehat{\boldsymbol{h}}_j\left(\mathbf{X}\left(T_{j,k}^-\right),T_{j,k}^-,Q_{j,k}\right),T_{j,k}^-\right)-\mathbf{F}\left(\mathbf{X}\left(T_{j,k}^-\right),T_{j,k}^-\right).$$

Example 5.26. ***Merton's Analysis of Black-Scholes Option Pricing Model:*** *A good application of multi-dimensional SDEs in finance is the survey of Merton's [201] (Merton [203, Chapter 8]) analysis of the Black-Scholes [34] financial options pricing model in Section 10.2 of Chapter 10. This treatment will serve as motivation for the study of SDEs and contains details not in Merton's paper.*

5.4 Distributed Jump SDE Models Exactly Transformable

Here, exactly transformable distributed jump-diffusion SDE models are listed, both in the scalar and vector cases, where conditions are applicable.

5.4.1 Distributed Jump SDE Models Exactly Transformable

- **Distributed Scalar Jump SDE:**

$$dX(t) = f(X(t),t)dt + g(X(t),t)dW(t) + \int_{\mathcal{Q}} h(X(t),t,q)\mathcal{P}(\mathbf{dt},\mathbf{dq}).$$

- **Transformed Scalar Process:** $Y(t) = F(X(t),t)$.
- **Transformed Scalar SDE:**

$$\begin{aligned} dY(t) = (F_t + F_x f + \frac{1}{2}F_{xx}g^2)dt + F_x g dW(t) \\ + \int_{\mathcal{Q}} (F(X(t)+h(X(t),t,q),t) - F(X(t),t))\mathcal{P}(\mathbf{dt},\mathbf{dq}). \end{aligned}$$

- **Target Explicit Scalar SDE:**

$$dY(t) = C_1(t)dt + C_2(t)dW(t) + \int_{\mathcal{Q}} C_3(t,q)\mathcal{P}(\mathbf{dt},\mathbf{dq}).$$

5.4.2 Vector Distributed Jump SDE Models Exactly Transformable

- **Vector Distributed Jump SDE:**

$$d\mathbf{X}(t) = \mathbf{f}(\mathbf{X}(t),t)dt + g(\mathbf{X}(t),t)d\mathbf{W}(t) + \int_{\mathcal{Q}} h(\mathbf{X}(t),t,\mathbf{q})\boldsymbol{\mathcal{P}}(\mathbf{dt},\mathbf{dq}).$$

- **Vector Transformed Process:** $\mathbf{Y}(t) = \mathbf{F}(\mathbf{X}(t),t)$.
- **Transformed Component SDE:**

$$\begin{aligned} dY_i(t) = (F_{i,t} + \sum_j F_{i,j}f_j + \frac{1}{2}\sum_j\sum_k\sum_l F_{i,jk}g_{jl}g_{kl})dt \\ + \sum_j F_{i,j}\sum_l g_{jl}dW_l(t) \\ + \sum_\ell \int_{\mathcal{Q}} (y_i(\mathbf{X}+\mathbf{h}_\ell,t) - F_i(\mathbf{X},t))\mathcal{P}_\ell(\mathbf{dt},\mathbf{dq}_\ell), \end{aligned}$$

$$\mathbf{h}_\ell(\mathbf{x},t,\mathbf{q}_\ell) \equiv [h_{i,\ell}(\mathbf{x},t,q_\ell)]_{m\times 1}$$

- **Transformed Vector SDE:**

$$\begin{aligned} d\mathbf{Y}(t) = (\mathbf{F}_t + (\mathbf{f}^T\nabla_x)\mathbf{F} + \frac{1}{2}(gg^T : \nabla_x\nabla_x)\mathbf{F})dt + ((gd\mathbf{W}(t))^T\nabla_x)\mathbf{F} \\ + \sum_\ell \int_{\mathcal{Q}} (\mathbf{F}(\mathbf{X}+\mathbf{h}_\ell,t) - \mathbf{F}(\mathbf{X},t))\mathcal{P}_\ell(\mathbf{dt},\mathbf{dq}_\ell). \end{aligned}$$

- **Vector Target Explicit SDE:**

$$d\mathbf{Y}(t) = \mathbf{C}_1(t)dt + C_2(t)d\mathbf{W}(t) + \sum_{\ell} \int_{\mathcal{Q}} \mathbf{C}_{3,\ell}(t, q_\ell)\mathcal{P}_\ell(\mathbf{dt}, \mathbf{dq}_\ell).$$

- **Original Coefficients:**

$$\begin{aligned}\mathbf{f}(\mathbf{x}, t) &= (\nabla_x \mathbf{F}^T)^{-T}(\mathbf{C}_1(t) - y_t \\ &\quad -\frac{1}{2}(\nabla_x \mathbf{F}^T)^{-T} C_2 C_2^T (\nabla_x \mathbf{F}^T)^{-1} : \nabla_x \nabla_x^T \mathbf{F});\end{aligned}$$

$$g(\mathbf{x}, t) = (\nabla_x \mathbf{F}^T)^{-T} C_2(t),$$

$$\mathbf{F}(\mathbf{x} + \mathbf{h}_\ell, t) = \mathbf{F}(\mathbf{x}, t) + \mathbf{C}_{3,\ell}(t, q_\ell) \quad \{\textit{note: left in implicit form}\}.$$

- **Vector Affine Transformation Example:**

$$\begin{aligned}\mathbf{F} &= A(t)\mathbf{x} + \mathbf{B}(t), \\ \mathbf{F}_t &= A'\mathbf{x} + \mathbf{B}', \\ (\nabla_x \mathbf{F}^T)^T &= A, \\ \mathbf{f}(\mathbf{x}, t) &= A^{-1}(\mathbf{C}_1(t) - A'\mathbf{x} - \mathbf{B}'), \\ g(\mathbf{x}, t) &= A^{-1} C_2(t), \\ \mathbf{h}_\ell(\mathbf{x}, t, q_\ell) &= A^{-1}\mathbf{C}_{3,\ell}(t, q_\ell).\end{aligned}$$

5.5 Exercises

1. Simulate $X(t)$ for the log-normally distributed jump-amplitude case with mean $\mu_j = \mathrm{E}[Q] = 0.28$ and variance $\sigma_j^2 = \mathrm{Var}[Q] = 0.15$ for the linear jump-diffusion SDE model (5.42) using $\mu_d(t) = 0.82\sin(2\pi t - 0.75\pi)$, $\sigma_d(t) = 0.88 - 0.44\sin(2\pi t - 0.75\pi)$ and $\lambda(t) = 8.0 - 1.82\sin(2\pi t - 0.75\pi)$, $N = 10000$ time-steps, $t_0 = 0$, $t_f = 1.0$, $X(0) - x_0$, for $k = 4$ random states, i.e., $\nu(t, Q) = \nu_0(Q) = \exp(Q) - 1$, with Q normally distributed. Plot the k sample states $X_j(t_i)$ for $j = 1 : k$, along with theoretical mean state path, $\mathrm{E}[X(t_i)]$ (5.49), and the sample mean state path, i.e., $M_x(t_i) = \sum_{j=1}^{k} X_j(t_i)/k$, all for $i = 1 : N + 1$.

   {*Hint: Modify the linear mark-jump-diffusion SDE simulator Example 5.25 with Appendix A MATLAB code C.15 and Corollary 5.9 for the discrete exponential expectation.* }

2. For the log-double-uniform jump distribution,

$$\phi_Q(q; t) \equiv \left\{ \begin{array}{ll} 0, & -\infty < q < a(t) \\ p_1(t)/|a|(t), & a(t) \le q < 0 \\ p_2(t)/b(t), & 0 \le q \le b(t) \\ 0, & b(t) < q < +\infty \end{array} \right\}, \tag{5.102}$$

where $p_1(t)$ is the probability of a negative jump and $p_2(t)$ is the probability of a posative jump on $a(t) < 0 \leq b(t)$, show that

(a) $\mathrm{E}_Q[Q] = \mu_j(t) = (p_1(t)a(t) + p_2(t)b(t))/2$;

(b) $\mathrm{Var}_Q[Q] = \sigma_j^2(t) = (p_1(t)a^2(t) + p_2(t)b^2(t))/3 - \mu_j^2(t)$;

(c) $E_Q\big[(Q-\mu_j(t))^3\big] = (p_1(t)a^3(t)+p_2(t)b^3(t))/4 - \mu_j(t)(3\sigma_j^2(t)+\mu_j^2(t))$;

(d) $\mathrm{E}[\nu(Q)] = \mathrm{E}[\exp(Q) - 1]$, where the answer needs to be derived.

3. Show that the Itô mean square limit for the integral of the product of two correlated mean-zero, dt-variance, differential diffusion processes, $dW_1(t)$ and $dW_2(t)$, symbolically satisfying the SDE,

$$dW_1(t)dW_2(t) \stackrel{\mathrm{dt}}{=} \rho(t)dt \,, \tag{5.103}$$

where

$$\mathrm{Cov}[\Delta W_1(t_i), \Delta W_2(t_i)] \simeq \rho(t_i)\Delta t_i$$

for sufficiently small Δt_i. Are there any modified considerations required if $\rho = 0$ or $\rho = \pm 1$? You may use the bivariate normal density in (B.146), boundedness Theorem B.59, Table B.1 of selected moments and nearby material of the Preliminaries Appendix B.

4. Finish the proof of Corollary 5.13 by showing the diffusion part using the techniques of Theorem 5.11 Eq. (5.53).

5. Prove the corresponding corollary for the variance of $X(t)$ from the solution of the linear SDE:

Corollary 5.27. ***Variance of X(t) for Linear SDE:***
Let $X(t)$ be the solution (5.45) with $\overline{\nu^2}(t) \equiv \mathrm{E}[\nu^2(t,Q)]$ of (5.42), then

$$\mathrm{Var}[dX(t)/X(t)] \stackrel{\mathrm{dt}}{=} \sigma_d^2(t) + \overline{\nu^2}(t)$$

and

$$\mathrm{Var}[X(t)] = \mathrm{E}^2[X(t)] \left(\exp\left(\int_{t_0}^{t} \mathrm{Var}[dX(s)/X(s)]ds \right) - 1 \right). \tag{5.104}$$

Be sure to state what extra conditions on processes and precision are needed that were not needed for proving Corollary 5.13 on $\mathrm{E}[X(t)]$.

6. Justify (5.94) for the covariance in multi-dimensions by giving the reasons for each step in the derivation. See the proof for (5.27).

Suggested References for Further Reading

- Çinlar, 1975 [55].
- Cont and Tankov, 2004 [59].
- Gihman and Skorohod, 1972 [94, Part 2, Chapter 2].
- Hanson, 1996 [108].
- Itô, 1951 [149].
- Kushner and Dupuis, 2001 [179].
- Øksendal and A. Sulem, 2005 [223].
- Snyder and Miller, 1991 [252, Chapter 4 and 5].
- Westman and Hanson, 1999 [276].
- Westman and Hanson, 2000 [277].
- Zhu and Hanson, 2006 [291].

Chapter 6

Stochastic Optimal Control - Stochastic Dynamic Programming

It was the owl that shriek'd, the fatal bellman,
Which gives the stern'st good-night.
—William Shakespeare (1564-1616) in *Macbeth*.

But the principal failing occurred in the sailing,
And the Bellman, perplexed and distressed,+
Said he had hoped, at least, when the wind blew due East,
That the ship would not travel due West!
—Lewis Carroll (1832-1898) in *The Bellman's Speech*.

6.1 Stochastic Optimal Control Problem

This main chapter introduces the optimal stochastic control problem. For many application systems, solving a SDE, or for that matter an ODE, to obtain its behavior is only part of the problem. The SDE is, in fact, a stochastic ordinary differential equation (SODE). Another, very significant part is finding out how to control the SDE or ODE as a model for controlling the application system.

Thus, the general jump-diffusion SDE (5.82) is reformulated with an additional process, the **vector control process** $\mathbf{U}(t) = [U_i(t)]_{n_u \times 1}$ on some n_u-dimensional control space $\mathcal{D}_u$,

$$d\mathbf{X}(t) \stackrel{\text{sym}}{=} \mathbf{f}(\mathbf{X}(t), \mathbf{U}(t), t)dt + g(\mathbf{X}(t), \mathbf{U}(t), t)d\mathbf{W}(t) + \int_{\mathcal{Q}} h(\mathbf{X}(t), \mathbf{U}(t), t, \mathbf{q})\boldsymbol{\mathcal{P}}(\mathbf{dt}, \mathbf{dq}; \mathbf{X}(t), \mathbf{U}(t), t) \,, \tag{6.1}$$

when $t_0 \leq t \leq t_f$ subject to a given initial state $\mathbf{X}(t_0) = \mathbf{x}_0$, where again $\mathbf{X}(t) = [X_i(t)]_{n_x \times 1}$ is the **vector state process** on some n_x-dimensional state space $\mathcal{D}_x$. The stochastic processes are the n_w-dimensional vector Wiener process or diffu-

sion process $\mathbf{W}(t) = [W_i(t)]_{n_w \times 1}$ and the n_p-dimensional vector Poisson process or jump process $\mathbf{P}(t; \mathbf{Q}, \mathbf{X}(t), \mathbf{U}(t), t) = [P_i(t; Q_i, \mathbf{X}(t), \mathbf{U}(t), t)]_{n_p \times 1}$, with IID jump-amplitude mark random vector $\mathbf{Q} = [Q_i]_{n_p \times 1}$ and Poisson random measure

$$\boldsymbol{\mathcal{P}}(\mathbf{dt}, \mathbf{dq}; \mathbf{X}(t), \mathbf{U}(t), t) = [\mathcal{P}_i(\mathbf{dt}, \mathbf{dq}; \mathbf{X}(t), \mathbf{U}(t), t)]_{n_p \times 1}.$$

The n_p-dimensional vector state-dependent compound Poisson process can also be defined as in Chapt. 5 in a zero-one law form,

$$\int_{\mathcal{Q}} h(\mathbf{X}(t), \mathbf{U}(t), t, \mathbf{q}) \boldsymbol{\mathcal{P}}(\mathbf{dt}, \mathbf{dq}; \mathbf{X}(t), \mathbf{U}(t), t)$$
$$\underset{zol}{\overset{dt}{=}} \left[\sum_{j=1}^{n_p} h_{i,j}(\mathbf{X}(t), \mathbf{U}(t), t, Q) dP_j(t; Q_j \mathbf{X}(t), \mathbf{U}(t), t) \right]_{n_x \times 1},$$

with

$$\mathrm{E}[d\mathbf{P}(t; \mathbf{Q}, \mathbf{X}(t), \mathbf{U}(t), t) | \mathbf{X}(t) = \mathbf{x}, \mathbf{U}(t) = \mathbf{u}] = \boldsymbol{\lambda}(t; \mathbf{x}, \mathbf{u}, t) dt,$$

and jump in the ith state component

$$[X_i](T_{j,k}) = h_{i,j}(\mathbf{X}(T_{j,k}^-), \mathbf{U}(T_{j,k}^-), T_{j,k}^-, Q_{j,k}),$$

where $\boldsymbol{\lambda}(t; \mathbf{x}, \mathbf{u}, t)$ is the jump rate vector and $T_{j,k}^-$ is the kth jump time of the jth differential Poisson process and $Q_{j,k}$ is the corresponding mark.

The coefficient functions are the $n_x \times 1$ plant function $\mathbf{f}(\mathbf{x}, \mathbf{u}, t)$, having the same dimension as the state $\mathbf{x}$, the $n_x \times n_w$ volatility function $g(\mathbf{x}, \mathbf{u}, t)$ or square root of the variance of the diffusion term, and the $n_x \times n_p$ jump amplitude of the jump term $h(\mathbf{x}, \mathbf{u}, t, \mathbf{Q})$, where $\mathbf{Q}$ is the underlying jump amplitude random mark process, the space part of the space-time Poisson process.

The optimization objective functional for a control formulation may be the combination of a final cost at time t_f and cumulative instantaneous costs, given the initial data $(\mathbf{x}_0, t_0)$. For instance,

$$V[\mathbf{X}, \mathbf{U}, t_f](\mathbf{x}_0, t_0) = \int_{t_0}^{t_f} C(\mathbf{X}(s), \mathbf{U}(s), s) ds + S(\mathbf{X}(t_f), t_f) \tag{6.2}$$

is a functional of the processes $\mathbf{X}(t)$ and $\mathbf{U}(t)$, where $C(\mathbf{x}, \mathbf{u}, t)$ is the scalar instantaneous or **running cost function** on the **time horizon** $(t_0, t_f]$ give the state at t_0 and $S(\mathbf{x}, t)$ is the **final cost function**; both are assumed continuous. This is the Bolza form of the objective. The objective $V[\mathbf{X}, \mathbf{U}, t_f](\mathbf{x}_0, t_0)$ is a functional of the state $\mathbf{X}$ and control process $\mathbf{U}$, i.e., a function of functions, while also dependent on the values of the initial data $(\mathbf{x}_0, t_0)$. The optimal control objective, in this case, is to minimize the expected total costs with respect to the control process on $(t_0, t_f]$. The feedback control of the multibody stochastic dynamical system (6.1) is illustrated in the block diagram displayed in Figure 6.1.

Prior to the optimization step, an averaging step, taking the conditional expectation, conditioned on some initial state, is **essential** to avoid the ill-posed problem

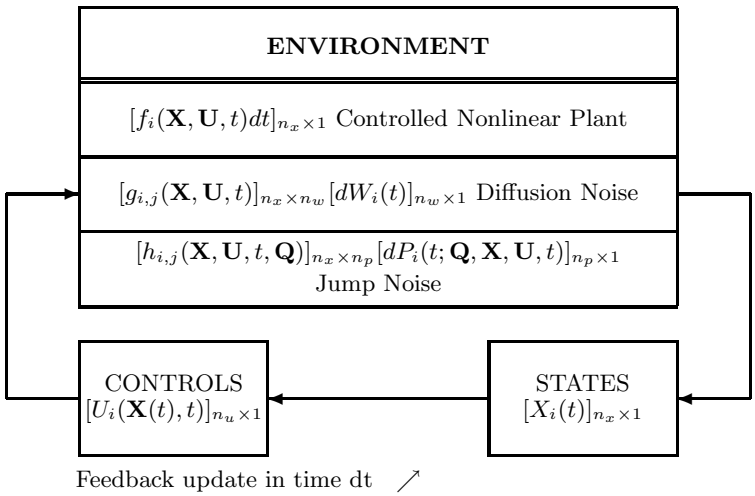


Figure 6.1. *Multibody Stochastic Dynamical System Under Feedback Control.*

of trying to optimize an uncertain, fluctuating objective. It is further assumed here that the running and terminal cost functions permit a unique minimum, subject to stochastic differential dynamics in the multi-dimensional jump-diffusion case (6.1). Hence, the optimal, expected cost for (6.2) is

$$v^*(\mathbf{x}_0,t_0) \equiv \min_{\boldsymbol{U}(t_0,t_f]}\Bigg[\underset{(\boldsymbol{W},\boldsymbol{P})(t_0,t_f]}{\mathrm{E}}\Big[V[\mathbf{X},\mathbf{U},t_f](\mathbf{x}_0,t_0) \Big| \mathbf{X}(t_0)=\mathbf{x}_0,\mathbf{U}(t_0)=\mathbf{u}_0\Big]\Bigg], \tag{6.3}$$

with the expectation preceding the minimization so that the minimization problem is better-posed by smoothing random fluctuations through averaging. In the optimization in (6.4), it is implicit that the stochastic dynamical system (6.1) is a constraint. The minimization over $\mathbf{U}(t_0,t_f]$ denotes the minimization over the control path $\mathbf{U}(t)$ for $t\in(t_0,t_f]$ and similarly the expectation over $\{W,P\}(t_0,t_f]$ denotes expectation over the joint stochastic pair $\{W(t),P(t)\}$ for $t\in(t_0,t_f]$.

Recall that the maximum problem, as in the maximization of profits, portfolio returns or utility, is an equivalent problem since

$$\max_{\mathbf{U}}[V[\mathbf{X},\mathbf{U},t_f](x_0,t_0)] = -\min_{\mathbf{U}}[-V[\mathbf{X},\mathbf{U},t_f](\mathbf{x}_0,t_0)]\,,$$

upon reversing the value. However, switching theoretical results from those for a minimum to get those of a maximum, basically requires just replacing the minimum function min for the maximum function max, along with replacing positive definiteness conditions for negative definite conditions, in the case of regular optima. For software optimization packages that are designed for minimizations, then the user

needs to use the negative of the function to be maximized and to take the negative of the final minimum output, for example, MATLAB's `fminsearch`.

In order to implement the dynamic part of dynamic programming, the fixed initial condition $\mathbf{X}(t_0) = \mathbf{x}_0$ for the SDE (6.1) needs to be replaced by a more arbitrary start, $\mathbf{X}(t) = \mathbf{x}$, so that the start can be analytically manipulated. This is a small but important step to produce a time-varying objective amenable to analysis. Hence, the optimal expected value as

$$v^*(\mathbf{x}, t) \equiv \min_{\boldsymbol{U}(t,t_f]} \left[\mathop{\mathrm{E}}_{(\boldsymbol{W},\boldsymbol{P})(t,t_f]} \left[V[\mathbf{X}, \mathbf{U}, t_f](\mathbf{x}, t) \middle| \mathbf{X}(t) = \mathbf{x}\,, \mathbf{U}(t) = \mathbf{u} \right] \right] . \quad (6.4)$$

Since the running cost integral vanishes when $t = t_f$, leaving only the terminal cost term conditioned on $\mathbf{X}(t_f) = \mathbf{x}$ and $\mathbf{U}(t_f) = \mathbf{u}$, a simple final condition for the optimal expected cost follows:

$$v^*(\mathbf{x}, t_f) = S(\mathbf{x}, t_f)\,, \quad (6.5)$$

for any $\mathbf{x}$ in the state domain $\mathcal{D}_x$, assuming that the terminal cost function $S(\mathbf{x}, t_f)$ is a deterministic function. This final condition is the first clue meaning that dynamic programming will use a backward program in time.

6.2 Bellman's Principle of Optimality

The basic assumption is that the optimization and expectation can be decomposed over increments in time. Bellman's Principle of Optimality can be systematically derived from optimization in time step proceeding backward from the final increment to the initial increment. Also, in the Markov processes case here, the independent increment properties of the Wiener and Poisson processes permit the decomposition of the expectation over time. This decomposition conveniently complements the decomposition of the optimization over time as in the deterministic case presented in Section A.4.

The semi-close-open time interval $(t, t_f]$ in the optimal expected cost formulation (6.4), given the state at time t, can be decomposed into disjoint increments $(t, t + \delta t]$ and $(t + \delta t, t_f]$ for fixed δt in $t < t + \delta t < t_f$. Symbolically, the **decomposition rules** are written:

Rules 6.1. ***Decomposition for Time, Integration, Expectation and Minimumization:***

- ***Time domain Decompostion into Subintervals:***

  $$(t, t_f] = (t, t + \delta t] + (t + \delta t, t_f],$$

  needs to be further decomposed for discrete approximations into sufficiently small increments Δt_i *for* $i = 1{:}n + 1$*, such that*

  $$t_{i+1} = t_i + \sum_{j=1}^{i} \Delta t_j,$$

$t_1 = t$, $t_\ell = t+\delta t$ *for some integer* $\ell \in [1, n+1]$, $t_{n+1} = t_f$, $\delta t_n = \max_i[\Delta t_i] \to 0$ *as* $n \to \infty$. *Recall that the approximation to the stochastic dynamics (6.1) is*

$$\mathbf{X}_{i+1} \simeq \mathbf{X}_i + \int_{t_i}^{t_i+\Delta t_i} d\mathbf{X}(s) \simeq \mathbf{X}_i + \mathbf{f}_i \Delta t_i + g_i \Delta \mathbf{W}_i + h_i \Delta \mathbf{P}_i,$$

for sufficiently small Δt_i, *where, for example,* $\mathbf{f}_i \equiv \mathbf{f}(\mathbf{X}_i, \mathbf{U}_i, t_i)$, *so that the change from* $\mathbf{X}_i$ *to* $\mathbf{X}_{i+1}$ *is due to the control* $\mathbf{U}_i$ *and random fluctuations* $(\Delta \mathbf{W}_i, \Delta \mathbf{P}_i)$ *determined from a prior stage.*

- ***Integration Additive Decomposition Rule:***

$$\int_t^{t_f} C(\mathbf{X}(s), \mathbf{U}(s), s)ds = \int_t^{t+\delta t} C(\mathbf{X}(s), \mathbf{U}(s), s)ds + \int_{t+\delta t}^{t_f} C(\mathbf{X}(s), \mathbf{U}(s), s)ds, \quad (6.6)$$

for the cumulative running costs by the regular additivity property of regular or Riemann-type integrals, or in terms of small increments in simplified notation. Let

$$V = \int_t^{t_f} Cds + S(\mathbf{X}(t_f), t_f) \simeq \sum_{i=1}^{n+1} \widehat{C}_i,$$

be the forward approximation, where $\widehat{C}_i \equiv C_i \Delta t_i = C(\mathbf{X}_i, \mathbf{U}_i, t_i)\Delta t_i$ *for* $i = 1 : n-1$ *and* $\widehat{C}_{n+1} \equiv S(\mathbf{X}(t_f), t_f) = S(X_{n+1}, t_{n+1}) = S_{n+1}$.

- ***Expectation Operator Multiplication Decomposition Rule:***

$$\overline{V} = \underset{(\boldsymbol{W},\boldsymbol{P})(t,t_f]}{\mathrm{E}}[V|\mathcal{C}(t)] = \underset{(\boldsymbol{W},\boldsymbol{P})(t,t+\delta t]}{\mathrm{E}}\left[\underset{(\boldsymbol{W},\boldsymbol{P})(t+\delta t,t_f]}{\mathrm{E}}[V|\mathcal{C}(t+\delta t)]\Big| \mathcal{C}(t)\right],$$

where V *is an objective function and* $\mathcal{C}(t) = \{\mathbf{X}(t), \mathbf{U}(t)\}$ *is the conditioning at time* t. *This decomposition relies on the corresponding decomposition of the Markov processes* $\mathbf{W}(t)$ *and* $\mathbf{P}(t; \mathbf{Q}, \mathbf{X}(t), \mathbf{U}(t), t)$ *into independent increments, so that the expectation over* $\{\mathbf{W}(s), \mathbf{P}(s)\}$ *for* $s \in (t, t_f]$ *is the product of expectation over* $\{\mathbf{W}(s), \mathbf{P}(s)\}$ *for* $s \in (t, t+\delta t]$ *and expectation over* $\{\mathbf{W}(r), \mathbf{P}(r)\}$ *for* $r \in (t+\delta t, t_f]$. *In order to compute the expectation over the path of a Markov process, we need to approximate the process by a sum of* n *independent increments for sufficiently large* n *to obtain sufficiently small* Δt_i *and then take the product of the expectations with respect to each of these independent increments, and finally taking the limit as* $n \to \infty$ *relying on mean square convergence in the result as in the first two chapters. In simple notation,*

$$\overline{V} = \mathrm{E}[V|\mathcal{C}(t)] \simeq \mathrm{E}\left[\sum_{i=1}^{n+1} \widehat{C}_i \Big| \mathbf{X}_1, \mathbf{U}_1\right],$$

where $\mathrm{E}\left[\widehat{C}_1\middle|\mathbf{X}_1,\mathbf{U}_1\right]\equiv\mathrm{E}_0\left[\widehat{C}_1\right]=\widehat{C}_1$ *since* $\widehat{C}_1=C(\mathbf{X}_1,\mathbf{U}_1,t_1)\Delta t_1$,

$$\mathrm{E}\left[\widehat{C}_2\middle|\mathbf{X}_1,\mathbf{U}_1\right]=\underset{(\boldsymbol{\Delta W}_1,\boldsymbol{\Delta P}_1)}{\mathrm{E}}\left[\widehat{C}_2\middle|\mathbf{X}_1,\mathbf{U}_1\right]\equiv\mathrm{E}_1\left[\widehat{C}_2\right]=\Pi_{j=0}^{1}\mathrm{E}_j\left[\widehat{C}\right],$$

$$\mathrm{E}\left[\widehat{C}_3\middle|\mathbf{X}_3,\mathbf{U}_3\right]=\mathrm{E}_1\left[\underset{(\boldsymbol{\Delta W}_2,\boldsymbol{\Delta P}_2)}{\mathrm{E}}\left[\widehat{C}_3\middle|\mathbf{X}_2,\mathbf{U}_2\right]\right]\equiv\Pi_{j=0}^{2}\mathrm{E}_j\left[\widehat{C}_3\right],$$

so in general,

$$\mathrm{E}\left[\widehat{C}_{i+1}\middle|\mathbf{X}_1,\mathbf{U}_1\right]=\Pi_{j=0}^{i}\mathrm{E}_j\left[\widehat{C}_{i+1}\right],$$

with

$$\mathrm{E}_j\left[\widehat{C}_{i+1}\right]\equiv\underset{(\boldsymbol{\Delta W}_j,\boldsymbol{\Delta P}_j)}{\mathrm{E}}\left[\widehat{C}_i\middle|\mathbf{X}_j,\mathbf{U}_j\right]$$

for $j=0:i$, $\mathrm{E}\left[\widehat{C}_{i+1}\right]=\widehat{C}_{i+1}$ *and finally,*

$$\overline{V}\simeq\sum_{i=1}^{n+1}\Pi_{j=0}^{i-1}\mathrm{E}_j\left[\widehat{C}_i\right]\longrightarrow\underset{(\boldsymbol{W},\boldsymbol{P})(t,t+\delta t]}{\mathrm{E}}\left[\int_t^{t+\delta t}Cds\right.$$
$$\left.+\underset{(\boldsymbol{W},\boldsymbol{P})(t+\delta t,t_f]}{\mathrm{E}}\left[\int_{t+\delta t}^{t_f}Cds+S(\mathbf{X}(t_f),t_f)\middle|(\mathbf{X},\mathbf{U})(t+\delta t)\right]\middle|(\mathbf{X},\mathbf{U})(t)\right],$$

as $n\to\infty$, *confirming the construction, assuming mean square convergence.*

- ***Minimization Operator Multiplication Decomposition Rule:***

$$\overline{V}^{\,*}=\min_{\boldsymbol{U}(t,t_f]}\left[\,\overline{V}\,\right]=\min_{\boldsymbol{U}(t,t+\delta t]}\left[\min_{\boldsymbol{U}(t+\delta t,t_f]}\left[\,\overline{V}\,\right]\right],\tag{6.7}$$

where $\overline{V}$ *is the expected value of an objective so that the decomposition rule is analogous to the use in deterministic dynamic programming. This decomposition depends on the reasonable heuristic idea that given a minimum on the later interval* $(t+\delta t,t_f]$, *taking the minimum of the given minimum over the small earlier interval* $(t,t+\delta t]$ *yields the minimum over the longer interval* $(t,t_f]$. *In terms of the small increments* (Δt_i) *construction,*

$$\overline{V}^{\,*}\simeq\sum_{i=1}^{n+1}\min_{\boldsymbol{U}(t,t_f]}\left[\Pi_{j=0}^{i-1}\mathrm{E}_j\left[\widehat{C}_i\right]\right]=\sum_{i=1}^{n+1}\left[\Pi_{j=0}^{i-1}\min_{\boldsymbol{U}_j}\mathrm{E}_j\left[\widehat{C}_i\right]\right]=\sum_{i=1}^{n+1}\Pi_{j=0}^{i-1}\mathrm{ME}_j\left[\widehat{C}_i\right]$$

where

$$\mathrm{ME}_0\equiv\min_{\boldsymbol{U}_1}\Big[\mathrm{E}_0\Big[\widehat{C}_0\Big|\,(\mathbf{X}_0,\mathbf{U}_0)\Big]\Big]$$

and

$$\mathrm{ME}_j\equiv\min_{\boldsymbol{U}_j}\Big[\mathrm{E}_j\Big[\widehat{C}_i\Big|\,\mathbf{X}_j,\mathbf{U}_j\Big]\Big]$$

for $j = 0 : i-1$. As $n \to \infty$ and $\delta t_n \to 0$, then

$$\overline{V}^* \to \min_{\boldsymbol{U}(t,t+\delta t]} \left[\mathop{\mathrm{E}}_{(\boldsymbol{W},\boldsymbol{P})(t,t+\delta t]} \left[\int_t^{t+\delta t} C ds + \min_{\boldsymbol{U}(t+\delta t,t_f]} \left[\mathop{\mathrm{E}}_{(\boldsymbol{W},\boldsymbol{P})(t+\delta t,t_f]} \left[\int_{t+\delta t}^{t_f} C ds + S(\mathbf{X}(t_f),t_f) \middle| (\mathbf{X},\mathbf{U})(t+\delta t) \right] \right] \middle| (\mathbf{X},\mathbf{U})(t) \right] \right].$$

The optimal decomposition seems to work for many examples. However, for empirical counterexamples, see Rust [240].

Thus, **optimal expected cost** (6.4) can be decomposed as follows:

$$\begin{aligned} v^*(\mathbf{x},t) &= \min_{\boldsymbol{U}(t,t+\delta t]} \left[\mathop{\mathrm{E}}_{(\boldsymbol{W},\boldsymbol{P})(t,t+\delta t]} \left[\int_t^{t+\delta t} C(\mathbf{X}(s),\mathbf{U}(s),s) ds \right.\right. \\ &\quad + \min_{\boldsymbol{U}(t+\delta t,t_f]} \left[\mathop{\mathrm{E}}_{(\boldsymbol{W},\boldsymbol{P})(t+\delta t,t_f]} \left[\int_{t+\delta t}^{t_f} C(\mathbf{X}(s),\mathbf{U}(s),s) ds + S(\mathbf{X}(t_f),t_f) \right.\right.\right. \\ &\quad \left.\left.\left. \middle| \{\mathbf{X}(t+\delta t), \mathbf{U}(t+\delta t)\} \right] \right] \left.\left. \middle| \mathbf{X}(t) = \mathbf{x}\,, \mathbf{U}(t) = \mathbf{u} \right] \right] \\ &= \min_{\boldsymbol{U}(t,t+\delta t]} \left[\mathop{\mathrm{E}}_{(\boldsymbol{W},\boldsymbol{P})(t,t+\delta t]} \left[\int_t^{t+\delta t} C(\mathbf{X}(s),\mathbf{U}(s),s) ds \right.\right. \\ &\quad \left.\left. + v^*(\mathbf{X}(t+\delta t), t+\delta t) \middle| \mathbf{X}(t) = \mathbf{x}\,, \mathbf{U}(t) = \mathbf{u} \right] \right], \end{aligned} \tag{6.8}$$

where the definition (6.4) for v^* has been reused with the arguments shifted by the time-step δt, since the inner part of the decomposition that is on $(t+\delta t, t_f]$ is precisely the definition of v^* in (6.4) but with arguments shifted from $(\mathbf{x}, t)$ to $(\mathbf{X}(t+dt), t+dt)$. Thus, Eq. (6.8) is a backward recursion relation for v^*. The subscript notation $\mathbf{U}(t, t+\delta t]$ under the min operator means that the minimum is with respect to $\mathbf{U}$ in the range $(t, t+\delta t]$, with similar subscript notation $\{\mathbf{W}, \mathbf{P}\}(t, t_f]$ for the expectation operator. Thus, we have formally derived the fundamental recursive formula of *stochastic dynamic programming*:

Lemma 6.2. ***Bellman's Principle of Optimality:***
Under the assumptions of the decomposition rules (6.7, 6.7, 6.6) and the properties of jump-diffusions,

$$\begin{aligned} v^*(\mathbf{x},t) &= \min_{\boldsymbol{U}(t,t+\delta t]} \left[\mathop{\mathrm{E}}_{(\boldsymbol{W},\boldsymbol{P})(t,t+\delta t]} \left[\int_t^{t+\delta t} C(\mathbf{X}(s),\mathbf{U}(s),s) ds \right.\right. \\ &\quad \left.\left. + v^*(\mathbf{X}(t+\delta t), t+\delta t) \middle| \mathbf{X}(t) = \mathbf{x}\,, \mathbf{U}(t) = \mathbf{u} \right] \right]. \end{aligned} \tag{6.9}$$

The argument of the minimum when it exists, within the control domain $\mathcal{D}_u$, is the optimal control $\mathbf{u}^* = \mathbf{u}^*(\mathbf{x}, t)$. Although the SDE is a forward differential

equation integrated forward from the initial condition, the optimal control problem is a backward general or functional equation integrated backward from the final time. The backward equation is quite basic, when one has a final objective, here optimal costs. Then the primary question is where to start initially to get that optimum. People do backward calculations all the time, such as when going to a scheduled meeting or a class, the meeting time is fixed and the problem is to estimate what time one should leave to get to the meeting. However, when economic decisions are made, the decision makers may not behave according to Bellman's principle of optimality according to the studies of Rust [240].

In general, **capital letters** are used for stochastic processes and **lower case letters** for conditioned or realized variables.

6.3 Hamilton-Jacobi-Bellman (HJB) Equation of Stochastic Dynamic Programming

Using the **Principle of Optimality** (6.9) and by taking the limit of small δt, replacing δt by dt, we can systematically derive the partial differential equation of stochastic dynamic programming, also called the stochastic **Hamilton-Jacobi-Bellman** (HJB) equation, for the general, multi-dimensional Markov dynamics case. From the increment form of the state differential $d\mathbf{X}(t) = \mathbf{X}(t+dt) - \mathbf{X}(t)$, we consider the expansion of the state argument

$$\mathbf{X}(t+dt) = \mathbf{X}(t) + d\mathbf{X}(t)$$

about $\mathbf{X}(t)$ for small $d\mathbf{X}(t)$ and about the explicit time argument $t+dt$ about t in the limit of small time increments dt, using an extension of Taylor approximations extended to include discontinuous (i.e, Poisson) and non-smooth (i.e., Wiener) processes. Sufficient differentiability of the optimal value function $v^*(\mathbf{x}, t)$, at least to first order in time and second order in state, is assumed except when its state argument has Poisson jumps. The spirit of the derivation of the multi-dimensional chain rule (5.98) is applied to the Principle of Optimality (6.9), except that the mean square limit substitution for the bilinear Wiener $W_i(t)W_j(t)$ process is not needed here because of the pre-optimization expectation operation. Then neglecting $\mathrm{o}(dt)$ terms as $dt \to 0^+$ (strictly, we are really working with finite increments δt) and substituting for the conditioning on $\mathbf{X}(t)$ and $\mathbf{U}(t)$, an intermediate reduction of the optimal expected value is

$$\begin{aligned} v^*(\mathbf{x},t) \stackrel{\mathrm{dt}}{=} \min_{\mathbf{u}} \Bigg[\underset{(\boldsymbol{dW},\boldsymbol{dP})(t)}{\mathrm{E}} \Big[& C(\mathbf{x},\mathbf{u},t)dt + v^*(\mathbf{x},t) + v_t^*(\mathbf{x},t)dt \\ & + \nabla_{\mathbf{x}}^\top [v^*](\mathbf{x},t) \cdot (\mathbf{f}(\mathbf{x},\mathbf{u},t)dt + g(\mathbf{x},\mathbf{u},t)d\mathbf{W}(t)) \\ & + \frac{1}{2} d\mathbf{W}^\top(t) g^\top(\mathbf{x},\mathbf{u},t) \nabla_{\mathbf{x}} [\nabla_{\mathbf{x}}^\top [v^*]](\mathbf{x},t)(g(\mathbf{x},\mathbf{u},t)d\mathbf{W}(t)) \\ & + \sum_{j=1}^{n_p} \int_{\mathcal{Q}} \Big(v^*(\mathbf{x} + \widehat{\boldsymbol{h}}_j(\mathbf{x},\mathbf{u},t,q_j),t) - v^*(\mathbf{x},t) \Big) \mathcal{P}_j(\mathbf{dt}, \mathbf{dq}_j; \mathbf{x},\mathbf{u},t) \Big] \Bigg], \end{aligned} \tag{6.10}$$

where it has been assumed that the random mark variables $Q_j = q_j$ are pair-wise independently distributed and the jump amplitude is separable in the marks. So

$$h(\mathbf{x}, \mathbf{u}, t, \mathbf{q}) = [h_{i,j}(\mathbf{x}, \mathbf{u}, t, q_j)]_{n_x \times n_p} , \tag{6.11}$$

with a corresponding multiplicative factoring of the Poisson random measure. Recall from Chapter 5 (5.100) that the jth vector component of the jump amplitude is

$$\widehat{\boldsymbol{h}}_j(\mathbf{x}, \mathbf{u}, t, q_j) \equiv [h_{i,j}(\mathbf{x}, \mathbf{u}, t, q_j)]_{n_x \times 1} , \tag{6.12}$$

for $j = 1 : n_p$, corresponding to the jth Poisson process

$$dP_j(t; \mathbf{x}, \mathbf{u}, t) = \int_{\mathcal{Q}} \mathcal{P}_j(\mathbf{dt}, \mathbf{dq}_j; \mathbf{x}, \mathbf{u}, t) ,$$

in terms of the jth Poisson mark-time random measure $\mathcal{P}_j$. Note that the first t argument of dP_j is the time implicit to the Poisson process, while the second t argument is an explicit time corresponding to the implicit state and control parametric dependence.

The next step is to take the conditional expectation over the now isolated differential Wiener and Poisson processes, but done by expanding them in components to facilitate understanding of the step and suppressing some arguments for simplicity,

$$\begin{aligned} v^*(\mathbf{x}, t) \stackrel{\mathrm{dt}}{=} & v^*(\mathbf{x}, t) + v_t^*(\mathbf{x}, t)dt + \min_{\boldsymbol{u}} [C(\mathbf{x}, \mathbf{u}, t)dt \\ & + \nabla_{\mathbf{x}}^\top [v^*](\mathbf{x}, t) \cdot \left(\mathbf{f}(\mathbf{x}, \mathbf{u}, t)dt + \sum_{i=1}^{n_w} g_i(\mathbf{x}, \mathbf{u}, t) \mathrm{E}_{dW_i}[dW_i(t)] \right) \\ & + \frac{1}{2} \sum_{i=1}^{n_w} \sum_{j=1}^{n_w} \mathrm{E}_{dW_i, dW_j} \left[dW_i(t) dW_j(t) \right] \left[g^\top(\mathbf{x}, \mathbf{u}, t) \nabla_{\mathbf{x}} [\nabla_{\mathbf{x}}^\top [v^*]] g(\mathbf{x}, \mathbf{u}, t) \right]_{i,j} \\ & + \sum_{j=1}^{n_p} \int_{\mathcal{Q}} \Big(v^*(\mathbf{x} + \widehat{\boldsymbol{h}}_j(\mathbf{x}, \mathbf{u}, t, q_j), t) - v^*(\mathbf{x}, t) \Big) \mathrm{E}_{\mathcal{P}_j} \left[\mathcal{P}_j(\mathbf{dt}, \mathbf{dq}_j; \mathbf{x}, \mathbf{u}, t) \right] \Big] \end{aligned}$$

$$\begin{aligned} \stackrel{ind}{=\!=\!=}_{inc} & v^*(\mathbf{x}, t) + v_t^*(\mathbf{x}, t)dt + \min_{\boldsymbol{u}} \Big[C(\mathbf{x}, \mathbf{u}, t)dt + \nabla_{\mathbf{x}}^\top [v^*](\mathbf{x}, t) \left(\mathbf{f}(\mathbf{x}, \mathbf{u}, t)dt + 0 \right) \\ & + \frac{1}{2} \sum_{i=1}^{n_w} \left[1 + \sum_{j=1}^{n_w} \rho_{i,j}(1 - \delta_{i,j}) \right] \left[g^\top(\mathbf{x}, \mathbf{u}, t) \nabla_{\mathbf{x}} \left[\nabla_{\mathbf{x}}^\top [v^*] \right](\mathbf{x}, t) g(\mathbf{x}, \mathbf{u}, t) \right]_{i,j} dt \\ & + \sum_{j=1}^{n_p} \lambda_j(t; \mathbf{x}, \mathbf{u}, t) \int_{\mathcal{Q}} \Big(v^*(\mathbf{x} + \widehat{\boldsymbol{h}}_j(\mathbf{x}, \mathbf{u}, t, q_j), t) - v^*(\mathbf{x}, t) \Big) \\ & \cdot \Phi_{Q_j}(\mathbf{dq}_j; \mathbf{x}, \mathbf{u}, t)dt \Big] , \end{aligned} \tag{6.13}$$

where we have used the expectations

$$\mathrm{E}[dW_i(t)] = 0, \quad \mathrm{E}[dW_i(t) dW_i(t)] = (\delta_{i,j} + \rho_{i,j}(1 - \delta_{i,j})dt$$

with correlation coefficient $\rho_{i,j}$ and

$$\begin{aligned}\mathrm{E}[\mathcal{P}_j(\mathbf{dt},\mathbf{dq}_j;\mathbf{x},\mathbf{u},t)] &= \lambda_j(t;\mathbf{x},\mathbf{u},t)dt\Phi_{Q_j}(\mathbf{dq}_j;\mathbf{x},\mathbf{u},t)\\ &= \lambda_j(t;\mathbf{x},\mathbf{u},t)\phi_{Q_j}(q_j;\mathbf{x},\mathbf{u},t)dq_j dt\,.\end{aligned}$$

Also, with sufficiently small dt, $\mathbf{U}(t,t+dt]$ has been replaced by the conditioned control vector $\mathbf{u}$ at t.

Note that the $v^*(\mathbf{x},t)$ value on both sides of the equation cancel and then the remaining common multiplicative factors of dt also cancel, so the **HJB equation** has been derived for this general case:

Theorem 6.3. ***Hamilton-Jacobi-Bellman Equation (HJBE) for Stochastic Dynamic Programming (SDP)***
If $v^*(\mathbf{x},t)$ is twice differentiable in $\mathbf{x}$ and once differentiable in t, while the operator decomposition rules (6.7-6.6) are valid, then

$$0 = v_t^*(\mathbf{x},t) + \min_{\mathbf{u}}\left[\mathcal{H}(\mathbf{x},\mathbf{u},t)\right] \equiv v_t^*(\mathbf{x},t) + \mathcal{H}^*(\mathbf{x},t) \tag{6.14}$$

where the ***Hamiltonian*** *(technically, a pseudo-Hamiltonian) functional is given by*

$$\begin{aligned}\mathcal{H}(\mathbf{x},\mathbf{u},t) \equiv\; & C(\mathbf{x},\mathbf{u},t) + \nabla_{\mathbf{x}}^{\top}[v^*](\mathbf{x},t)\cdot\mathbf{f}(\mathbf{x},\mathbf{u},t)\\ & + \frac{1}{2}\left(gR'g^{\top}\right)(\mathbf{x},\mathbf{u},t) : \nabla_{\mathbf{x}}\left[\nabla_{\mathbf{x}}^{\top}[v^*]\right](\mathbf{x},t)\\ & + \sum_{j=1}^{n_p}\lambda_j(t;\mathbf{x},\mathbf{u},t)\int_{\mathcal{Q}}\left[v^*\left(\mathbf{x}+\widehat{\boldsymbol{h}}_j(\mathbf{x},\mathbf{u},t,q_j),t\right) - v^*(\mathbf{x},t)\right]\\ & \cdot\phi_{Q_j}(q_j;\mathbf{x},\mathbf{u},t)dq_j\,,\end{aligned} \tag{6.15}$$

where the correlation modified indentity R' is defined in (5.95) as

$$R' \equiv \left[\delta_{i,j} + \rho_{i,j}(1-\delta_{i,j})\right]_{n_w\times n_w}\,, \tag{6.16}$$

and where the correlation coefficient between i and j components is

$$\rho_{i,j}dt = \mathrm{Cov}[dW_i(t), dW_j(t)]\,, \tag{6.17}$$

provided $j\neq i$ for $i,j = 1:n_x$. The double-dot product $A:B$ is defined in (5.99).
The optimal control, if it exists, is given by

$$\mathbf{u}^* = \mathbf{u}^*(\mathbf{x},t) = \underset{\mathbf{u}}{\mathrm{argmin}}\;\left[\mathcal{H}(\mathbf{x},\mathbf{u},t)\right]\,, \tag{6.18}$$

subject to any control constraints.

This **HJB equation** (6.14) is no ordinary PDE, but but has the following properties or attributes:

Properties 6.4.

- *The HJBE is a functional PDE or PIDE due to the presence of the minimum operator* min *and the Poisson integral term (the last term) with steps in the state argument of the optimal value function v^* due to the jump amplitude.*

- *The HJBE is a scalar valued equation, but has a $(nu+1)$-**dimensional solution** consisting of the scalar optimal value function $v^* = v^*(\mathbf{x},t)$ and the optimal control vector $\mathbf{u}^* = \mathbf{u}^*(\mathbf{x},t)$ as well. These **dual solutions** are generally tightly coupled in functional dependence. In general, this tight coupling requires a number of iterations between v^* and $\mathbf{u}^*$ to obtain a reasonable approximation to the $(nu+1)$-dimensional solution. However, it should be noted that the optimal control $\mathbf{u}(\mathbf{x},t)$ in (6.18) is deterministic and if the $\mathbf{x}$ dependence is genuine then it is also feedback optimal control. In fact, the HJB equation is a deterministic equation as well.*

- *A further complication in this functional PDE or PIDE is that the HJB equation (6.14) has **global state dependence** due to the Poisson jump functional integral term, whereas the HJB equation for purely Gaussian or Wiener processes is essentially a diffusion equation that has only **local state dependence** since it depends only on the values $v^*(\mathbf{x},t)$, $\mathbf{u}^*(\mathbf{x},t)$, the gradient vector $\nabla_{\mathbf{x}}[v^*](\mathbf{x},t)$, and the Hessian matrix of 2nd order derivatives $\nabla_{\mathbf{x}}[\nabla_{\mathbf{x}}^\top[v^*]](\mathbf{x},t)$ at $(\mathbf{x},t)$. Contrast this with the random noise case including the Poisson random measure disturbance, with local dependence at $\mathbf{x}$, but global dependence on a range of points at $\mathbf{x} + \widehat{\boldsymbol{h}}_j(\mathbf{x},\mathbf{u},t,q_j)$ depending on the Poisson mark distribution.*

While letting $C^*(\mathbf{x},t) \equiv C(\mathbf{x},\mathbf{u}^*,t)$, $\mathbf{f}^*(\mathbf{x},t) \equiv \mathbf{f}(\mathbf{x},\mathbf{u}^*,t)$, $g^*(\mathbf{x},t) \equiv g(\mathbf{x},\mathbf{u}^*,t)$, $\widehat{\boldsymbol{h}}_j^*(\mathbf{x},t,q_j) \equiv \widehat{\boldsymbol{h}}_j(\mathbf{x},\mathbf{u}^*,t,q_j)$, and so forth for all control-dependent functions, then the HJB equation (HJBE) takes the form of a backward parabolic partial differential equation except that it has an additional integral term:

$$\begin{aligned} 0 &= v_t^*(\mathbf{x},t) + \mathcal{H}(\mathbf{x},\mathbf{u}^*(\mathbf{x},t),t) \\ &= v_t^*(\mathbf{x},t) + C^*(\mathbf{x},t) + \nabla_{\mathbf{x}}^\top[v^*](\mathbf{x},t)\cdot\mathbf{f}^*(\mathbf{x},t) \\ &\quad + \frac{1}{2}\left(g^* R\, {g^*}^\top\right)(\mathbf{x},t) : \nabla_{\mathbf{x}}\left[\nabla_{\mathbf{x}}^\top[v^*]\right](\mathbf{x},t) \\ &\quad + \sum_{j=1}^{n_p} \lambda_j^*(t;\mathbf{x},t)\int_{\mathcal{Q}} \Delta_j[v^*](\mathbf{x},t,q_j)\phi_{Q_j}^*(q_j;\mathbf{x},t)dq_j\ , \end{aligned} \tag{6.19}$$

where the jth jump increment is defined as

$$\Delta_j[v^*](\mathbf{x},t,q_j) \equiv v^*\left(\mathbf{x} + \widehat{\boldsymbol{h}}_j^*(\mathbf{x},t,q_j),t\right) - v^*(\mathbf{x},t) \tag{6.20}$$

and the double-dot product $(A:B)$ is defined in (5.99). The final condition is given by $v^*(\mathbf{x},t_f) = S(\mathbf{x},t_f)$.

The Hamilton-Jacobi-Bellman name of the equation comes from the fact that Bellman [25, 26] was the founding developer of dynamic programming and the fact

that the general evolution equation, $v_t^*(\mathbf{x},t) + \mathcal{H}^*(\mathbf{x},t) = 0$, is called a Hamilton-Jacobi equation and where $\mathcal{H}(\mathbf{x},\mathbf{u},t)$ is like a classical Hamiltonian. Sometimes, the HJB equation (6.14) is called simply the Bellman equation, or the stochastic dynamic programming equation or the PDE of stochastic dynamic programming, or in particular, the PIDE of stochastic dynamic programming where PIDE denotes a partial integral differential equation).

6.4 Linear Quadratic Jump-Diffusion (LQJD) Problem

The linear quadratic jump-diffusion (LQJD) problem is also called a linear quadratic Gaussian Poisson (LQGP) problem or jump linear quadratic Gaussian (JLQG) problem. The Markov property of the jump-diffusion processes described in this book leads to an analogous dynamic programming formulation to dynamic programming for deterministic processes as in the deterministic linear quadratic (LQ) problem of Subsection A.4.4. In this chapter, the LQJD problem is presented in more generality than in Chapter A.

The linear quadratic problem in both state and control leads to a quadratic decomposition of the optimal value function with respect to the state and a linear or feedback decomposition of the optimal control. However, first the LQJD problem is examined for a special case that is linear quadratic in control only to show how much an advantage is gained by the control dependence alone. For many applications it is not appropriate to have the problem linear quadratic in the state.

6.4.1 LQJD in Control Only (LQJD/U) Problem

A general variant of the LQJD problem is the LQJD/U problem that is LQJD in control only. Just having a control problem linear quadratic in control retains an important feature of the full linear quadratic control problem in that the optimal control can be solved for exactly in terms of the optimal value, even though the state decomposition property does not follow. The restricted linear quadratic problem in the control only will be treated first to examine how far the analysis can be taken before treating the full linear quadratic problem in the state and the control. In many control problems, the state dependence of the plant function $\mathbf{f}(\mathbf{x},\mathbf{u},t)$ is dictated by the application and may be significantly nonlinear, but the control dependence of the dynamics is up to the control designer who might chose to make the control simple, e.g., linear, so that the control process will be manageable for the control manager. Hence, the LQ problem in control only, may be more appropriate for some applications. In the past, linear systems were preferred since linear methods were well-known, but now nonlinear methods and problems have become more prevalent as we try to make more realistic models for applications.

Let the jump-diffusion linear quadratic model, in the control only, be given with the plant function for the deterministic or non-noise dynamics term,

$$\mathbf{f}(\mathbf{x},\mathbf{u},t) = \mathbf{f}_0(\mathbf{x},t) + f_1(\mathbf{x},t)\mathbf{u}\,, \tag{6.21}$$

with the diffusion term,

$$g(\mathbf{x}, \mathbf{u}, t) = g_0(\mathbf{x}, t) \,, \tag{6.22}$$

assumed control-independent for simplicity, with a jump term decomposition corresponding to independent sources of n_p-type jumps

$$h(\mathbf{x}, \mathbf{u}, t, \mathbf{q}) = h_0(\mathbf{x}, t, \mathbf{q}) = [h_{0,i,j}(\mathbf{x}, t, q_j)]_{n_x \times n_p} \,, \tag{6.23}$$

also assumed control-independent along with the very simplified Poisson noise

$$d\mathbf{P}(t; \mathbf{Q}, \mathbf{x}, \mathbf{u}, t) = d\mathbf{P}(t; \mathbf{Q}, \mathbf{x}, \mathbf{u}, t) \,, \quad \mathrm{E}[d\mathbf{P}(t; \mathbf{Q}, \mathbf{x}, \mathbf{u}, t)] = \boldsymbol{\lambda}(t; \mathbf{x}, \mathbf{u}, t)dt \tag{6.24}$$

and finally with the quadratic running cost function,

$$C(\mathbf{x}, \mathbf{u}, t) = C_0(\mathbf{x}, t) + \mathbf{C}_1^\top(\mathbf{x}, t)\mathbf{u} + \frac{1}{2}\mathbf{u}^\top C_2(\mathbf{x}, t)\mathbf{u} \,. \tag{6.25}$$

It is assumed that all right hand side coefficients are commensurate in multiplication and that the product is the same type at that on the left hand side. A crucial assumption in case of a minimum objective is that the quadratic control $C_2(\mathbf{x}, t)$ is positive definite, but $C_2(\mathbf{x}, t)$ can be assumed to be symmetric without loss of generality by the symmetric property of quadratic forms (B.135).

Thus, the pseudo-Hamiltonian is quadratic in the control,

$$\mathcal{H}(\mathbf{x}, \mathbf{u}, t) = \mathcal{H}_0(\mathbf{x}, t) + \boldsymbol{\mathcal{H}}_1^\top(\mathbf{x}, t)\mathbf{u} + \frac{1}{2}\mathbf{u}^\top \mathcal{H}_2(\mathbf{x}, t)\mathbf{u} \,, \tag{6.26}$$

where the scalar coefficient is

$$\begin{aligned} \mathcal{H}_0(\mathbf{x}, t) = &\left[C_0 + \mathbf{f}_0^\top \nabla_{\mathbf{x}}[v^*] + \frac{1}{2} g_0 g_0^\top : \nabla_{\mathbf{x}}[\nabla_{\mathbf{x}}[v^*]] \right](\mathbf{x}, t) \\ &+ \sum_{j=1}^{n_p} \lambda_j(t; \mathbf{x}, t) \int_{\mathcal{Q}_j} \Delta_j[v^*](\mathbf{x}, t, q_j)\phi_{\mathcal{Q}_j}(q_j)dq_j \,, \end{aligned} \tag{6.27}$$

where the double-dot product (5.99) is $GG^\top : A = \mathrm{Trace}[G^\top AG]$, while the jump increment is

$$\Delta_j[v^*](\mathbf{x}, t, q_j) \equiv v^* \left(\mathbf{x} + \widehat{\boldsymbol{h}}_j(\mathbf{x}, t, q_j), t \right) - v^*(\mathbf{x}, t) \,,$$

the linear control coefficient n_u-vector is

$$\boldsymbol{\mathcal{H}}_1(\mathbf{x}, t) = \left[\mathbf{C}_1 + f_1^\top \nabla_{\mathbf{x}}[v^*] \right](\mathbf{x}, t) \,, \tag{6.28}$$

and the quadratic control coefficient $n_u \times n_u$-matrix is simply

$$\mathcal{H}_2(\mathbf{x}, t) = C_2(\mathbf{x}, t) \,, \tag{6.29}$$

where $\mathcal{H}_2(\mathbf{x}, t)$ is assumed to be symmetric along with $C_2(\mathbf{x}, t)$. If the minimum cost is the objective, then $\mathcal{H}_2(\mathbf{x}, t)$ is positive definite since $C_2(\mathbf{x}, t)$ is assumed to be positive definite.

Thus, in search of a regular control minimum, the critical points of the pseudo-Hamiltonian $\mathcal{H}(\mathbf{x},\mathbf{u},t)$ is considered by examining the zeros of its gradient,

$$\nabla_{\mathbf{u}}[\mathcal{H}](\mathbf{x},\mathbf{u},t) = \boldsymbol{\mathcal{H}}_1(\mathbf{x},t) + \mathcal{H}_2(\mathbf{x},t)\mathbf{u} = \mathbf{0}\,, \tag{6.30}$$

yielding the regular control,

$$\begin{aligned}\mathbf{u}^{(\mathrm{reg})}(\mathbf{x},t) &= -\mathcal{H}_2^{-1}(\mathbf{x},t)\boldsymbol{\mathcal{H}}_1(\mathbf{x},t)\\ &= -C_2^{-1}(\mathbf{x},t)\left(\mathbf{C}_1 + f_1^{\top}\nabla_{\mathbf{x}}[v^*]\right)(\mathbf{x},t)\,,\end{aligned} \tag{6.31}$$

with the existence of the inverse being guaranteed by positive definiteness. The fact that the regular control can be solved for exactly in terms of the optimal value $v^*(\mathbf{x},t)$ is a major benefit of having an LQJD problem that is just quadratic in the control. If the usual LQ assumption it made that the control is unconstrained, then the regular control is also the optimal control:

$$\mathbf{u}^*(\mathbf{x},t) = \mathbf{u}^{(\mathrm{reg})}(\mathbf{x},t) \tag{6.32}$$

and the optimal Hamiltonian using (6.31) is

$$\begin{aligned}\mathcal{H}^*(\mathbf{x},t) &\equiv \mathcal{H}(\mathbf{x},\mathbf{u}^*,t)\\ &= \left[\mathcal{H}_0 - \boldsymbol{\mathcal{H}}_1^{\top}\mathcal{H}_2^{-1}\boldsymbol{\mathcal{H}}_1 + \frac{1}{2}\boldsymbol{\mathcal{H}}_1^{\top}\mathcal{H}_2^{-\top}\mathcal{H}_2\mathcal{H}_2^{-1}\boldsymbol{\mathcal{H}}_1\right](\mathbf{x},t)\\ &= \left[\mathcal{H}_0 - \frac{1}{2}\boldsymbol{\mathcal{H}}_1^{\top}\mathcal{H}_2^{-1}\boldsymbol{\mathcal{H}}_1\right](\mathbf{x},t)\,,\end{aligned} \tag{6.33}$$

where by symmetry the inverse transpose $\mathcal{H}_2^{-\top} = \mathcal{H}_2^{-1}$. Since the difference of the quadratic $\mathcal{H}$ in control from the designated minimum using the Taylor approximation form and the critical condition (6.30) is

$$\begin{aligned}\mathcal{H}(\mathbf{x},\mathbf{u},t) - \mathcal{H}^*(\mathbf{x},t) &= \mathcal{H}_0 - \mathcal{H}^*(\mathbf{x},t) + (\mathbf{u}-\mathbf{u}^*)^{\top}\nabla_{\mathbf{u}}[\mathcal{H}](\mathbf{x},\mathbf{u}^*,t)\\ &\quad + \frac{1}{2}(\mathbf{u}-\mathbf{u}^*)^{\top}\nabla_{\mathbf{u}}[\nabla_{\mathbf{u}}^{\top}[\mathcal{H}]](\mathbf{x},\mathbf{u}^*,t)(\mathbf{u}-\mathbf{u}^*)\\ &= \frac{1}{2}\boldsymbol{\mathcal{H}}_1^{\top}\mathcal{H}_2^{-1}\boldsymbol{\mathcal{H}}_1 + \frac{1}{2}(\mathbf{u}-\mathbf{u}^*)^{\top}\mathcal{H}_2(\mathbf{u}-\mathbf{u}^*)\\ &= \frac{1}{2}\left(\boldsymbol{\mathcal{H}}_1^{\top}\mathcal{H}_2^{-1}\boldsymbol{\mathcal{H}}_1\right)(\mathbf{x},t) + \frac{1}{2}(\mathbf{u}-\mathbf{u}^*)^{\top}\mathcal{H}_2(\mathbf{x},t)(\mathbf{u}-\mathbf{u}^*)\\ &\geq \frac{1}{2}\left(\boldsymbol{\mathcal{H}}_1^{\top}\mathcal{H}_2^{-1}\boldsymbol{\mathcal{H}}_1\right)(\mathbf{x},t) \geq 0\,,\end{aligned} \tag{6.34}$$

it is always possible to solve the optimal control in the minimum problem if $C_2(\mathbf{x},t)$ and thus $\mathcal{H}_2(\mathbf{x},t)$ are symmetric, positive definite. This corresponds to the minimum principle discussed for deterministic optimal control problems in Chapter A.

Within the generality of this linear quadratic problem in control only, the optimal control will generally be nonlinear in the state, so the corresponding HJB equation,

$$v_t^*(\mathbf{x},t) + \mathcal{H}^*(\mathbf{x},t) = 0\,, \tag{6.35}$$

will be highly nonlinear in the state, with $\mathcal{H}^*(\mathbf{x},t)$ given by (6.33) and coefficients (6.27, 6.28, 6.29). This requires careful solution by numerical PDE or PIDE methods or the computational methods of Chapter 8.

These LQJD/U derived results are summarized in the following theorem:

Theorem 6.5. ***LQJD/U Equations:***
Let the problem be the LQJD in control only problem, so that the deterministic plant function $\mathbf{f}(\mathbf{x},\mathbf{u},t)$ *is linear in the control as given in (6.21), the coefficient* $g(\mathbf{x},\mathbf{u},t)$ *of the Wiener process* $d\mathbf{W}(t)$ *is given in (6.22), the jump amplitude* $h(\mathbf{x},\mathbf{u},t,\mathbf{q})$ *of the Poisson jump process* $d\mathbf{P}(t;\mathbf{Q},\mathbf{x},t)$ *is given by (6.23), and the quadratic running cost* $C(\mathbf{x},\mathbf{u},t)$ *is given in (6.25).*

Then, the Hamiltonian $\mathcal{H}(\mathbf{x},\mathbf{u},t)$ *is a quadratic in the control (6.26) with coefficients* $\{\mathcal{H}_0(\mathbf{x},t),\boldsymbol{\mathcal{H}}_1(\mathbf{x},t),\mathcal{H}_2(\mathbf{x},t)\}$ *given in (6.27, 6.28, 6.29), respectively. The optimal control vector, in absence of control constraints, has the linear feedback control form,*

$$\mathbf{u}^*(\mathbf{x},t)=\mathbf{u}^{(\mathrm{reg})}(\mathbf{x},t)=-C_2^{-1}(\mathbf{x},t)\left[\mathbf{C}_1+f_1^\top\nabla_{\mathbf{x}}[v^*]\right](\mathbf{x},t)\,, \tag{6.36}$$

as long as the quadratic control coefficient $C_2(\mathbf{x},t)$ *is positive definite in case of a minimum expected objective and in absence of constraints on the control. Assuming that an optimal value* $v^*(\mathbf{x},t)$ *solution exists, then* $v^*(\mathbf{x},t)$ *satisfies the Hamilton Jacobi Bellman equation,*

$$v_t^*(\mathbf{x},t)+\left(\mathcal{H}_0-\frac{1}{2}\boldsymbol{\mathcal{H}}_1^\top\mathcal{H}_2^{-1}\boldsymbol{\mathcal{H}}_1\right)(\mathbf{x},t)=0\,. \tag{6.37}$$

The solution $v^*(\mathbf{x},t)$ *is subject to the final condition*

$$v^*(\mathbf{x},t_f)=S(\mathbf{x},t_f)\,, \tag{6.38}$$

and any necessary boundary conditions.

For solutions of LQJD/U problems, computational methods are quite essential; see Hanson's 1996 chapter [108] or Chapter 8.

6.4.2 LLJD/U or the Case $\mathbf{C_2\equiv 0}$:

If the quadratic cost coefficient $C_2(\mathbf{x},t)\equiv 0$, then

$$\mathcal{H}(\mathbf{x},\mathbf{u},t)=\mathcal{H}_0(\mathbf{x},t)+\boldsymbol{\mathcal{H}}_1^\top(\mathbf{x},t)\mathbf{u}\,, \tag{6.39}$$

the linear linear jump-diffusion (LLJD/U) problem in control only. The minimum with respect to the control depends on the linear cost coefficient

$$\mathcal{H}^*(\mathbf{x},t)=\min_{\mathbf{u}}\left[\mathcal{H}_0(\mathbf{x},t)+\boldsymbol{\mathcal{H}}_1^\top(\mathbf{x},t)\mathbf{u}\right]=\mathcal{H}_0(\mathbf{x},t)+\min_{\mathbf{u}}\left[\boldsymbol{\mathcal{H}}_1^\top(\mathbf{x},t)\mathbf{u}\right]\,. \tag{6.40}$$

Since this is a problem of linear or singular control, it makes sense only if the control is constrained, e.g., component-wise constraints,

$$U_i^{(\min)}\le u_i\le U_i^{(\max)}\,. \tag{6.41}$$

For this type of constraint the minimum is separable by component and the optimal control ia a n_u-dimensional bang-bang control

$$\begin{aligned}\mathcal{H}^*(\mathbf{x},t) &= \mathcal{H}_0(\mathbf{x},t) + \sum_{i=1}^{n_u} \min\left[\mathcal{H}_{1,i}(\mathbf{x},t)u_i\right] \\ &= \mathcal{H}_0(\mathbf{x},t) + \sum_{i=1}^{n_u} \left\{ \begin{array}{ll} \mathcal{H}_{1,i}(\mathbf{x},t)U_i^{(\max)}, & \mathcal{H}_{1,i}(\mathbf{x},t) < 0 \\ 0, & \mathcal{H}_{1,i}(\mathbf{x},t) = 0 \\ \mathcal{H}_{1,i}(\mathbf{x},t)U_i^{(\min)}, & \mathcal{H}_{1,i}(\mathbf{x},t) > 0 \end{array} \right\} \\ &= \mathcal{H}_0(\mathbf{x},t) + \frac{1}{2}\mathcal{H}_1(\mathbf{x},t).*\Big[\mathbf{U}^{(\min)}.*(\mathbf{1}+\mathbf{sgn}_1) \\ &\quad + \mathbf{U}^{(\max)}.*(\mathbf{1}-\mathbf{sgn}_1)\Big]\,, \end{aligned} \tag{6.42}$$

where $\mathbf{1} \equiv [1]_{n_u\times 1}$, $\mathbf{sgn}_1 \equiv [\mathrm{sgn}(\mathcal{H}_{1,i}(\mathbf{x},t)]_{n_u\times 1}$,

$$\mathrm{sgn}(x) \equiv \left\{ \begin{array}{ll} -1, & x < 0 \\ 0, & x = 0 \\ +1, & x > 0 \end{array} \right\} \tag{6.43}$$

is the sign or signum function, $\mathbf{U}^{(\min)} \equiv [U_i^{(\min)}]_{n_u\times 1}$, $\mathbf{U}^{(\max)} \equiv [U_i^{(\max)}]_{n_u\times 1}$, and $\mathbf{v}.*\mathbf{u} \equiv [v_i u_i]_{n_u\times 1}$ is the dot-star or element-by-element product. The optimal control is undefined for components for which $\mathcal{H}_{1,i}(\mathbf{x},t) = 0$, but otherwise is given in composite form:

$$u_i^*(\mathbf{x},t) = \left\{ \begin{array}{ll} U_i^{(\max)}, & \mathcal{H}_{1,i}(\mathbf{x},t) < 0 \\ U_i^{(\min)}, & \mathcal{H}_{1,i}(\mathbf{x},t) > 0 \end{array} \right\}\,. \tag{6.44}$$

If the components of $\mathcal{H}_1$ change sign often, then that can lead to **chattering control**.

6.4.3 Canonical LQJD Problem

The standard or canonical LQJD problem is linear in the dynamics and quadratic in the costs with respect to both state and control vectors. This LQJD problem is a special case of the LQJD problem in control only and results in substantial simplifications of the solution with a quadratic state decomposition of the optimal value function and the a linear or feedback decomposition of the optimal control vector. The decomposition of optimal value and control is similar to that of the deterministic LQ problem, but here the more general quadratic state and linear control decompositions is presented.

Let the more general jump-diffusion linear quadratic model be given with the plant function for the deterministic or non-noise dynamics term and be linear in both state $\mathbf{X}(t)$ and $\mathbf{U}(t)$,

$$\mathbf{f}(\mathbf{x},\mathbf{u},t) = \mathbf{f}_0(t) + f_1^\top(t)\mathbf{x} + f_2^\top(t)\mathbf{u}\,, \tag{6.45}$$

the first subscript indicating the degree and the subsequent subscripts, if present, indicating either state (1) or control (2), with the diffusion term,

$$g(\mathbf{x}, \mathbf{u}, t) = g_0(t) \,, \tag{6.46}$$

assumed state-independent and control-independent for simplicity, and with the jump term,

$$h(\mathbf{x}, \mathbf{u}, t, \mathbf{q}) = h_0(t, \mathbf{q}) \,, \tag{6.47}$$

also assumed state-independent and control-independent for simplicity. The current form of the linear SDE (6.1) is written here as

$$d\mathbf{X}(s) \overset{\text{sym}}{=} \mathbf{f}(\mathbf{X}(s), \mathbf{U}(s), s)ds + g_0(s)d\mathbf{W}(s) + h_0(s, \mathbf{Q})d\mathbf{P}(s; \mathbf{Q}, s) \,, \tag{6.48}$$

on $t \le s \le t_f$, with $\mathrm{E}[d\mathbf{P}(t; \mathbf{Q}, t)] = [\lambda_{0,j}(t)dt]_{n_p \times 1}$.

The quadratic running cost function is

$$\begin{aligned} C(\mathbf{x}, \mathbf{u}, t) = C_0(t) + \mathbf{C}_1^\top(t)\mathbf{x} + \mathbf{C}_2^\top(t)\mathbf{u} \\ + \frac{1}{2}\mathbf{x}^\top C_{1,1}(t)\mathbf{x} + \mathbf{x}^\top C_{1,2}(t)\mathbf{u} + \frac{1}{2}\mathbf{u}^\top C_{2,2}(t)\mathbf{u} \end{aligned} \tag{6.49}$$

and the terminal cost also has a general quadratic form

$$S(\mathbf{X}(t_f), t_f) = S_0(t_f) + \mathbf{S}_1^\top(t_f)\mathbf{X}(t_f) + \frac{1}{2}\mathbf{X}^\top(t_f)S_{1,1}(t_f)\mathbf{X}(t_f) \,, \tag{6.50}$$

in the state vector. It is assumed that all right hand side coefficients are commensurate in multiplication and the product is the same type as that on the left hand side. It is assumed that all coefficients are well-defined, but in particular that $C_{2,2}(t)$ is positive definite for the minimum problem, a crucial assumption, and symmetric due to the quadratic form, while $C_{1,1}(t)$ and $C_{1,2}(t)$ need to be positive semi-definite. Also, $S_{1,1}(t_f)$ is symmetric, positive semi-definite.

As in the deterministic LQ problem in Section A.4.4, a quadratic function of the state vector is sought. However, due to the extra linear terms in the quadratic cost beyond the pure quadratic form in (A.126) a more general quadratic decomposition is heuristically assumed for the optimal value,

$$v^*(\mathbf{x}, t) = v_0(t) + \mathbf{v}_1^\top(t)\mathbf{x} + \frac{1}{2}\mathbf{x}^\top v_{1,1}(t)\mathbf{x} \,, \tag{6.51}$$

where the optimal value coefficients $\{v_0(t), \mathbf{v}_1(t), v_{1,1}(t)\}$ are compatible in multiplication and any product is scalar valued. Without loss of generality, the quadratic coefficient $v_{1,1}(t)$ is taken to be symmetric. Consequently, the partial derivative with respect to time is

$$v_t^*(\mathbf{x}, t) = \dot{v}_0(t) + \dot{\mathbf{v}}_1^\top(t)\mathbf{x} + \frac{1}{2}\mathbf{x}^\top \dot{v}_{1,1}(t)\mathbf{x} \,,$$

where $\{\dot{v}_0(t), \dot{\mathbf{v}}_1(t), \dot{v}_{1,1}(t)\}$ denote the state time derivatives, the state gradient is

$$\nabla_{\mathbf{x}}[v^*](\mathbf{x}, t) = \mathbf{v}_1(t) + v_{1,1}(t)\mathbf{x} \,,$$

the state Hessian is

$$\nabla_{\mathbf{x}}\left[\nabla_{\mathbf{x}}^{\top}[v^*]\right](\mathbf{x},t) = v_{1,1}(t)$$

and the jump increment is

$$\begin{aligned}\Delta_j[v^*](\mathbf{x},t,q_j) = \mathbf{v}_1^{\top}(t)\widehat{\boldsymbol{h}}_{0,j}(t,q_j) + \frac{1}{2}\widehat{\boldsymbol{h}}_{0,j}^{\top}(t,q_j)v_{1,1}(t)\widehat{\boldsymbol{h}}_{0,j}(t,q_j)\\ +\mathbf{x}^{\top}v_{1,1}(t)\widehat{\boldsymbol{h}}_{0,j}(t,q_j)\,,\end{aligned}$$

where

$$\widehat{\boldsymbol{h}}_{0,j}(t,q_j) = [h_{0,i,j}(t,q_j)]_{n_x\times 1}$$

for $j = 1:n_p$.

With the proposed general quadratic decomposition (6.51) of $v^*(\mathbf{x},t)$, the pseudo-Hamiltonian has a quadratic decomposition in both state and control vectors like the cost coefficient $C(\mathbf{x},\mathbf{u},t)$ decomposition (6.49),

$$\begin{aligned}\mathcal{H}(\mathbf{x},\mathbf{u},t) = \mathcal{H}_0(t) + \boldsymbol{\mathcal{H}}_1^{\top}(t)\mathbf{x} + \boldsymbol{\mathcal{H}}_2^{\top}(t)\mathbf{u}\\ +\frac{1}{2}\mathbf{x}^{\top}\mathcal{H}_{1,1}(t)\mathbf{x} + \mathbf{x}^{\top}\mathcal{H}_{1,2}(t)\mathbf{u} + \frac{1}{2}\mathbf{u}^{\top}\mathcal{H}_{2,2}(t)\mathbf{u}\,,\end{aligned} \tag{6.52}$$

where the scalar coefficient is

$$\begin{aligned}\mathcal{H}_0(t) = C_0(t) + \mathbf{f}_0^{\top}(t)\mathbf{v}_1(t) + \frac{1}{2}\left(g_0 g_0^{\top}\right)(t) : v_{1,1}(t)\\ +\mathbf{v}_1^{\top}(t)\overline{h}_0(t).*\boldsymbol{\lambda}_0(t) + \frac{1}{2}v_{1,1}(t).*\overline{(h_0\Lambda h_0)}(t)\,,\end{aligned} \tag{6.53}$$

where

$$\overline{h}_0(t) \equiv \left[\int_{\mathcal{Q}_j} h_{0,i,j}(t,q_j)\phi_{Q_j}(q_j;t)dq_j\right]_{n_x\times n_p}\,, \tag{6.54}$$

$$\boldsymbol{\lambda}_0(t) \equiv [\lambda_{0,i}(t)]_{n_p\times 1}\,, \tag{6.55}$$

$$\Lambda_0(t) \equiv [\lambda_{0,i}(t)\delta_{i,j}]_{n_p\times n_p}\,, \tag{6.56}$$

$$\overline{h_0\Lambda_0 h_0^{\top}}(t) \equiv \left[\sum_{k=1}^{n_p}\lambda_{0,k}\int_{\mathcal{Q}_j} h_{0,i,k}(t,q_k)h_{0,j,k}(t,q_k)\phi_{Q_k}(q_k;t)dq_k\right]_{n_x\times n_x}, \tag{6.57}$$

the linear state coefficients is

$$\boldsymbol{\mathcal{H}}_1(t) = \mathbf{C}_1(t) + f_1(t)\mathbf{v}_1(t) + v_{1,1}(t)\mathbf{f}_0(t) + v_{1,1}(t)\overline{h}_0(t).*\boldsymbol{\lambda}_0(t)\,, \tag{6.58}$$

the linear control coefficient is

$$\boldsymbol{\mathcal{H}}_2(t) = \mathbf{C}_2(t) + f_2(t)\mathbf{v}_1(t)\,, \tag{6.59}$$

and the quadratic coefficients are

$$\mathcal{H}_{1,1}(t) = C_{1,1}(t) + 2f_1(t)v_{1,1}(t)\,, \tag{6.60}$$

$$\mathcal{H}_{1,2}(t) = C_{1,2}(t) + v_{1,1}^{\top}(t)f_2^{\top}(t)\,, \tag{6.61}$$

$$\mathcal{H}_{2,2}(t) = C_{2,2}(t)\,. \tag{6.62}$$

Since quadratic forms only operate on the symmetric part of the quadratic coefficient (B.135), $\mathcal{H}_{2,2}(t)$ will be symmetric, positive definite with $C_{2,2}(t)$.

The optimal control is the same as the regular control in the absence of control constraints, so the zero of

$$\nabla_{\mathbf{u}}[\mathcal{H}](\mathbf{x},\mathbf{u},t) = \boldsymbol{\mathcal{H}}_2(t) + \mathcal{H}_{1,2}^{\top}(t)\mathbf{x} + \mathcal{H}_{2,2}(t)\mathbf{u}$$

results in

$$\begin{aligned}\mathbf{u}^*(\mathbf{x},t) =& -\mathcal{H}_{2,2}^{-1}(t)\left(\boldsymbol{\mathcal{H}}_2(t) + \mathcal{H}_{1,2}^{\top}(t)\mathbf{x}\right)\\ =& -C_{2,2}^{-1}(t)\left(\mathbf{C}_2(t) + f_2(t)\mathbf{v}_1(t) + \left(C_{1,2}^{\top}(t) + f_2(t)v_{1,1}(t)\right)\mathbf{x}\right)\ .\end{aligned} \tag{6.63}$$

Hence, the optimal control vector is a linear or affine function of the state vector, the general form of linear feedback control. This completes the preliminary work on the LQJD problem for the feedback control state dependence.

Upon substituting the preliminary reduction of the linear optimal control (6.63) into the Hamilton Jacobi Bellman equation (6.35), then the HJB equation becomes

$$\begin{aligned}0 =& \ \dot{v}_0(t) + \dot{\mathbf{v}}_1^{\top}(t)\mathbf{x} + \tfrac{1}{2}\mathbf{x}^{\top}\dot{v}_{1,1}(t)\mathbf{x} + \mathcal{H}_0(t) + \boldsymbol{\mathcal{H}}_1^{\top}(t)\mathbf{x}\\ &-\boldsymbol{\mathcal{H}}_2^{\top}(t)\mathcal{H}_{2,2}^{-1}(t)\left(\boldsymbol{\mathcal{H}}_2(t) + \mathcal{H}_{1,2}^{\top}(t)\mathbf{x}\right)\\ &+\tfrac{1}{2}\mathbf{x}^{\top}\mathcal{H}_{1,2}(t)\mathcal{H}_{2,2}^{-1}(t)\mathbf{x} - \mathbf{x}^{\top}\mathcal{H}_{1,2}(t)\mathcal{H}_{2,2}^{-1}(t)\left(\boldsymbol{\mathcal{H}}_2(t) + \mathcal{H}_{1,2}^{\top}(t)\mathbf{x}\right)\\ &+\tfrac{1}{2}\left(\boldsymbol{\mathcal{H}}_2^{\top}(t) + \mathbf{x}^{\top}\mathcal{H}_{1,2}(t)\right)\mathcal{H}_{2,2}^{-1}(t)\left(\boldsymbol{\mathcal{H}}_2(t) + \mathcal{H}_{1,2}^{\top}(t)\mathbf{x}\right)\ .\end{aligned} \tag{6.64}$$

Next, separating this LQJD form of the HJBE (6.64) into purely quadratic terms, purely linear terms and state-independent terms leads to a set of three uni-directionally coupled ordinary matrix differential equations for the optimal control coefficients $v_{1,1}(t)$, $\mathbf{v}_1(t)$ and $v_0(t)$ which are summarized in the following theorem which we have just derived.

Theorem 6.6. ***LQJD Equations:***
Let the $n_x \times 1$ jump-diffusion state process $\mathbf{X}(t)$ satisfy dynamics linear in both the state and the $n_u \times 1$ control $\mathbf{U}(t)$ with $n_x \times 1$ linear deterministic plant term

$$\mathbf{f}(\mathbf{x},\mathbf{u},t) = \mathbf{f}_0(t) + f_1^{\top}(t)\mathbf{x} + f_2^{\top}(t)\mathbf{u}$$

from (6.45), with $n_x \times n_w$ state and control independent diffusion coefficient $g_0(t)$ of the $n_w \times 1$ Wiener process $d\mathbf{W}(t)$, and with $n_x \times n_p$ state and control independent jump amplitude $h_0(t,q)$ (6.46) of the $n_p \times 1$ Poisson process $d\mathbf{P}(t;\mathbf{Q},t)$. Let the scalar quadratic running cost be

$$\begin{aligned}C(\mathbf{x},\mathbf{u},t) = C_0(t) + \mathbf{C}_1^{\top}(t)\mathbf{x} + \mathbf{C}_2^{\top}(t)\mathbf{u}\\ +\frac{1}{2}\mathbf{x}^{\top}C_{1,1}(t)\mathbf{x} + \mathbf{x}^{\top}C_{1,2}(t)\mathbf{u} + \frac{1}{2}\mathbf{u}^{\top}C_{2,2}(t)\mathbf{u}\end{aligned}$$

and terminal cost be

$$S(\mathbf{X}(t_f), t_f) = S_0(t_f) + \mathbf{S}_1^\top(t_f)\mathbf{X}(t_f) + \frac{1}{2}\mathbf{X}^\top(t_f)S_{1,1}(t_f)\mathbf{X}(t_f)\,.$$

Then the optimal stochastic control problem admits a solution quadratic in the state vector

$$v^*(\mathbf{x}, t) = v_0(t) + \mathbf{v}_1^\top(t)\mathbf{x} + \frac{1}{2}\mathbf{x}^\top v_{1,1}(t)\mathbf{x}\,,$$

with optimal control vector that in linear in the state vector

$$\mathbf{u}^*(\mathbf{x}, t) = -C_{2,2}^{-1}(t)\left(\mathbf{C}_2(t) + f_2(t)\mathbf{v}_1(t) + \left(C_{1,2}^\top(t) + f_2(t)v_{1,1}(t)\right)\mathbf{x}\right)\,.$$

The optimal value $v^*(\mathbf{x}, t)$ coefficients satisfy a uni-directionally coupled set of matrix ordinary differential equations, which are solved starting from the $n_x \times n_x$ quadratic coefficient equation

$$\begin{aligned} 0 = {} & \dot{v}_{1,1}(t) + C_{1,1}(t) + 2f_1(t)v_{1,1}(t) \\ & - \left(C_{1,2}(t) + v_{1,1}(t)f_1^\top(t)\right)C_{2,2}^{-1}(t)\left(C_{1,2}^\top(t) + f_1(t)v_{1,1}(t)\right) \end{aligned} \tag{6.65}$$

for $v_{1,1}(t)$, then the $n_x \times 1$ linear coefficient equation

$$\begin{aligned} 0 = {} & \dot{\mathbf{v}}_1(t) + \mathbf{C}_1(t) + f_1(t)\mathbf{v}_1(t) \\ & - \left(C_{1,2}(t) + v_{1,1}(t)f_1^\top(t)\right)C_{2,2}^{-1}(t)\left(\mathbf{C}_2(t) + f_2(t)\mathbf{v}_1(t)\right) \\ & + v_{1,1}(t)\overline{h}_0(t)\boldsymbol{\lambda}_0(t) \end{aligned} \tag{6.66}$$

for $\mathbf{v}_1(t)$ using the existing solution for $v_{1,1}(t)$, and finally the scalar state-independent coefficient equation

$$\begin{aligned} 0 = {} & \dot{v}_0(t) + C_0(t) + \mathbf{f}_0^\top(t)\mathbf{v}_1(t) + \frac{1}{2}g_0(t)g_0^\top(t) : v_{1,1}(t) \\ & - \frac{1}{2}\left(\mathbf{C}_2^\top(t) + \mathbf{v}_1^\top(t)f_2(t)\right)C_{2,2}^{-1}(t)\left(\mathbf{C}_2(t) + f_2(t)\mathbf{v}_1(t)\right) \\ & + \mathbf{v}_1^\top(t)\overline{h}_0(t).*\boldsymbol{\lambda}_0(t) + \frac{1}{2}\overline{\left(h_0\Lambda_0h_0^\top\right)}(t) : v_{1,1}(t). \end{aligned} \tag{6.67}$$

Remarks 6.7.

- *The nonlinear differential equation (6.65) for the quadratic coefficient $v_{1,1}(t)$ is called a **matrix Riccati equation** due to the quadratic linearity in $v_{1,1}(t)$. Since $v_{1,1}(t)$ can be assumed to be symmetric without loss of generality since it is defined as the coefficient of a quadratic from, computational effort can be reduced to just finding the upper or lower triangular part, i.e., just $n_x(n_x+1)/2$ elements.*
- *Once $v_{1,1}(t)$ is known or a reasonable approximation is found, the equation (6.66) for the linear coefficient $\mathbf{v}_1(t)$ will be a linear matrix equation which is relatively simpler to solve than the matrix Riccati equation.*

- *Similarly, once both $v_{1,1}(t)$ and $\mathbf{v}_1(t)$ are found to reasonable approximations, then equation (6.67) for the state-independent coefficient $v_0(t)$ will be a linear scalar equation.*
- *Once the solutions to the time-dependent coefficients $v_{1,1}(t)$, $\mathbf{v}_1(t)$ and $v_0(t)$ are obtained, then the optimal value $v^*(\mathbf{x}, t)$ quadratic decomposition (6.51) is justified, at least heuristically.*

6.5 Exercises

1. For the scalar linear jump-diffusion dynamics with arithmetic rather than geometric diffusion,

$$dX(t) = (\mu_0 X(t) + \beta_0 U(t))dt + \sigma_0 dW(t) + \nu_0 X(t)dP(t),$$

for $0 \leq t \leq t_f$ and initial state $X(0) = x_0 > 0$ and the control process $-\infty < U(t) < +\infty$ is unconstrained. The coefficients $\mu_0 \neq 0$, $\beta_0 \neq 0$, $\sigma_0 > 0$, $\nu_0 \geq 0$ and $\lambda_0 > 0$ are constants, where $\mathrm{E}[dP(t)] = \lambda_0 dt$ (note that the jump process here is a discrete, Poisson process, since there is no mark process). The costs are quadratic, i.e.,

$$V[X,U](X(t),t) = \frac{1}{2}\int_t^{t_f} \left(q_0 X^2(s) + r_0 U^2(s)\right) ds + \frac{1}{2} S_f X^2(t_f)$$

for $q_0 > 0$, $r_0 > 0$, and $S_f > 0$. Let the optimal, expected value be

$$v^*(x,t) = \min \left[\mathrm{E}\left[V[X,U](X(t),t) \,|X(t) = x, U(t) = u \right] \right].$$

   (a)
   (b) Derive the PDE of Stochastic Dynamic Programming for the optimal expected value:

$$v^*(x,t) = \min_u \left[\mathrm{E}\left[V[X,U](X(t),t) \,|X(t) = x, U(t) = u \right] \right],$$

   starting from the Principle of Optimality;
   (c) Specify the final condition for $v^*(x,t)$ fully qualified;
   (d) Formally find the optimal (unconstrained) control $u^*(x,t)$ in terms of the *shadow "cost"* $v_x^*(x,t)$;
   (e) Obtain an LQJD solution form for $v^*(x,t)$ and an explicit linear feedback control law for $u^*(x,t)$;

2. Derive the modifications necessary in the set of Riccati-like equations for the scalar Linear-Quadratic Jump-Diffusion (LQJD) problem when the dynamics are scalar and linear (affine), i.e.,

$$dX(t) = f(X(t),U(t),t)dt + g(X(t),U(t),t)dW(t) + h(X(t),U(t),t)dP(t)\,,$$

where

$$\mathrm{E}[dP(t)] = \lambda(t)dt\,,$$
$$f(x,u,t) = f_0(t) + f_1(t)x + f_2(t)u\,,$$
$$g(x,u,t) = g_0(t) + g_1(t)x\,,$$
$$h(x,u,t) = h_0(t) + h_1(t)x\,,$$

the jump amplitude being independent of any mark process. The running and terminal costs for a maximum objective are quadratic,

$$C(x,u,t) = C_0(t) + C_1(t)x + C_2(t)u + 0.5C_{1,1}(t)x^2 + C_{1,2}(t)xu + 0.5C_{2,2}(t)u^2\,,$$

where $C_{2,2}(t) < 0$, and

$$S(x,t) = S_0(t) + S_1(t)x + 0.5 * S_{1,1}(t)x^2\,,$$

where $S_{1,1}(t) < 0$.

If the objective is to maximize the expected total utility in the unconstrained control case, then find the Riccati ODEs for the coefficient functions $v_0(t)$, $v_1(t)$, $v_{1,1}(t)$, in the solution form,

$$v^*(x,t) = v_0(t) + v_1(t)x + 0.5v_{1,1}(t)x^2$$

and $u_0(t)$ and $u_1(t)$ in the form

$$u^*(x,t) = u_0(t) + u_1(t)x$$

explicitly in terms of the $\{v_0(t), v_1(t), v_{1,1}(t)\}$, dynamical and cost coefficient functions. Do not try to solve the Riccati equation system of ODEs for $\{v_0(t), v_1(t), v_{1,1}(t)\}$.

3. Let $\beta(t)$ be the discount rate at time t and

$$\exp\left(-\widehat{\beta}(t,s)\right) = \exp\left(-\int_t^s \beta(r)dr\right) = \widehat{\beta}(0,s) - \widehat{\beta}(0,t) \qquad (6.68)$$

be the cumulative discount factor for the time-interval $[t,s]$, so the optimal, expected, discounted costs are

$$v^*(x,t) = \min_u \left[\mathrm{E}\left[\int_t^{t_f} e^{-\widehat{\beta}(t,s)} C(X(s),U(s),s)ds + e^{-\widehat{\beta}(t,t_f)} S(X(t_f),t_f)\,\middle|\, \mathcal{C}(t)\right]\right],$$

where $\mathcal{C}(t) = \{X(t) = x, U(t) = u\}$ is the conditioning set. Noting that this $v^*(x,t)$ does not have the form to satisfy the Principle of Optimality given in (6.9) because of the dual-time dependence of the discount factor on (t,s), so

(a) show that $w^*(x,t) = \exp\left(-\widehat{\beta}(t)\right) v^*(x,t)$ properly satisfies the usual form of the Principle of Optimality (6.9) and hence

(b) show that proper modification of the Principle of Optimality for discounted costs is

$$v^*(x,t) = \min_u \left[\mathrm{E}\left[\int_t^{t+\delta t} e^{-\widehat{\beta}(t,s)} C(X(s),U(s),s)ds + e^{-\widehat{\beta}(t,t+\delta t)} v^*(X(t+\delta t), t+\delta t) \Big| \mathcal{C}(t) \right] \right]. \tag{6.69}$$

4. Derive the Hamilton-Jacobi-Bellman PDE for the scalar optimal stochastic control problem (a simplified jump-diffusion optimal portfolio and consumption problem), with stochastic dynamical system,

$$dX(t) = X(t)\left(\mu_0(t)dt + U_1(t)\left(\mu_1(t)dt + \sigma(t)dW(t) + \left(e^Q - 1\right)dP(t)\right)\right) - U_2(t)dt\,,$$

where $t \in [0, t_f]$, $X(0) = x_0 > 0$, $\mathrm{E}[dP(t)] = \lambda(t)dt = \mathrm{Var}[dP(t)]$, $\mathrm{E}[dW(t)] = 0$, $\mathrm{Var}[dW(t)] = dt$, Q is an IID uniformly distributed mark on $[a,b]$, $a < 0 < b$,

$$\{\mu_0(t), \mu_1(t), \sigma(t), \lambda(t)\}$$

are specified time-dependent coefficients, $X(t) \geq 0$ is the state, $\{U_1(t), U_2(t)\}$ is the control set, $0 \leq U_2(t) \leq K_2 X(t)$, $K_2 > 0$, $-U_N \leq U_1(t) \leq U_P$, $U_N > 0$, $U_P > 0$ and the optimal objective is

$$v^*(x,t) = \max_{\{u_1,u_2\}} \left[\mathrm{E}_{\{W,P\}} \left[\int_t^{t_f} e^{-\widehat{\beta}(t,s)} \frac{U_2^\gamma(s)}{\gamma} ds + e^{-\widehat{\beta}(t,t_f)} \frac{X^\gamma(t_f)}{\gamma} \Big| \mathcal{C}(t) \right] \right],$$

where $\mathcal{C}(t) \equiv \{X(t) = x, U_1(t) = u_1(t), U_2(t) = u_2(t)\}$ is the conditioning set, $\beta(t) > 0$ is the discount rate with the cumulative discount $\widehat{\beta}(t,s)$ defined in (6.68), $\gamma \in (0,1)$ is a constant utility power and the zero-state absorbing boundary condition for this problem is $v^*(0^+, t) = 0$.

(a) If Exercise 3 on the form of the Principle of Optimality of discounting has not been done, then do it now, otherwise proceed to next item.

(b) Derive the modified HJBE for time-discounting from the discount form of the Principle of Optimality in (6.69), with the minimum merely replaced by a maximum. Be sure to point out the difference from the non-discounting form.

(c) Derive the relationship of the optimal controls to the *shadow utility* $v_x^*(x,t)$, accounting for the control constraints.

(d) Test the validity of the CRRA (constant relative risk aversion) canonical separated form of the regular solution,

$$v^*(x,t) = v_0(t)x^\gamma/\gamma,$$

determining what reduced ODE the time-dependent solution factor satisfies, specifying what side (final and boundary) conditions need to be satisfied for the problem.

Suggested References for Further Reading

- Bellman, 1957 [25].
- Cont and Tankov, 2004 [59].
- Fleming and Rishel, 1975 [85].
- Gihman and Skorohod, 1979 [95].
- Hanson, 1996 [108].
- Jazwinski, 1970 [154].
- Kushner, 1967 [173].
- Kushner and Dupuis, 2001 [179].
- Lewis, 1986 [184].
- Øksendal and Sulem, 2005 [223].
- Runggaldier, 2003 [239].
- Stengel, 1994 [258].
- Yong and Zhou, 1999 [288].

Chapter 7

Kolmogorov Forward and Backward Equation and Their Applications

The theory of probability as mathematical discipline can and should be developed from axioms in exactly the same way as Geometry and Algebra.
—Andrey Nikolaevich Kolmogorov (1903-1987), *Wikipedia*, March 2006.

Here, the Kolmogorov forward (Fokker-Planck) and backward equations are treated, including their inter-relationship and their use in finding transition distributions, densities, moments and optimal state trajectories. There is a close relationship between the PDE representations in the Kolmogorov equations and the SDE representation. Unlike the SDE which is a symbolic representation that requires specification of the stochastic ne integration rule to be well posed, the Kolmogorov equations are deterministic. They can be derived from an SDE using expectations and a chain rule such as Itô's chain rule. Some investigators prefer to solve problems with the Kolmogorov PDEs rather than directly from the underlying SDEs.

7.1 Dynkin's Formula and the Backward Operator

Prior to deriving the Kolmogorov PDEs, a useful formula due to Dynkin is derived. Dynkin's formula relates the expectation of a function of a jump-diffusion process and a functional of the backward jump-diffusion operator. There are many variants of Dynkin's formula [77], but here a derivation of Schuss [244] for pure-diffusions is modified for jump-diffusions in the time-inhomogeneous case and in one-dimension to start.

Theorem 7.1. ***Dynkin's Formula for Jump-Diffusions on*** $[t_0, t]$ ***in One Space Dimension:***

Let $X(t)$ be a jump-diffusion process satisfying the SDE,

$$dX(t) \stackrel{\text{sym}}{=} f(X(t),t)dt + g(X(t),t)dW(t) + h(X(t),t,Q)dP(t;Q,X(t),t)\,, \quad (7.1)$$

with smooth (continuously differentiable) coefficients $\{f,g,h\}$ with bounded spatial gradients. The diffusion process is the Wiener process $W(t)$ and the jump process is the Poisson process $P(t;Q,X(t),t)$ such that $\mathrm{E}[dP(t;Q,X(t),t)|X(t)=x] = \lambda(t;x,t)dt$ and Q is the jump amplitude mark random variable with density $\phi_Q(q;X(t),t)$. Let $v(x,t)$ be twice continuously differentiable in x and once in t, while bounded at infinity. Then the conditional expectation of the composite process $v(X(t),t)$ satisfies Dynkin's formula in integral form,

$$\begin{aligned} u(x_0,t_0) &= \overline{v}(x_0,t_0;t) \equiv \mathrm{E}[v(X(t),t)|X(t_0)=x_0] \\ &= v(x_0,t_0) + \mathrm{E}\left[\int_{t_0}^{t}\left(\frac{\partial v}{\partial t}(X(s),s) + \mathcal{B}_x[v](X(s),s)\right)ds \middle| X(t_0)=x_0\right], \end{aligned} \quad (7.2)$$

*where the dependence on the parameter t is suppressed in $u(x_0,t_0)$. The jump-diffusion **backward operator** $\mathcal{B}_{x_0}$ with respect to the state x_0 for time t dependent coefficients, in backward coordinates, is*

$$\begin{aligned} \mathcal{B}_{x_0}[v](x_0,t_0) \equiv\, & f(x_0,t_0)\frac{\partial v}{\partial x_0}(x_0,t_0) + \frac{1}{2}g^2(x_0,t_0)\frac{\partial^2 v}{\partial x_0^2}(x_0,t_0) \\ & + \widehat{\lambda}(x_0,t_0)\int_{\mathcal{Q}} \Delta_h[v](x_0,t_0,q)\phi_Q(q;x_0,t_0)dq, \end{aligned} \quad (7.3)$$

where $\widehat{\lambda}(x_0,t_0) \equiv \lambda(t;x_0,t_0)$ suppresses the forward time t and the Poisson h-jump is

$$\Delta_h[v](x_0,t_0,q) \equiv v(x_0+h(x_0,t_0,q),t) - v(x_0,t_0)\,. \quad (7.4)$$

Note that the subscript x_0 on the backward operator $\mathcal{B}_{x_0}$ only denotes that the operator operates with respect to the backward state variable x_0 for jump-diffusions and only denotes partial differentiation in the pure-diffusion ($h(x_0,t_0,q) \equiv 0$) case.

In the time-homogeneous case, $f(x,t) = f(x)$, $g(x,t) = g(x)$ and $h(x,t,q) = h(x,q)$, so $v(x,t) = v(x)$ and

$$\begin{aligned} u(x_0) &\equiv \mathrm{E}[v(X(t))|X(t_0)=x_0] \\ &= v(x_0) + \mathrm{E}\left[\int_{t_0}^{t}\mathcal{B}_x[v](X(s))ds \middle| X(t_0)=x_0\right], \end{aligned} \quad (7.5)$$

dropping the t dependence of the backward operator here.

Proof. Dynkin's formula follows from Itô's chain rule for jump-diffusions here. Thus,

$$\begin{aligned} dv(X(t),t) \stackrel{\text{dt}}{=}\, & \left(\frac{\partial v}{\partial t} + f\frac{\partial v}{\partial x} + \frac{1}{2}g^2\frac{\partial^2 v}{\partial x^2}\right)(X(t),t)dt + \left(g\frac{\partial v}{\partial x}\right)(X(t),t)dW(t) \\ & + \int_{\mathcal{Q}} \Delta_h[v](X(t),t,q)\mathcal{P}(\mathbf{dt},\mathbf{dq};X(t),t)\,, \end{aligned} \quad (7.6)$$

where common arguments have been condensed. Upon integrating in t,

$$v(X(t),t) = v(x_0,t_0) + \int_{t_0}^{t} \left(\left(\frac{\partial v}{\partial t} + f\frac{\partial v}{\partial x} + \frac{1}{2}g^2\frac{\partial^2 v}{\partial x^2} \right)(X(s),s)ds \right. \tag{7.7}$$
$$\left. + \left(g\frac{\partial v}{\partial x} \right)(X(s),s)dW(s) + \int_{\mathcal{Q}} \Delta_h[v](X(s),s,q)\mathcal{P}(\mathbf{ds},\mathbf{dq};X(s),s) \right).$$

Next taking expectations while using the facts that follow from the independent increment property of Markov processes,

$$\mathrm{E}\left[\int_{t_0}^{t} G(X(s),s)dW(s) \right] = 0$$

after (2.43) and with the zero mean jump process

$$\mathrm{E}\left[\int_{t_0}^{t} H(X(s),s)\widehat{\mathcal{P}}(\mathbf{ds},\mathbf{dq};X(s),s) \right] = 0,$$

generalized from (3.27) with $d\widehat{P}(s)$, where here the mean-zero Poisson random measure is

$$\widehat{\mathcal{P}}(\mathbf{dt},\mathbf{dq};X(t),t) \equiv \mathcal{P}(\mathbf{dt},\mathbf{dq};X(t),t) - \lambda(t;X(t),t)\phi_Q(q;X(t),t)dqdt, \tag{7.8}$$

then using the definition of the backward operator $\mathcal{B}_x[v]$,

$$\begin{aligned} E[v(X(t),t)|X(t_0)=x_0] = \; & v(x_0,t_0) \\ & + \mathrm{E}\Big[\textstyle\int_{t_0}^{t} \left(\frac{\partial v}{\partial t} + \mathcal{B}_x[v] \right)(X(s),s)ds \,|X(t_0)=x_0 \Big]. \end{aligned} \tag{7.9}$$

In the time-homogeneous case, without time-dependent coefficients, we need only use the x-dependent test function $v = v(x)$ and the Dynkin formula reduces to (7.5). □

Example 7.2. ***Application of Dynkin's Formula to Final Value Problems:*** *Consider the final value problem for the backward problem with PDE*

$$\begin{gathered} \frac{\partial v}{\partial t_0}(x_0,t_0) + \mathcal{B}_{x_0}[v](x_0,t_0) = \alpha(x_0,t_0) \quad x_0 \in \Omega, \quad t_0 < t_f\,, \\ v(x_0,t_f) = \gamma(x_0,t_f) \quad x_0 \in \Omega\,, \end{gathered} \tag{7.10}$$

where the general functions $\alpha(x,t)$ and $\gamma(x,t)$ are given, while $\mathcal{B}_{x_0}[v](x_0,t_0)$ is the jump-diffusion backward operator defined in (7.3). From Dynkin's formula (7.2) with $t = t_f$,

$$\mathrm{E}[\gamma(X(t_f),t_f)|X(t_0)=x_0] = v(x_0,t_0) + \mathrm{E}\left[\int_{t_0}^{t_f} \alpha(X(s),s) \,\middle|\, X(t_0)=x_0 \right],$$

where the jump-diffusion process is given by the SDE (7.1). By simple rearrangement, the formal solution to the final value problem is given by

$$v(x_0,t_0) = \mathrm{E}\bigg[\gamma(X(t_f),t_f) - \int_{t_0}^{t_f} \alpha(X(s),s)\bigg| X(t_0)=x_0\bigg], \tag{7.11}$$

in a more useful form, suitable for stochastic simulations using the given problem functions and the SDE.

The final problem (7.10) can be called the ***Dynkin's equation*** *corresponding to* ***Dynkin's formula*** *(7.2).*

7.2 Backward Kolmogorov Equations

Many exit and stopping time problems rely on the backward Kolmogorov equations, since they represent perturbations of the initial condition when the final condition for exit or stopping is known. Another very useful application is a PDE governing the behavior of the transition density as a function of the initial state. First the general backward equation in the sense of Kolmogorov is derived using an infinitesimal form of Dynkin's equation.

Theorem 7.3. ***General Backward Kolmogorov Equation for Jump-Diffusions on*** $[t_0,t]$ ***in One Space Dimension:***
Let the jump-diffusion process $X(t)$ *at time* t *with* $X(t_0)=x_0$ *at initial or backward time* t_0 *satisfy (7.1) along with associated conditions and let the test function* $v(X(t))$ *also satisfy relevant conditions. Let*

$$u(x_0,t_0) = \overline{v}(x_0,t_0;t) \equiv \mathrm{E}[v(X(t))|X(t_0)=x_0] = \mathrm{E}_{(t_0,t]}[v(X(t))|X(t_0)=x_0], \tag{7.12}$$

suppressing the ***forward time*** t *in favor of the* ***backward time*** t_0*. Then* $u(x_0,t_0)$ *satisfies the following backward PDE with backward arguments,*

$$0 = \frac{\partial u}{\partial t_0}(x_0,t_0) + \mathcal{B}_{x_0}[u](x_0,t_0)\,, \tag{7.13}$$

where the backward operator with respect to x_0 *operating on* u *is*

$$\begin{aligned}\mathcal{B}_{x_0}[u](x_0,t_0) = f(x_0,t_0)\frac{\partial u}{\partial x_0}(x_0,t_0) + \frac{1}{2}g^2(x_0,t_0)\frac{\partial^2 u}{\partial x_0^2}(x_0,t_0) \\ +\widehat{\lambda}(x_0,t_0)\int_{\mathcal{Q}} \Delta_h[u](x_0,t_0,q)\phi_Q(q;x_0,t_0))dq\,,\end{aligned} \tag{7.14}$$

the h*-jump of* u *is*

$$\Delta_h[u](x_0,t_0,q) \equiv u(x_0+h(x_0,t_0,q),t_0) - u(x_0,t_0)\,, \tag{7.15}$$

with final condition

$$\lim_{t_0\uparrow t} u(x_0,t_0) = v(x_0). \tag{7.16}$$

Proof. This formal proof is a modified version of the one for pure diffusions in Schuss [244] modified to include Poisson jump processes. First, the objective is to calculate the backward time partial derivative

$$u(x_0,t_0)-u(x_0,t_0-dt)\stackrel{\rm dt}{=}\frac{\partial u}{\partial t_0}dt\equiv\left.\frac{\partial u}{\partial t_0}\right|_{x_0\ \text{fixed}}dt\,,$$

so consider the infinitesimal backward difference in the spirit of Dynkin's formula, noting that the initial time t_0 is perturbed one step backward in time to t_0-dt with fixed x_0. On the other hand, using the representation (7.12), splitting the expectation at t_0 using the new random variable $X(t_0)$ and expanding by the stochastic chain rule,

$$\begin{aligned}
u(x_0,t_0)-u(x_0,t_0-dt) &= u(x_0,t_0)-E[v(X(t))|X(t_0-dt)=x_0]\\
&= u(x_0,t_0)-\mathrm{E}[\mathrm{E}[v(X(t))|X(t_0)]|X(t_0-dt)=x_0]\\
&= u(x_0,t_0)-\mathrm{E}[u(X(t_0),t_0)|X(t_0-dt)=x_0]\\
&= \mathrm{E}[u(x_0,t_0)-u(X(t_0),t_0)|X(t_0-dt)=x_0]\\
&\stackrel{\rm dt}{=} \mathrm{E}[\mathcal{B}_{x_0}[u](x_0,t_0)dt+g(x_0,t_0)dW(t_0)\\
&\quad+\int_{\mathcal{Q}}\Delta_h[u](X(s),s,q)\widehat{\mathcal{P}}(\mathbf{ds},\mathbf{dq};X(s),s)|X(t_0-dt)=x_0]\\
&= \mathrm{E}[\mathcal{B}_{x_0}[u](x_0,t_0)dt|X(t_0-dt)=x_0]\\
&= \mathcal{B}_{x_0}[u](x_0,t_0)dt\\
&= \left[f(x_0,t_0)\frac{\partial u}{\partial x_0}(x_0,t_0)+\frac{1}{2}g^2(x_0,t_0)\frac{\partial^2 u}{\partial x_0^2}(x_0,t_0)\right.\\
&\quad\left.+\,\widehat{\lambda}(x_0,t_0)\int_{\mathcal{Q}}\Delta_h[u](x_0,t_0,q)\phi_Q(q;x_0,t_0)dq\right]dt\,,
\end{aligned}$$

where the stochastic chain rule (5.41) was used, marked by the dt-precision step, along with expectations over the zero-mean jump-diffusion differentials. Just equating the two about results for $u(x_0,t_0)-u(x_0,t_0-dt)$ and eliminating the dt factor yields the backward Kolmogorov equation (7.13) result. The final condition (7.16) simply follows from the definition of $u(x_0,t_0)$ in (7.12) and taking the indicated limit from the backward time t_0 to the forward time t for fixed x_0,

$$\lim_{t_0\uparrow t}u(x_0,t_0)=\lim_{t_0\uparrow t}\mathrm{E}[v(X(t))|X(t_0)=x_0]=\mathrm{E}[v(X(t))|X(t)=x_0]=v(x_0).$$

□

Transition Probability Distribution $\Phi_{X(t)}(x,t;x_0,t_0)$:

One of the most important applications of the backward Kolmogorov equation is for the transition probability whose distribution is given by

$$\Phi_{X(t)}(x,t;x_0,t_0)\equiv\mathrm{Prob}[X(t)\le x|X(t_0)=x_0] \tag{7.17}$$

with density

$$\phi_{X(t)}(x,t;x_0,t_0) = \frac{\partial \Phi_{X(t)}}{\partial x}(x,t;x_0,t_0) \tag{7.18}$$

or alternatively by

$$\begin{aligned}\phi_{X(t)}(x,t;x_0,t_0)dx &\stackrel{\text{dx}}{=} \text{Prob}[x < X(t) \le x+dx | X(t_0)=x_0] \\ &= \text{Prob}[X(t) \le x+dx | X(t_0)=x_0] \\ &\quad -\text{Prob}[X(t) \le x | X(t_0)=x_0]\,,\end{aligned} \tag{7.19}$$

in dx-precision, provided the density exists, including the case of generalized functions (see Section B.12) as assumed in this book. In terms of the transition density, the conditional expectation can be rewritten such that

$$\begin{aligned}u(x_0,t_0) = \overline{v}(x_0,t_0;t) &= \text{E}_{(t_0,t]}[v(X(t))|X(t_0)=x_0] \\ &= \int_{-\infty}^{+\infty} v(x)\phi_{X(t)}(x,t;x_0,t_0)dx\,.\end{aligned} \tag{7.20}$$

Thus, if we let

$$v(x) \stackrel{\text{gen}}{=} \delta(x-\xi),$$

then

$$u(x_0,t_0) = \overline{v}(x_0,t_0;t) = \phi_{X(t)}(\xi,t;x_0,t_0)$$

by definition of the Dirac delta function, and so the transition density satisfies the general backward Kolmogorov equation (7.13) in the backward or initial arguments (x_0,t_0).

Corollary 7.4. ***Backward Kolmogorov Equation for Jump-Diffusion Transition Density:***
Let $\widehat{\phi}(x_0,t_0) \equiv \phi_{X(t)}(x,t;x_0,t_0)$, suppressing the parametric dependence on the forward coordinates (x,t), where the process satisfies the jump-diffusion SDE (7.1) under the specified conditions. Then

$$0 = \frac{\partial \widehat{\phi}}{\partial t_0}(x_0,t_0) + \mathcal{B}_{x_0}[\widehat{\phi}](x_0,t_0) \tag{7.21}$$

$$\begin{aligned}&= \frac{\partial \widehat{\phi}}{\partial t_0}(x_0,t_0) + f(x_0,t_0)\frac{\partial \widehat{\phi}}{\partial x_0}(x_0,t_0) + \frac{1}{2}g^2(x_0,t_0)\frac{\partial^2 \widehat{\phi}}{\partial x_0^2}(x_0,t_0) \\ &\quad +\widehat{\lambda}(x_0,t_0)\int_{\mathcal{Q}} \Delta_h\left[\widehat{\phi}\right](x_0,t_0,q)\phi_Q(q;x_0,t_0)dq\,,\end{aligned} \tag{7.22}$$

subject to the final condition,

$$\lim_{t_0 \uparrow t} \widehat{\phi}(x_0,t_0) = \delta(x_0-x)\,. \tag{7.23}$$

The final condition (7.23) follows from the alternate, differential definition (7.19) of the transition probability density.

Often the transition density backward equation (7.21) is referred to as the **backward Kolmogorov equation**. It is useful for problems in which the final state is known, such as an exit time problem or a stopping time problem where a state boundary is reached, in the case of finite state domains. For some stochastic researchers, the backward equation is considered more basic than the forward equation, since in the backward equation some final goal may be reached as in stochastic dynamic programming, or some significant event may occur, such as the extinction time for a species. The evolution of the moments or expectations of powers of the state are governed by transition probability density.

7.3 Forward Kolmogorov Equations

In contrast to the backward time problems of the previous section, the forward equation will be needed to find the evolution of the transition density forward in time given an initial state. The basic idea is that the **forward operator** $\mathcal{F}_x$ and the **backward operator** are (formal) **adjoint** operators, i.e., under suitable conditions on the transition density

$$\phi(x,t) = \phi_{X(t)}(x,t;x_0,t_0),$$

with truncated arguments to focus on forward variables, and a well-behaved test function $v(x)$, well-behaved particularly at infinity. Then the operators are related through an inner product equality,

$$(\mathcal{B}_x[v],\phi) = (\mathcal{F}_x[\phi],v)\,, \tag{7.24}$$

which is derived in Theorem 7.5 below. The conditional expectations in Dynkin's formula can be considered an inner product over a continuous state space with the transition density such that

$$(v,\phi) = \mathrm{E}[v(X(t))|X(t_0)=x_0] = \int_{-\infty}^{+\infty} v(x)\phi(x,t)dx\,,$$

emphasizing forward variables (x,t).

Theorem 7.5. ***Forward Kolmogorov Equation or Fokker-Planck Equation for the Transition Density*** $\phi(\mathbf{x},\mathbf{t};\mathbf{x_0},\mathbf{t_0})$***:***
Let $\phi(x,t;x_0,t_0)$ be the transition probability density for the jump-diffusion process $X(t)$ that is symbolically represented by the SDE (7.1) along with the coefficient conditions specified in Dynkin's Formula Theorem 7.1. Let $v(x)$ be a bounded and twice differentiable but otherwise arbitrary test function such that the integrated ***conjunct*** *vanishes, i.e.,*

$$\left[\left((f\phi)(x,t) - \frac{1}{2}\frac{\partial(g^2\phi)}{\partial x}(x,t)\right)v(x) + \frac{1}{2}(g^2\phi)(x,t)v'(x)\right]_{-\infty}^{+\infty} = 0\,, \tag{7.25}$$

where $(f\phi)(x,t) \equiv f(x,t)\phi(x,t)$, $g^2(x,t) \equiv (g(x,t))^2$ and $v'(x) \equiv (dv/dx)(x)$. Then, in the weak sense, ϕ satisfies the forward Kolmogorov equation in forward

space-time variables (x,t),

$$\frac{\partial\phi}{\partial t}(x,t) = \frac{1}{2}\frac{\partial^2(g^2\phi)}{\partial x^2}(x,t) - \frac{\partial(f\phi)}{\partial x}(x;t) - (\widehat{\lambda}\phi)(x,t) + \int_{\mathcal{Q}} (\widehat{\lambda}\phi)(x-\eta,t)|1-\eta_x|\phi_Q(q;x-\eta,t)dq\,, \tag{7.26}$$

where $\eta = \eta(x;t,q)$ *is related to the inverse jump amplitude such that*

$$x = \xi + h(\xi,t,q)$$

is the new state value corresponding to the old state value ξ*, such that*

$$\eta(x;t,q) = h(\xi,t,q),$$

assuming h *is monotonic in* ξ *so that* h *is invertible with respect to* ξ*, that the Jacobian*

$$(1-\eta_x) = \left(1 - \frac{\partial\eta}{\partial x}(x;t,q)\right)\,,$$

is non-vanishing, and that the inverse transformation from ξ *to* x *maps* $(-\infty,+\infty)$ *onto* $(-\infty,+\infty)$.

The transition probability density satisfies the delta function intial condition,

$$\phi(x,t_0^+) = \phi_{X(t_0^+)}(x,t_0^+;x_0,t_0) = \delta(x-x_0)\,. \tag{7.27}$$

Proof. The main idea of this proof is to perform several integrations by parts to move the partial differentiation from the backward operator on the arbitrary test function $v(x)$ to differentiation of the jump-diffusion transition probability $\phi(x,t) = \phi_{X(t)}(x,t;x_0,t_0)$, deriving the adjoint backward-forward operator relation (7.24) in principle. Differentiating Dynkin's formula (7.2) in forward time t for fixed initial conditions (x_0,t_0) and for some well-behaved test function $v(x)$,

$$\begin{aligned}\frac{\partial\overline{v}}{\partial t}(x_0,t_0;t) &= \mathrm{E}\left[\frac{\partial}{\partial t}\int_{t_0}^{t}\mathcal{B}_x[v](X(s))ds\,\middle|\, X(t_0)=x_0\right] \\ &= \mathrm{E}\left[\mathcal{B}_x[v](X(t))|\, X(t_0)=x_0\right]\end{aligned} \tag{7.28}$$

assuming that differentiation and expectation can be interchanged, where the backward operator $\mathcal{B}$ is given in (7.3). However, the conditional expectation of $\mathcal{B}$ on the RHS of (7.28) can be written in terms of the transition probability ϕ (7.20),

$$\mathrm{E}[\mathcal{B}_x[v](X(t))|X(t_0)=x_0] = \int_{-\infty}^{+\infty}\mathcal{B}_x[v](x)\phi(x,t)dx\,. \tag{7.29}$$

Combining (7.28) and (7.29) , substituting for $\mathcal{B}$ using (7.3), and using two integration by parts on the spatial derivatives to move the spatial derivatives from v to ϕ,

then

$$\begin{aligned}
\frac{\partial \overline{v}}{\partial t}(x_0,t_0;t) &= \int_{-\infty}^{+\infty} v(x)\frac{\partial \phi}{\partial t}(x,t)dx = \int_{-\infty}^{+\infty} \mathcal{B}_x[v](x)\phi(x,t)dx \\
&= \int_{-\infty}^{+\infty} \Big(f(x,t)v'(x) + \frac{1}{2}g^2(x,t)v''(x) \\
&\quad + \widehat{\lambda}(x,t)\int_{\mathcal{Q}} \Delta_h[v](x,t,q)\phi_Q(q;x,t)dq \Big)\phi(x,t)dx \\
&= \int_{-\infty}^{+\infty} \Big(-v(x)\frac{\partial (f\phi)}{\partial x}(x,t) - \frac{1}{2}\frac{\partial (g^2\phi)}{\partial x}(x,t)v'(x) \\
&\quad + (\widehat{\lambda}\phi)(x,t)\int_{\mathcal{Q}} \Delta_h[v](x,t,q)\phi_Q(q;x,t)dq \Big)dx \\
&\quad + \Big[(f\phi)(x,t)v(x) + \frac{1}{2}(g^2\phi)(x,t)v'(x) \Big]_{-\infty}^{+\infty} \\
&= \int_{-\infty}^{+\infty} \Big(v(x)\Big(\frac{1}{2}\frac{\partial^2 (g^2\phi)}{\partial x^2}(x,t) - \frac{\partial (f\phi)}{\partial x}(x,t) \Big) \\
&\quad + (\widehat{\lambda}\phi)(x,t)\int_{\mathcal{Q}} \Delta_h[v](x,t,q)\phi_Q(q;x,t)dq \Big)dx \\
&\quad + \Big[\Big(f\phi - \frac{1}{2}\frac{\partial (g^2\phi)}{\partial x} \Big)(x,t)v(x) + \frac{1}{2}(g^2\phi)(x,t)v'(x) \Big]_{-\infty}^{+\infty}.
\end{aligned}$$

The last term is the integrated **conjunct** from two integrations by parts. By the hypothesis in (7.25), this conjunct is required to be zero, so that the forward and backward operators will be genuine adjoint operators. Otherwise, the forward and backward operators would be called *formal adjoints*.

So far only the adjoint diffusion part of the forward operator has been formed with respect to the test function v as an integration weight. There still remains more work to form the corresponding adjoint jump part and this is done inverting the jump amplitude function $h(x,t,q)$ with respect to x, assuming that $h(x,t,q)$ is monotonic x. Let the post-jump state value be $y = x + h(x,t,q)$ for each fixed (t,q) with inverse written as $x = y - \eta(y;t,q)$ relating the pre-jump state to the post-jump state. Technically, with fixed (t,q), if $y = (I+h)(x)$ where here I denotes the identity function so $I(x) = x$, then the inverse argument is $x = (I+h)^{-1}(y) = (I-\eta)(y)$ for convenience and $\eta \stackrel{\text{op}}{=} I - (I+h)^{-1}$. Thus, $dx = (1 - \eta_y(y;t,q))dy$, where $(1-\eta_y(y;t,q))$ is the Jacobian of the inverse transformation. Further, it is assumed that the state domain $(-\infty,+\infty)$ is transformed back onto itself, modulo

the sign of the Jacobian. Consequently, we have

$$\int_{-\infty}^{+\infty} v(x)\frac{\partial\phi}{\partial t}(x,t)dx = \int_{-\infty}^{+\infty} v(x)\left(\frac{1}{2}\frac{\partial^2(g^2\phi)}{\partial x^2}(x,t) - \frac{\partial(f\phi)}{\partial x}(x,t) - (\widehat{\lambda}\phi)(x,t)\right.$$
$$+\int_{\mathcal{Q}}(\widehat{\lambda}\phi)(x-\eta(x;t,q),t)|1-\eta_x(x;t,q,t)|$$
$$\left.\cdot\phi_Q(q;x-\eta(x;t,q),t)dq\right)dx\,,$$

upon replacing y as a dummy variable in the state integral back to x so a common factor of the test function $v(x)$ can be collected. Finally, since the test function is assumed to be arbitrary, then the coefficients of $v(x)$ must be equivalent on the left and right sides of the equation *in the weak sense*. The argument is that of the *Fundamental Lemma of the Calculus of Variations* [40, 15, 163]. This leads to the forward Kolmogorov equation for the transition density $\phi(x,t) = \phi_{X(t)}(x,t;x_0,t_0)$ given in the concluding equation (7.26) of Theorem 7.5,

$$\begin{aligned}\frac{\partial\phi}{\partial t}(x,t) &= \mathcal{F}_x[\phi](x,t)\\ &\equiv \frac{1}{2}\frac{\partial^2(g^2\phi)}{\partial x^2}(x,t) - \frac{\partial(f\phi)}{\partial x}(x;t) - (\widehat{\lambda}\phi)(x,t) \qquad (7.30)\\ &\quad + \int_{\mathcal{Q}}(\widehat{\lambda}\phi)(x-\eta(x;t,q),t)|1-\eta_x(x;t,q)|\phi_Q(q;x-\eta(x;t,q),t)dq.\end{aligned}$$

Note that the subscript x on the forward operator $\mathcal{F}_x$ only denotes that the operator operates with respect to the forward variable x for jump-diffusions and only denotes partial differentiation in the pure-diffusion ($h(x,t,q) \equiv 0$) case.

The initial condition (7.27), $\phi_{X(t_0^+)}(x,t_0^+;x_0,t_0) = \delta(x-x_0)$, is very obvious for the continuous pure diffusion process, but the jump-diffusion processes undergo jumps triggered by the Poisson process $P(t;Q,X(t),t)$ and so $X(t)$ can be discontinuous. However, a jump is very unlikely in a small time interval since by (1.42) modified by replacing $\lambda(t)$ by the composite time dependence $\lambda(t;X(t),t)$,

$$\text{Prob}[dP(t;Q,X(t),t)=0] = p_0(\lambda(t;X(t),t)dt) = e^{-\lambda(t;X(t),t)dt} = 1+\text{O}(dt) \sim 1\,,$$

as $dt \to 0^+$, so the initial state is certain with probability one by conditioning, i.e.,

$$\phi(x,t) = \phi_{X(t)}(x,t;x_0,t_0) \to \delta(x-x_0) \;\text{ as }\; t \to t_0^+\,.$$

□

Remarks 7.6.

- *Another applied approach to derive the forward equation for pure diffusions is to use the diffusion approximation as given by Feller [84], but this requires strong assumptions about truncating a Taylor expansion just for diffusion processes alone. This approach does not apply to jump-diffusions, since the jump difference term $D_h[\phi]$ would require an infinite expansion.*

- *For the jump amplitude, a good illustration could be the affine model that is the sum of a state-independent term plus a term purely linear in the state, i.e.,* $h(x,t,q) = \nu_0(t,q) + \nu_1(t,q)x$ *for suitable time-mark coefficients, so the inverse of* $y = x + h(x,t,q)$ *is* $x = (y - \nu_0(t,q))/(1 + \nu_1(t,q)) = y - \eta(y;t,q)$ *and* $\eta(y;t,q) = (\nu_0(t,q) + \nu_1(t,q)y)/(1 + \nu_1(t,q))$*. For comparison, different cases of this model are tabulated in Table 7.1.*

Table 7.1. *Some Simple jump amplitude models and inverses.*

State Dependence	Direct $h(x,t,q)$	Forward Arg. x=$y - \eta(y;t,q)$	Inverse $\eta(y;t,q)$
constant	$\nu_0(t,q)$	$y - \nu_0(t,q)$	$\nu_0(t,q)$
pure linear	$\nu_1(t,q)x$	$\frac{y}{1+\nu_1(t,q)}$	$\frac{\nu_1(t,q)y}{1+\nu_1(t,q)}$
affine	$\nu_0(t,q) + \nu_1(t,q)x$	$\frac{y-\nu_0(t,q)}{1+\nu_1(t,q)}$	$\frac{\nu_0(t,q)+\nu_1(t,q)y}{1+\nu_1(t,q)}$

A mistake is sometimes made by incorrectly generalizing the inverse of the linear jump case $x + \nu_1(t,q)x = y$*, so that* $(1 - \nu_1(t,q))y$ *is incorrectly used for the forward argument (x) in the linear case instead of the correct argument, which is* $x = y/(1 + \nu_1(t,q))$*.*

- *The difference in the jump argument between the backward and forward equation is that in the backward case the post-jump or forward value* $y = x + h(x,t,q)$ *is used, while in the forward case the pre-jump or backward value* $x = y - h(x,t,q) = y - \eta(y;t,q)$ *is used.*

7.4 Multi-dimensional Backward and Forward Equations

For many applications, there can be multiple state variables and multiple sources of random disturbances. In biological problems there can be several interacting species each suffering from species specific and common random changes, that can be detrimental or beneficial in effect and range in magnitude from small to large fluctuations. Such effects may be due to the weather, diseases, natural disasters or inter-species predation. In finance, there are the usual background fluctuations in market values, and then there is the occasional market crash or buying frenzy. In manufacturing systems, there may be a large number of machines which randomly fail with the time to repair being randomly distributed due to the many causes of failure.

Consider again the multi-dimensional SDE from Chapter 5 for the n_x-dimensional state process $\mathbf{X}(t) = [X_i(t)]_{n_x \times 1}$,

$$d\mathbf{X}(t) \overset{\text{sym}}{=} \mathbf{f}(\mathbf{X}(t),t)dt + g(\mathbf{X}(t),t)d\mathbf{W}(t) + h(\mathbf{X}(t),t,\mathbf{Q})d\mathbf{P}(t;Q,\mathbf{X}(t),t)\,, \quad (7.31)$$

where

$$\mathbf{W}(t) = [W_i(t)]_{n_w \times 1}$$

is an n_w-dimensional vector diffusion process and

$$\mathbf{P}(t; Q, \mathbf{X}(t), t) = [P_i(t; Q_i, \mathbf{X}(t), t)]_{n_p \times 1}$$

is an n_p-dimensional vector state-dependent Poisson jump process. The state-dependent coefficient functions are dimensionally specified by

$$\begin{aligned} \mathbf{f} &= [f_i(\mathbf{X}(t), t)]_{n_x \times 1}\,, \\ g(\mathbf{X}(t), t) &= [g_{i,j}(\mathbf{X}(t), t)]_{n_x \times n_w}\,, \\ h(\mathbf{X}(t), t, \mathbf{Q}) &= [h_{i,j}(\mathbf{X}(t), t, Q_j)]_{n_x \times n_p} \end{aligned}$$

and have dimensions that are commensurate in multiplication. The mark vector, $\mathbf{Q} = [Q_i)]_{n_p \times 1}$, in the last coefficient function is assumed to have components corresponding to all Poisson vector process components. The coefficient $h(\mathbf{X}(t), t, \mathbf{Q})$ of $d\mathbf{P}(t; Q, \mathbf{X}(t), t)$ is merely the mark $\mathbf{Q}$ dependent symbolic form of the jump amplitude operator-coefficient $h(\mathbf{X}(t), t, \mathbf{q})$, using similar notation, in the corresponding Poisson random mark integral (5.83), i.e.,

$$h(\mathbf{X}(t), t, \mathbf{Q}) d\mathbf{P}(t; Q, \mathbf{X}(t), t) \stackrel{\text{sym}}{=} \int_{\mathcal{Q}} h(\mathbf{X}(t), t, \mathbf{q}) \boldsymbol{\mathcal{P}}(\mathbf{dt}, \mathbf{dq}; \mathbf{X}(t), t).$$

Dynkin's formula remains unchanged, except for converting the state variable $X(t)$ to a vector $\mathbf{X}(t)$ and making the corresponding change in the backward operator $\mathcal{B}_{\mathbf{x}}[v]$ using the multi-dimensional stochastic chain rule (5.98),

$$\begin{aligned} \overline{v}(\mathbf{x}_0, t_0; t) &\equiv \mathrm{E}[v(\mathbf{X}(t))|\mathbf{X}(t_0) = \mathbf{x}_0] \\ &= v(\mathbf{x}_0) + \mathrm{E}\left[\int_{t_0}^{t} \mathcal{B}_{\mathbf{x}}[v](\mathbf{X}(s); \mathbf{X}(s), s) ds \middle| \mathbf{X}(t_0) = \mathbf{x}_0\right], \end{aligned} \tag{7.32}$$

where the backward operator is given below. The multi-dimensional backward and forward Kolmogorov equations are summarized in the following theorem, with the justification left as an exercise for the reader.

Theorem 7.7. ***Kolmogorov Equations for Jump-Diffusions in Multi-dimensions on*** $[t_0, t]$***:***
Let

$$u(\mathbf{x}_0, t_0) = \overline{v}(\mathbf{x}_0, t_0; t) = \mathrm{E}[v(\mathbf{X}(t))|\mathbf{X}(t_0) = \mathbf{x}_0].$$

Then $u(\mathbf{x}_0, t_0)$ satisfies the following multi-dimensional backward Kolmogorov PDE with backward arguments,

$$0 = \frac{\partial u}{\partial t_0}(\mathbf{x}_0, t_0) + \mathcal{B}_{\mathbf{x}_0}[u](\mathbf{x}_0, t_0; \mathbf{x}_0, t_0)\,, \tag{7.33}$$

where the backward Kolmogorov operator is defined as

$$\begin{aligned}\mathcal{B}_{\mathbf{x}_0}[u](\mathbf{x}_0,t_0;\mathbf{x}_0,t_0) \equiv\ & \mathbf{f}^\top(\mathbf{x}_0,t_0)\nabla_{\mathbf{x}_0}[u](\mathbf{x}_0,t_0)\\ &+\frac{1}{2}\left(gR'g^\top\right):\nabla_{\mathbf{x}_0}\left[\nabla^\top_{\mathbf{x}_0}[u]\right](\mathbf{x}_0,t_0)\\ &+\sum_{j=1}^{n_p}\widehat{\lambda}_j(\mathbf{x}_0,t_0)\int_{\mathcal{Q}}\Delta_j[u](\mathbf{x}_0,t_0,q_j)\phi_{Q_j}(q_j;\mathbf{x}_0,t_0)dq_j,\end{aligned} \tag{7.34}$$

where R' is a correlation matrix defined in (5.95), $A:B$ is the double dot product (5.99),

$$\Delta_j[u](\mathbf{x}_0,t_0,q_j) \equiv u(\mathbf{x}_0+\widehat{\boldsymbol{h}}_j(\mathbf{x}_0,t_0,q_j),t_0)-u(\mathbf{x}_0,t_0)$$

is the jump of u corresponding to the jump amplitude

$$\widehat{\boldsymbol{h}}_j(\mathbf{x},t,q_j) \equiv [h_{i,j}(\mathbf{x},t,q_j)]_{n_x\times 1}$$

of the jth Poisson process P_j at the jth mark for $j=1:n_p$ and with final condition

$$u(\mathbf{x}_0,t^-) = \overline{v}(\mathbf{x}_0,t^-;t) = v(\mathbf{x}_0)\,.$$

Similarly, the forward Kolmogorov PDE in the multi-dimensional transition density $\phi(\mathbf{x},t;\mathbf{x}_0,t_0)$ as the adjoint of the backward equation is

$$\frac{\partial\phi}{\partial t}(\mathbf{x},t) = \mathcal{F}_{\mathbf{x}}[\phi](\mathbf{x},t)\,, \tag{7.35}$$

where the forward Kolmogorov operator is defined as

$$\begin{aligned}\mathcal{F}_{\mathbf{x}}[\phi](\mathbf{x},t) \equiv\ & \frac{1}{2}\nabla_{\mathbf{x}}\left[\nabla^\top_{\mathbf{x}}:\left[gR'g^\top\phi\right]\right](\mathbf{x},t)\\ &-\nabla^\top_{\mathbf{x}}[\mathbf{f}\phi](\mathbf{x};t)-\sum_{j=1}^{n_p}(\widehat{\lambda}_j\phi)(\mathbf{x},t)\\ &+\sum_{j=1}^{n_p}\int_{\mathcal{Q}}(\widehat{\lambda}_j\phi)(\mathbf{x}-\boldsymbol{\eta}_j(\mathbf{x};t,q_j),t)\left|1-\frac{\partial(\boldsymbol{\eta}_j(\mathbf{x};t,q_j))}{\partial(\mathbf{x})}\right|\\ &\cdot\phi_{Q_j}(q_j;\mathbf{x}-\boldsymbol{\eta}_j(\mathbf{x};t,q_j),t)dq_j\,,\end{aligned} \tag{7.36}$$

where the backward to forward transformation and its Jacobian are

$$\mathbf{x}-\mathbf{x}_0 = \boldsymbol{\eta}_{j'}(\mathbf{x},t,q_{j'}) = \widehat{\boldsymbol{h}}_{j'}(\mathbf{x}_0,t,q_{j'})\,;$$

$$\frac{\partial(\boldsymbol{\eta}_{j'}(\mathbf{x};t,q_{j'}))}{\partial(\mathbf{x})} = \mathrm{Det}\left[\left[\frac{\partial\eta_{j',i}(\mathbf{x};t,q_{j'})}{\partial x_j}\right]_{n_x\times n_x}\right] = \mathrm{Det}\left[\left(\nabla_{\mathbf{x}}\left[\boldsymbol{\eta}^\top_{j'}\right]\right)^\top\right]$$

for $j'=1:n_p$.

7.5 Chapman-Kolmogorov Equation for Markov Processes in Continuous Time

Alternate methods for deriving the Kolmogorov equations are based upon a fundamental functional equation of Chapman and Kolmogorov (see Bharucha-Reid [31] or other references at the end of this chapter). Let $\mathbf{X}(t)$ be a $n_x \times 1$ Markov process in continuous time, i.e., a jump-diffusion, on the state space Ω. The transition probability distribution function is given by

$$\Phi(\mathbf{x}, t; \mathbf{x}_0, t_0) = \text{Prob}[\mathbf{X}(t) < \mathbf{x} \mid \mathbf{X}(t_0) = \mathbf{x}_0] \,, \tag{7.37}$$

provided $t > t_0$, $\mathbf{X}(t) < \mathbf{x}$ means $X_i(t) < x_i$ for $i = 1 : n_x$, and assuming the probability density exists even if in the generalized sense,

$$\phi(\mathbf{x}, t; \mathbf{x}_0, t_0) = \left(\prod_{i=1}^{n_x} \frac{\partial \phi}{\partial x_i} \right) (\mathbf{x}, t; \mathbf{x}_0, t_0) \,. \tag{7.38}$$

Expressed as a Markov property for distributions, the *Chapman-Kolmogorov equation* for the transition between the start $(\mathbf{x}_0, t_0)$ and the current position $(\mathbf{x}, t)$ through all possible intermediate positions $(\mathbf{y}, s)$ is

$$\begin{aligned} \Phi(\mathbf{x}, t; \mathbf{x}_0, t_0) &= \int_\Omega \Phi(\mathbf{y}, s; \mathbf{x}_0, t_0) \Phi(\mathbf{x}, t; d\mathbf{y}, s) \\ &= \int_\Omega \Phi(\mathbf{y}, s; \mathbf{x}_0, t_0) \phi(\mathbf{x}, t; \mathbf{y}, s) d\mathbf{y} \,, \end{aligned} \tag{7.39}$$

where $t_0 < s < t$. Alternately, the *Chapman-Kolmogorov equation* solely in terms of transition probability densities is

$$\phi(\mathbf{x}, t; \mathbf{x}_0, t_0) = \int_\Omega \phi(\mathbf{y}, s; \mathbf{x}_0, t_0) \phi(\mathbf{x}, t; \mathbf{y}, s) d\mathbf{y} \,, \tag{7.40}$$

upon differentiating (7.39) according to (7.38), again with $t_0 < s < t$. See Bharucha-Reid [31] or other references at the end of this chapter for applications.

7.6 Jump-Diffusion Boundary Conditions

Many boundary value problems for stochastic diffusion processes are similar to their deterministic counterparts, but the stochastic justifications are different. When jump processes are included, then the situation is even more complicated. Since jump processes are discontinuous, jumps may over shoot the boundary making it more difficult to construct an auxiliary process that will implement the boundary with proper probability law.

7.6.1 Absorbing Boundary Condition

If the boundary is absorbing, i.e., the process that hits the boundary stays there [84, 98, 244, 162], it is quite easy to specify since the process can not reenter the

interior and the transition probability for the process initially at $\mathbf{X}(0) = \mathbf{x}_0$ on the boundary $\Gamma = \partial\Omega$ can not reach $\mathbf{X}(t) = \mathbf{y}$ in the interior of the domain Ω. Thus, for pure-diffusions

$$\phi_{\mathbf{X}(t)}(\mathbf{x}, t; \mathbf{x}_0, t_0) = \text{Prob}[\mathbf{X}(t) = \mathbf{x} \in \Omega | \mathbf{X}(t_0) = \mathbf{x}_0 \in \Gamma, t > 0] = 0, \qquad (7.41)$$

whereas for jump-diffusions

$$\phi_{\mathbf{X}(t)}(\mathbf{x}, t; \mathbf{x}_0, t_0) = \text{Prob}[\mathbf{X}(t) = \mathbf{x} \in \Omega | \mathbf{X}(0) = \mathbf{x}_0 \notin \text{Interior}[\Omega], t > 0] = 0, \quad (7.42)$$

since it is assumed that a jump over-shoot into the boundary or exterior of the region is absorbed. Kushner and Dupuis [179] have a more elaborate treatment of the absorbing boundary by stopping the process once it hits the boundary, assumed to be smooth and reachable in finite time (also called attainable or accessible). These are boundary conditions for the transition probability density backward equations, since they are specified on the backward variable x_0.

7.6.2 Reflecting Boundary Conditions

The reflecting boundary is much more complicated and the smoothness of the boundary, i.e., the boundary is continuously differentiable, is important for defining the reflection. Since a simple reflection at a boundary point, $\mathbf{x}_b$, will be in the plane of the nearby incoming trajectory at $\mathbf{x}_0$ and the normal vector $\mathbf{N}_b$ to the tangent plane of the boundary at $\mathbf{x}_b$. Let $\delta\mathbf{x} = \mathbf{x}_0 - \mathbf{x}_b$ be the distance vector to the point of contact and let $\mathbf{T}_b$ a tangent vector in the intersection of the tangent plane and the trajectory-normal plane. Using *stochastic reflection principle*, similar to the reflection principle used in PDEs, a stochastic reflection process is constructed such that $\delta\mathbf{x}_r = \mathbf{x}_r - \mathbf{x}_b$ is its current increment at the same time as $\delta\mathbf{x}$. The only difference is the opposite sign of its normal component, i.e., $\delta\mathbf{x}_r = -\delta_n\mathbf{N}_b + \delta_t\mathbf{T}_b$ if $\delta\mathbf{x}_0 = +\delta_n\mathbf{N}_b + \delta_t\mathbf{T}_b$, for sufficiently small and positive components d_n and δ_t. Since the reflected process at $\mathbf{x}_r$ by its construction must have the same probability as the original process at $\mathbf{x}_0$, then

$$\mathbf{N}_b^\top \nabla_{x_0} \left[\phi_{\mathbf{X}(t)}\right](\mathbf{x}, t\mathbf{x}_b, t_0) = \mathbf{N}_b^\top \nabla_{x_0} \left[\widehat{\phi}\right](\mathbf{x}_b, t_0) = 0\,, \qquad (7.43)$$

upon expanding the difference between the two probability densities

$$\widehat{\phi}(\mathbf{x}_0, t_0') - \widehat{\phi}(\mathbf{x}_r, t_0') = \widehat{\phi}(\mathbf{x}_b + \delta_n\mathbf{N}_b + \delta_t\mathbf{T}_b, t_0') - \widehat{\phi}(\mathbf{x}_b - \delta_n\mathbf{N}_b + \delta_t\mathbf{T}_b, t_0') = 0,$$

in simplified backward notation at pre-hit time t_0' here, to order δ_n. The order δ_t cancels out.

See Kushner and Dupuis [179] about more reflecting boundary conditions and systematically constructing reflecting jump-diffusion processes. Also, see Karlin and Taylor [162] for a thorough discussion of other boundary conditions such as sticky and elastic, as well as an extensive boundary classification for pure diffusion problems.

7.7 Stopping Times: Expected Exit and First Passage Times

In many problems, an *exit time*, also called a *stopping time* or a *first passage time*, is of interest. For instance when a population falls to the zero level and thus ceases to exist, it is said to be extinct and the time of extinction is of interest. If it is a stochastic population, then the expected extinction time is of interest (Hanson and Tuckwell [119, 121]). For a neuron, stochastic fluctuations can be important and then the time to reach a threshold to fire a nerve pulse is of interest and in particular the expected firing time can be calculated (Stein [257], Tuckwell [269], Hanson and Tuckwell [120]). In cancer growth studies, the expected doubling time for the size of a tumor is often calculated (Hanson and Tier [117]). There are many other example of stopping times. First deterministic exit time problems are introduced as examples and as a basic reference.

Examples 7.8. ***Deterministic Exit Time Problems***

- ***Forward Exit Time Formulation:***
  Let $X(t)$ be the state of the system at time t and be governed by the ODE

$$\frac{dX}{dt}(t) = f(X(t)), \quad X(0) = x_0 \in (a, b), \tag{7.44}$$

  where $f(x)$ is strictly positive or strictly negative, $f(x)$ is continuous and $1/f(x)$ is integrable on $[a, b]$. Thus inverting 7.44, the forward running time is

$$dt = dT_F(x) = dx/f(x), \quad T_F(x_0) = 0,$$

  so

$$T_F(x) = \int_{x_0}^{x} dy/f(y),$$

  and the forward exit time is

$$T_F(b) \quad \text{if} \quad f(x) > 0 \quad \text{or} \quad T_F(a) \quad \text{if} \quad f(x) < 0.$$

- ***More Relevant Backward Exit Time Formulation:***
  Since the stochastic exit time problem is more conveniently formulated as a backward time problem, let $x = c$ be the point of exit, so when $x_0 = c$ then we know the state $X(t)$ is already at the exit and the final condition is $T_B(c) \equiv 0$. Consequently, the backward exit time $T_B(x)$ problem is formulated with $T_B(x) = T_F(c) - T_F(x)$ or $T_B'(x) = -T_F'(x)$ as

$$dT_B(x) = -dx/f(x), \quad T_B(c) = 0$$

  or in the more conventional backward form,

$$f(x)T_B'(x) = -1, \quad T_B(c) = 0, \tag{7.45}$$

so

$$T_B(x) = -\int_c^x dy/f(y)$$

or the backward exit time ending at $x = c$ is

$$T_B(x_0) = \int_{x_0}^c dy/f(y)$$

where $c = b$ if $f(x) > 0$ or $c = a$ if $f(x) < 0$.

7.7.1 Expected Stochastic Exit Time

First, the exit time is analytically defined, relevant for the piece-wise continuous jump-diffusion. For continuous, pure diffusion processes, it is sufficient to consider when the process hits a boundary. However, when the stochastic process also includes jumps, then it is possible that the process overshoots the boundary and ends up in the exterior of the domain. Here the domain will simply be an open interval in one state dimension.

Again let $X(t)$ be a jump-diffusion process satisfying the SDE,

$$dX(t) \stackrel{\text{sym}}{=} f(X(t),t)dt + g(X(t),t)dW(t) + h(X(t),t,Q)dP(t;Q,X(t),t) \quad (7.46)$$

with smooth (continuously differentiable) coefficients $\{f, g, h\}$ with bounded spatial gradients.

Definition 7.9. *In one state dimension, the exit time for the Markov process $X(t)$ in continuous time (7.46) from the open interval (a, b) is*

$$\tau_e(x_0, t_0) \equiv \inf_t [\ t\ |X(t) \notin (a,b); X(t_0) = x_0 \in (a,b)]\,, \quad (7.47)$$

if it exists.

Before considering a more general formulation using probability theory, some applications of Dynkin's formula will be used to compute the expected extinction time and some higher moments.

Examples 7.10. ***Expected Exit Time Applications of Dynkin's Formula:***

- ***Small modification of Dynkin's formula for exit times:***
  Consider the following boundary value problem of inhomogeneous backward Kolmogorov equation,

$$\frac{\partial v}{\partial t_0}(x_0,t_0) + \mathcal{B}_{x_0}[v](x_0,t_0) = \alpha(x_0,t_0), \quad x_0 \in (a,b)\,, \quad (7.48)$$

$$v(x_0,t_0) = \beta(x_0,t_0), \quad x_0 \notin (a,b)\,, \quad (7.49)$$

where $\mathcal{B}_{x_0}[v](x_0,t_0)$ (7.14) is the jump-diffusion backward operator, $\alpha(x_0,t_0)$ is a given general state-independent homogeneous term and $\beta(x_0,t_0)$ is a given general exit boundary value. Both $\alpha(x_0,t_0)$ and $\beta(x_0,t_0)$ depend on the application. Sometimes (7.48) is called Dynkin's equation due to its relationship with Dynkin's formula.

Prior to taking expectations, the integral form (7.9) of the stochastic chain rule was

$$\begin{aligned} v(X(t),t) = v(x_0,t_0) + \int_{t_0}^{t} \Bigg(& \left(\frac{\partial v}{\partial t} + f\frac{\partial v}{\partial x} + \frac{1}{2}g^2\frac{\partial^2 v}{\partial x^2} \right)(X(s),s)ds \\ & + \left(g\frac{\partial v}{\partial x} \right)(X(s),s)dW(s) \\ & + \int_{\mathcal{Q}} \Delta_h[v](X(s),s,q)\mathcal{P}(\mathbf{ds},\mathbf{dq};X(s),s) \Bigg), \end{aligned} \tag{7.50}$$

but now make the random exit time substitution $t = \tau_e(x_0,t_0)$ for the deterministic time variable which is simply abbreviated as $t = \tau_e$ and then take expectations getting an exit time version of Dynkin's formula,

$$\begin{aligned} E\left[v(X(\tau_e),\tau_e)|X(t_0)=x_0\right] = v(x_0,t_0) & \\ + \mathrm{E}\left[\int_{t_0}^{\tau_e} \left(\frac{\partial v}{\partial t} + \mathcal{B}_x[v] \right)(X(s),s)ds \right]. & \end{aligned} \tag{7.51}$$

Upon substituting Dynkin's equation (7.48) into Dynkin' Formula, it reduces to

$$E\left[\beta(X(\tau_e),\tau_e)|X(t_0)=x_0\right] = v(x_0,t_0) + \mathrm{E}\left[\int_{t_0}^{\tau_e} \alpha(X(s),s)ds \right]. \tag{7.52}$$

- ***Ultimate Exit Time Distribution:***
  Let $\alpha(x_0,t_0) = 0$, while $\beta(X(\tau_e),\tau_e) = 1$ since if x_0 starts at an exit, i.e., $x_0 \notin (a,b)$, then exit is certain and the distribution function is 1. Hence, due to the jump-diffusion $v(x_0,t_0) = 1 = \Phi_{\tau_e(x_0,t_0)}(+\infty)$ on (a,b) under reasonable conditions for the existence of an exit.

- ***Expected Exit Time:***
  Assuming that exit is certain, $\Phi_{\tau_e(x_0,t_0)}(+\infty) = 1$, let $\alpha(x_0,t_0) = -1 = -\Phi_{\tau_e(x_0,t_0)}(+\infty)$ and $\beta(X(\tau_e),\tau_e) = 0$, corresponding to $x_0 \notin (a,b)$ implying zero exit time, then

$$\mathrm{E}[\tau_e(x_0,t_0)] = t_0 + v^{(1)}(x_0,t_0)\,, \tag{7.53}$$

  where $v^{(1)}(x_0,t_0)$ is the solution to the problem (7.48-7.49) with $\alpha(x_0,t_0) = 0$ and $\beta(X(\tau_e),\tau_e) = 0$.

- ***Second Moment of Exit Time:***
  Assuming that exit is certain, let $\alpha(x_0,t_0) = -2t_0$ and $\beta(X(\tau_e),\tau_e) = 0$ again, then

$$\mathrm{E}[\tau_e^2(x_0,t_0)] = t_0^2 + v^{(2)}(x_0,t_0)\,, \tag{7.54}$$

where $v^{(2)}(x_0,t_0)$ is the solution to the problem (7.48-7.49) with $\alpha(x_0,t_0) = -2t_0$ and $\beta(X(\tau_e),\tau_e) = 0$. Hence, the variance of the exit time on (a,b) is

$$\begin{aligned}\mathrm{Var}[\tau_e(x_0,t_0)] &= \mathrm{E}[\tau_e^2(x_0,t_0)] - \mathrm{E}^2[\tau_e(x_0,t_0)] \\ &= v^{(2)}(x_0,t_0) - 2t_0 v^{(1)}(x_0,t_0) - (v^{(1)})^2(x_0,t_0)\end{aligned}$$

and the coefficient of variation (CV) of the exit time is

$$\begin{aligned}\mathrm{CV}[\tau_e(x_0,t_0)] &= \frac{\sqrt{\mathrm{Var}[\tau_e(x_0,t_0)]}}{\mathrm{E}[\tau_e(x_0,t_0)]} \\ &= \frac{\sqrt{v^{(2)}(x_0,t_0) - 2t_0 v^{(1)}(x_0,t_0) - (v^{(1)})^2(x_0,t_0)}}{v^{(1)}(x_0,t_0) + t_0}\,.\end{aligned}$$

- ***Higher Moments of Exit Time:***
  Assuming that exit is certain, let $\alpha(x_0,t_0) = -nt_0^{n-1}$ and again $\beta(X(\tau_e),\tau_e) = 0$, then

$$\mathrm{E}[\tau_e^n(x_0,t_0)] = t_0^n + v^{(n)}(x_0,t_0)\,, \tag{7.55}$$

  where $v^{(n)}(x_0,t_0)$ is the solution to the problem (7.48-7.49) with $\alpha(x_0,t_0) = -nt_0^{n-1}$ and $\beta(X(\tau_e),\tau_e) = 0$.

Often conditional exit time moments are of interest, but then the inhomogeneous term $\alpha(x_0,t_0)$ genuinely depends on the state x_0 which makes the (7.51) form of Dynkin's formula not too useful since then the $\alpha(X(s),s)$ in the integrand genuinely depends on the stochastic process $X(s)$ and the integral is no longer simple. Hence, for more conditional and more general problems a more general form is needed. This more general form is based upon a generalization of the time-homogeneous derivations in Schuss [244] and in the appendix of Hanson and Tier [117] to the time dependent coefficient case, obtaining a hybrid backward or Dynkin equation for the exit time density $\phi_{\tau_e(x_0,t_0)}(t)$.

Lemma 7.11. ***Exit Time Distribution and Density:***
Given the exit time $\tau_e(x_0,t_0)$ (7.47), then its probability distribution can be related to the distribution for $X(t)$ by

$$\Phi_{\tau_e(x_0,t_0)}(t) = 1 - \int_a^b \phi_{X(t)}(x,t;x_0,t_0)dx, \tag{7.56}$$

where $\phi_{X(t)}(x,t;x_0,t_0)$ is the transition probability density for the Markov process $X(t) = x$ in continuous time conditionally starting at $X(t_0) = x_0$, as given in (7.18). The density $\phi_{X(t)}(x,t;x_0,t_0)$ is assumed to exist.

Assuming the exit time distribution and the transition density are differentiable even in a generalized sense, the exit time probability density is

$$\phi_{\tau_e(x_0,t_0)}(t) = \frac{\partial \Phi_{\tau_e(x_0,t_0)}}{\partial t}(t)\,.$$

The $\phi_{X(t)}$ transition density is assumed to be twice differentiable in x_0 and once in t, leading to the Kolmogorov equation in the forward time but with the backward operator $\mathcal{B}_{x_0}$,

$$\begin{aligned}\frac{\partial \phi_{\tau_e(x_0,t_0)}}{\partial t}(t) &= \mathcal{B}_{x_0}\left[\phi_{\tau_e(x_0,t_0)}(t)\right] \\ &= f(x_0,t_0)\frac{\partial \phi_{\tau_e(x_0,t_0)}}{\partial x_0}(t) + \frac{1}{2}g^2(x_0,t_0)\frac{\partial^2 \phi_{\tau_e(x_0,t_0)}}{\partial x_0^2}(t) \\ &\quad +\widehat{\lambda}(x_0,t_0)\int_{\mathcal{Q}} \Delta_h[\phi_{\tau_e(x_0,t_0)}(t)](x_0,t_0,q)\phi_Q(q;x_0,t_0)dq,\end{aligned} \tag{7.57}$$

where the jump function Δ_h is given in (7.4).

Proof. The Eq. (7.56) for the exit time distribution follows from the probability definitions

$$\begin{aligned}\Phi_{\tau_e(x_0,t_0)}(t) &= \text{Prob}[\tau_e(x_0,t_0) < t] = \text{Prob}[X(t) \notin (a,b)|X(t_0) = x_0] \\ &= 1 - \text{Prob}[X(t) \in (a,b)|X(t_0) = x_0] \\ &= 1 - \int_a^b \phi_{X(t)}(x,t;x_0,t_0)dx,\end{aligned}$$

i.e., the fact that the exit time probability is the complement of the probability that the process $X(t)$ is in the interval (a,b) and thus yields the right-hand side of (7.56).

Under differentiability assumptions, the exit time density can be related to an integral of the forward operator $\mathcal{F}_x$ using the forward Kolomogorov

$$\begin{aligned}\phi_{\tau_e(x_0,t_0)}(t) = \frac{\partial \Phi_{\tau_e(x_0,t_0)}}{\partial t}(t) &= -\int_a^b \phi_{X(t),t}(x,t;x_0,t_0)dx \\ &= -\int_a^b \mathcal{F}_x[\phi](x,t;x_0,t_0)dx\,.\end{aligned}$$

Manipulating partial derivatives, first in forward form,

$$\phi_{X(t),t}(x,t;x_0,t_0)) = \phi_{X(t),t-t_0}(x,t;x_0,t_0) = -\phi_{X(t),t_0-t}(x,t;x_0,t_0)$$

and then in backward form,

$$\phi_{X(t),t_0}(x,t;x_0,t_0) = \phi_{X(t),t_0-t}(x,t;x_0,t_0)\,,$$

leads to

$$\phi_{\tau_e(x_0,t_0)}(t) = +\int_a^b \phi_{X(t),t_0}(x,t;x_0,t_0)dx = -\int_a^b \mathcal{B}_{x_0}[\phi](x,t;x_0,t_0)dx\,.$$

Again assuming sufficient differentiability along with the interchange of integral and

differential operators,

$$\begin{aligned}\phi_{\tau_e(x_0,t_0),t}(t) &= -\int_a^b \mathcal{B}[\phi_{X(t),t}(x,t;x_0,t_0)]dx \\ &= -\int_a^b \mathcal{B}_{x_0}[\mathcal{F}[\phi_{X(t)}]](x,t;x_0,t_0)dx \\ &= -\mathcal{B}_{x_0}\left[\int_a^b \mathcal{F}[\phi_{X(t)}](x,t;x_0,t_0)dx\right] = +\mathcal{B}_{x_0}\left[\phi_{\tau_e(x_0,t_0)}(t)\right] .\end{aligned}$$

This is a hybrid Kolmogorov equation (7.57), since it is in forward time t on the left and the backward operator is on the far right. □

Examples 7.12. ***Conditionally Expected Exit Time Applications:***

- ***Ultimate Probability of Exit:***
  The ultimate probability of exit is

$$\Phi_e(x_0,t_0) \equiv \Phi_{\tau_e(x_0,t_0)}(+\infty) = \int_0^\infty \phi_{\tau_e(x_0,t_0)}(t)dt\,, \tag{7.58}$$

  assuming that the distribution is bounded for all t. Also under the same conditions,

$$\int_0^\infty \phi_{\tau_e(x_0,t_0),t}(t)dt = \phi_{\tau_e(x_0,t_0)}(t)\Big|_0^{+\infty} = 0$$

  and then from the exit time density equation (7.57), integration-operator interchange and (7.58) for $\Phi_e(x_0,t_0)$,

$$\int_0^\infty \mathcal{B}[\phi_{\tau_e(x_0,t_0)}(t)]dt = \mathcal{B}[\Phi_e(x_0,t_0)] = 0\,. \tag{7.59}$$

  For certain exit at both endpoints a and b, the obvious boundary conditions are $\Phi_e(a,t_0) = 1$ and $\Phi_e(b,t_0) = 1$ for continuous diffusion processes, but $[\Phi_e(x_0,t_0)]$ =1 for $x_0 \notin (a,b)$ for jump-diffusions. Presuming uniqueness, then the solution to the boundary value problem is $\Phi_e(x_0,t_0) = 1$.

- ***Conditional Exit on the Right of (a,b):*** *Now suppose the statistics of ultimate exit on one side of (a,b), say $x_0 \in [b,+\infty)$, i.e., on the right. The corresponding random exit time variable is*

$$\tau_e^{(b)}(x_0,t_0) = \inf_t[\ t\ |X(t) \geq b,\ X(s) \in (a,b),\ t_0 \leq s < t,\ X(t_0) = x_0]\,,$$

  and the exit time distribution function is

$$\Phi_{\tau_e^{(b)}(x_0,t_0)}(t) \equiv \text{Prob}[\tau_e^{(b)}(x_0,t_0) < t]$$

and the corresponding density is $\phi_{\tau_e^{(b)}(x_0,t_0)}(t)$*. Thus, the ultimate conditional distribution,*

$$\Phi_e^{(b)}(x_0,t_0) \equiv \int_0^{+\infty} \phi_{\tau_e^{(b)}(x_0,t_0)}(t)dt,$$

for counting only exits on the right, has boundary conditions $\Phi_e^{(b)}(x_0,t_0) = 1$ *if* $x_0 \in [b,+\infty)$*, but* $\Phi_e^{(b)}(x_0,t_0) = 0$ *if* $x_0 \in (-\infty,a]$*. (For counting only exits at the left,* $(-\infty,a]$*, then the boundary conditions are interchanged for* $\Phi_e^{(a)}(x_0,t_0)$*.) In general, the conditional distribution* $\Phi_e^{(b)}(x_0,t_0)$ *will not be one as in the certain ultimate probability in the prior item, so it is necessary to work in exit time moments rather than expected exit times. Let the conditional exit time first moment be*

$$M_e^{(b)}(x_0,t_0) \equiv \int_0^{+\infty} t\phi_{\tau_e^{(b)}(x_0,t_0)}(t)dt \tag{7.60}$$

and the expected conditional exit time is

$$T_e^{(b)}(x_0,t_0) \equiv M_e^{(b)}(x_0,t_0)/\Phi_e^{(b)}(x_0,t_0) \tag{7.61}$$

if $x_0 > a$*. Upon integration of both sides of (7.57), making the reasonable assumption*

$$t\phi_{\tau_e^{(b)}(x_0,t_0)}(t)\Big|_0^{+\infty} = 0$$

when apply integration by parts on the left, then the conditional moment equation, interchanging left and right sides, is

$$\mathcal{B}_{x_0}\left[M_e^{(b)}\right](x_0,t_0) = -\Phi_e^{(b)}(x_0,t_0) \tag{7.62}$$

with boundary condition $M_e^{(b)}(x_0,t_0) = 0$ *if* $x_0 \notin (a,b)$*. The conditions are zero on either side of* (a,b) *for different reasons, due to instant exit for* $x_0 \in [b,+\infty)$ *and due to excluded exit for* $x_0 \in (-\infty,a]$*.*

7.8 Diffusion Approximation Basis

Up until this point, stochastic diffusions have almost been taken as given. There are many derivations for physical diffusions in physics and engineering, such as the diffusion of a fluid concentration in a liquid or gas according to Fick's law for the flux or flow of concentration or the diffusion of heat in a conduction medium according to Fourier's law for the flux of heat. These types of physical diffusions lead to the same or similar diffusion equations as seen in this chapter when the jump terms are omitted. However, the stochastic diffusions are usually postulated on a different basis.

A fundamental property that distinguishes the pure diffusion process from the discontinuous jump process among Markov processes in continuous time is that the

diffusion process is a continuous process. Let $\mathbf{X}(t) = [X_i(t)]_{n_x \times 1}$ be a continuous process, then it must satisfy the following continuity condition, given some $\delta > 0$,

$$\lim_{\Delta t \to 0} \frac{\text{Prob}[|\Delta \mathbf{X}(t)| > \delta \mid \mathbf{X}(t) = \mathbf{x}]}{\Delta t} = 0\,, \tag{7.63}$$

so jumps in the process are unlikely.

In addition, two basic moment properties are needed for the continuous process to have a diffusion limit and these are that the conditional mean increment process satisfy

$$\begin{aligned} \text{E}[\Delta \mathbf{X}(t) | \mathbf{X}(t) = \mathbf{x}] &= \int_\Omega \phi_{X(t)}(\mathbf{y}, t + \Delta t; \mathbf{x}, t) d\mathbf{y}) \\ &= \boldsymbol{\mu}(\mathbf{x}, t)\Delta t + \text{o}(\Delta t) \quad \text{as} \quad \Delta t \to 0\,, \end{aligned} \tag{7.64}$$

where $\boldsymbol{\mu}(\mathbf{x}, t) = [\mu_i(\mathbf{x}, t)]_{n_x \times 1}$, and that the conditional variance increment process satisfy

$$\text{Cov}[\Delta \mathbf{X}(t), \Delta \mathbf{X}^\top(t) | \mathbf{X}(t) = \mathbf{x}] = \sigma(\mathbf{x}, t)\Delta t + \text{o}(\Delta t) \quad \text{as} \quad \Delta t \to 0\,, \tag{7.65}$$

where $\sigma(\mathbf{x}, t) = [\sigma_{i.j}(\mathbf{x}, t)]_{n_x \times n_x} > 0$, i.e., positive definite, and $\phi_{X(t)}(\mathbf{x}, t; \mathbf{x}_0, x_0) d\mathbf{y})$ is the transition probability density for $\mathbf{X}(t)$. Alternatively, these two *infinitesimal moment conditions* can be written

$$\lim_{\Delta t \to 0} \frac{\text{E}[\Delta \mathbf{X}(t) | \mathbf{X}(t) = \mathbf{x}]}{\Delta t} = \boldsymbol{\mu}(\mathbf{x}, t)$$

and

$$\lim_{\Delta t \to 0} \frac{\text{Cov}[\Delta \mathbf{X}(t), \Delta \mathbf{X}^\top(t) | \mathbf{X}(t) = \mathbf{x}]}{\Delta t} = \sigma(\mathbf{x}, t)\,.$$

There are other technical conditions that are needed and the reader should consult references like Feller [84, Chapt. 10] or Karlin and Taylor [162, Chapt. 15] for the history and variations in these conditions. Another technical condition implies that higher order moments are negligible,

$$\lim_{\Delta t \to 0} \frac{\text{E}[|\Delta \mathbf{X}(t)|^m \mid \mathbf{X}(t) = \mathbf{x}]}{\Delta t} = 0\,, \tag{7.66}$$

for $m \geq 3$.

Remarks 7.13.

- *Note that since the focus is on diffusion, the mth central moment could be used here as in [84, 162], instead of the uncentered mth moment in (7.66), just as the 2nd moment could have been used in (7.65) instead of the covariance. For high moments, the central moment form may be easier to use since means of deviation are trivially zero.*

- *Karlin and Taylor [162] show that from the Chebyshev inequality (Chapter 1, Exercise 4),*

$$\frac{\text{Prob}[|\Delta \mathbf{X}(t)| > \delta \mid \mathbf{X}(t) = \mathbf{x}]}{\Delta t} \leq \frac{\text{E}[|\Delta \mathbf{X}(t)|^m \mid \mathbf{X}(t) = \mathbf{x}]}{\delta^m \Delta t}, \qquad (7.67)$$

that the high moment condition (7.66) for ***any*** $m \geq 3$ *can imply the continuity condition (7.63) for* $\delta > 0$*. Depending on the problem formulation, the high moment condition may be easier to demonstrate than estimating the tail of the probability distribution in the continuity condition.*

In terms of the general multi-dimensional jump-diffusion model (7.31), the corresponding infinitesimal parameters, in absence of the jump term ($h = 0$), are the infinitesimal vector mean

$$\boldsymbol{\mu}(\mathbf{x}, t) = \mathbf{f}(\mathbf{x}, t)$$

and the infinitesimal matrix covariance

$$\sigma(\mathbf{x}, t) = (gg^\top)(\mathbf{x}, t)\,.$$

These infinitesimal properties by themselves do not make a diffusion process, since adding jump processes to diffusion process invalidates the continuity condition (7.63). For instance, examining this continuity condition for the simplest case of a simple Poisson process $X(t) = P(t)$ but with a time-dependent jump rate $\lambda(t) > 0$, yields

$$\frac{\text{Prob}[|\Delta P(t)| > \delta \mid P(t) = j]}{\Delta t} = \sum_{k=1}^{\infty} e^{-\Delta\Lambda(t)} \frac{(\Delta\Lambda)^k(t)}{k!\Delta t} = \frac{1 - e^{-\Delta\Lambda(t)}}{\Delta t}$$

assuming for continuity's sake that $0 < \delta < 1$ and where

$$\Delta\Lambda(t) = \int_t^{t+\Delta t} \lambda(s)ds \to \lambda(t)\Delta t \;\text{ as }\; \Delta t \to 0^+\,.$$

Thus,

$$\lim_{\Delta t \to 0} \frac{\text{Prob}[|\Delta P(t)| > \delta \mid P(t) = j]}{\Delta t} = \lambda(t) > 0$$

invalidating the continuity condition as expected, although the two basic infinitesimal moments can be calculated. In general, the higher moment criterion (7.66) will not be valid either, since for example,

$$\begin{aligned}
\lim_{\Delta t \to 0} \frac{\text{E}[|\Delta P(t)|^3 \mid \mathbf{X}(t) = \mathbf{x}]}{\Delta t} &= \lim_{\Delta t \to 0} \sum_{k=1}^{\infty} e^{-\Delta\Lambda(t)} \frac{(\Delta\Lambda)^k(t)}{k!\Delta t} k^3 \\
&= \lim_{\Delta t \to 0} \frac{\Delta\Lambda(t)(1 + 3\Delta\Lambda(t) + (\Delta\Lambda)^2(t))}{\Delta t} \\
&= \lambda(t) > 0\,,
\end{aligned}$$

where incremental moment Table 1.2 has been used. It is easy to guess that the number of infinitesimal moments of the Poisson process will be infinite, extrapolating from Table 1.2, unlike the limit of two infinitesimal moments for diffusion processes. However, the table only can be used to confirm that cases $m = 3:5$ yield the infinitesimal expectation of $\lambda(t)$.

So far these conditions are merely general formulations of diffusion processes for which similar properties have been derived in the earlier chapters of this book. Where their power lies is when they are used to approximate other stochastic processes, such as in the stochastic tumor application using a diffusion approximation that can be solved for tumor doubling times in Subsection 11.2.1.

7.9 Exercises

1. *Derivation of the Forward Kolmogorov Equation in the Generalized Sense.* Let the jump-diffusion process $X(t)$ satisfy the SDE,

$$dX(t) = f(X(t),t)dt + g(X(t),t)dW(t) + h(X(t),t,Q)dP(t;Q,X(t),t)) \quad (7.68)$$

$X(t_0) = x_0$, where the coefficient functions (f, g, h) are sufficiently well-behaved, Q is the jump-amplitude random mark with density $\phi_Q(q; X(t), t)$ and $\mathrm{E}[dP(t;Q,X(t),t)|X(t) = x] = \lambda(t;Q,x,t)dt$.

   (a) Show (easy) that, in the generalize sense,

$$\phi(x,t) \stackrel{\text{gen}}{=} \mathrm{E}[\delta(X(t) - x)|X(t_0) = x_0], \quad t_0 < t,$$

   where $\phi(x,t) = \phi_{X(t)}(x,t;x_0,t_0)$ is the transition probability density for the process $X(t)$ conditioned on the starting at $X(t_0) = x_0$ and $\delta(x)$ is the Dirac delta function.

   (b) Show that the Dirac delta function with composite argument satisfies

$$\int_{-\infty}^{+\infty} F(y)\delta(\gamma(y) - x)dy \stackrel{\text{gen}}{=} F\left(\gamma^{-1}(x)\right)\left|(\gamma^{-1})'(x)\right|,$$

   where $\gamma(y)$ is a monotonic function with non-vanishing derivative and inverse $y = \gamma^{-1}(z)$, such that $(\gamma^{-1})'(z) = 1/\gamma'(y)$ and $|\gamma^{-1}(\pm\infty)| = \infty$.

   (c) Apply the previous two results and other delta function properties from Section B.12 to derive the forward Kolmogorov equation (7.26) in the generalized sense.
   Hint: Regarding the proof of (7.26), the diffusion part is much easier given the delta function properties for the derivation, but the jump part is similar and is facilitated by the fact that $\gamma(y) = y + h(y;t,q)$ for fixed (t,q).

2. *Derivation of the **Feynman-Kac** (Dynkin with Integrating Factor) Formula for Jump-Diffusions.*
   Consider the jump-diffusion process,

$$dX(t) = f(X(t),t)dt + g(X(t),t)dW(t) + h(X(t),t,Q)dP(t;Q,X(t),t),$$

$X(t_0) = x_0 \in \Omega$, $t_0 < t < t_f$ and related backward Feynman-Kac (pronounced Fineman-Katz) final value problem,

$$\frac{\partial v}{\partial t_0}(x_0, t_0) + \mathcal{B}[v](x_0, t_0) + \theta(x_0, t_0)v(x_0, t_0) = \alpha(x_0, t_0)\,, \tag{7.69}$$

$x_0 \in \Omega$, $0 \le t_0 < t_f$, with final condition

$$v(x_0, t_f) = \gamma(x_0, t_f)\,, \quad x_0 \in \Omega\,, \quad 0 \le t_0 < t_f\,,$$

where $\mathcal{B}[v](x_0, t_0)$ is the backward operator corresponding to the jump-diffusion process (7.3). The given coefficients, $\theta(x_0, t_0)$, $\alpha(x, t)$ and $\gamma(x, t)$ are bounded and continuous. The solution $v(x_0, t_0)$ is assumed to be twice continuously differentiable in x_0 while once in t.

(a) In preparation, apply the stochastic chain rule to the auxiliary function

$$w(X(t), t) = v(X(t), t)\exp(\Theta(t_0, t))$$

to use an integrating factor technique to remove the non-Dynkin linear source term $\theta(x_0, t_0)v(x_0, t_0)$ from (7.69) with integrating factor exponent process

$$\Theta(t_0, t) = \int_{t_0}^{t} \theta(X(s), s)ds.$$

Then show (best done using the usual time-increment form of the stochastic chain rule) that

$$\begin{aligned} dw(X(t), t) \stackrel{\text{dt}}{=} e^{\Theta(t_0,t)}\Bigg(&\left(\frac{\partial v}{\partial t} + \mathcal{B}[v] + \theta v\right)(X(t), t)dt \\ &+ (gv\frac{\partial v}{\partial x})(X(t), t)dW(t) \\ &+ \int_{\mathcal{Q}} \delta_h[v](X(t), t, q)\widehat{\mathcal{P}}(\mathbf{dt}, \mathbf{dq}; X(t), t)\Bigg)\,, \end{aligned} \tag{7.70}$$

where $\delta_h[v]$ is defined in (7.4) and $\widehat{\mathcal{P}}$ is defined in (7.8).

(b) Next integrate the SDE (7.70) on $[t_0, t_f]$, solve for $v(x_0, t_0)$, then take expectations and finally apply the final value problem to obtain the Feynman-Kac formula corresponding to (7.69),

$$\begin{aligned} v(x_0, t_0) = \mathrm{E}\Bigg[&e^{+\Theta(t_0,t_f)}\gamma(X(t_f), t_f) \\ &- \int_{t_0}^{t_f} e^{+\Theta(t_0,s)}\alpha(X(s), s)ds \,\Bigg|\, X(t_0) = x_0\Bigg]\,. \end{aligned} \tag{7.71}$$

Hint: Follow the procedure in the derivation proof of Theorem 7.3 for this Feynman-Kac formula. See Schuss [244] or Yong and Zhou [288] for pure diffusion processes.

3. *Moments of Stochastic Dynamical Systems.* Consider first the linear stochastic dynamical system,

$$dX(t) = \mu_0 X(t)dt + \sigma_0 X(t)dW(t) + \nu_0 X(t)h(Q)dP(t;Q), \quad X(t_0) = x_0 \ ,$$

where $\{\mu_0, \sigma_0, \nu_0\}$ is a set of constant coefficients, x_0 is specified and $h(q)$ has finite moments with respect to a Poisson mark amplitude density $\phi_Z(z)$. Starting with a Dynkin's Formula (*or the Forward Kolmogorov Equation if you like deriving results the hard way*),

(a) Show that the conditional first moment of the process

$$\overline{X}(t) = \mathrm{E}[X(t)|X(t_0) = x_0]$$

satisfies a first order ODE in $\overline{X}(t)$ only, (x_0, t_0) fixed, corresponding to the mean (quasi-deterministic) analog of the SDE. Solve the ODE in terms of the given initial conditions.

(b) Derive the ODE for second moment

$$\overline{X^2}(t) = \mathrm{E}[X^2(t)|X(t_0) = x_0]$$

for the more general SDE

$$dX(t) = f(X(t))dt + g(X(t))dW(t) + h(X(t), q)dP(t;Q) \ ,$$

$X(t_0) = x_0$, in terms of expected coefficient values over both state and mark spaces.

(c) Use the general second moment ODE of part (b) to derive the corresponding ODE for the state variance

$$\mathrm{Var}[X(t)] = \overline{X^2}(t) - (\overline{X})^2(t)$$

for the linear dynamical system in the part (a). Your result should show that the ODE is linear in $\mathrm{Var}[X](t)$ with an inhomogeneous term depending on the $\overline{X}(t)$ first moment solution and constants, so the ODE is closed in that it is independent of any higher moments beyond the second. Solve the ODE.

Suggested References for Further Reading

- Arnold, 1974 [13].
- Bharucha-Reid, 1960 [31].
- Feller, 1971 [84, II].
- Gihman and Skorohod, 1972 [94].
- Goel and Richter-Dyn, 1974 [98].

- Hanson and Tier, 1982 [117].
- Jazwinski, 1970 [154].
- Karlin and Taylor, 1981 [162, II].
- Kushner and Dupuis, 2001 [179].
- Ludwig, 1975 [188].
- Øksendal, 1998 [222].
- Schuss, 1980 [244].

Chapter 8

Computational Stochastic Control Methods

God does not care about our mathematical difficulties.
He integrates empirically.
—Albert Einstein (1879-1955).

An idea which can be used once is a trick.
If it can be used more than once it becomes a method.
—George Polya and Gabor Szego.

"That's when I realized that research was my true calling, not software," he says. Developing software so other people could answer the big questions wasn't for him. He wanted to get back to answering them himself.
—Ajay Royyuruat , IBM Genographer, *Dream Jobs, IEEE Spectrum, vol. 43, no. 2, February 2006, pp. 40-41.*

The use of stochastic models, on the other hand, can result in gigantic increases in the complexity of data volume, storage, manipulation, and retrieval requirements.
—Simulation-Based Engineering Science, Report of the National Science Foundation Blue Ribbon Panel on Simulation-Based Engineering Science, *J. T. Oden, Chair, February 2006, 85 pages.*

Stochastic dynamic programming is not easy since the **PDE of stochastic dynamic programming** or the **Hamilton-Jacobi equation** given in (6.14-6.17) of Chapter 6 is not a standard PDE (partial differential equation). In fact, it is a functional PDE with just diffusion owing to the presence of a maximum with respect to the control. Also, for the more general jump-diffusion, the additional jump integrals make the PDE of stochastic dynamic programming a functional

partial integral differential equation or functional PIDE (partial integral differential equation). The analytic complexity of this functional PIDE means that for the usual finite difference or finite element methods, numerical convergence conditions are unknown or not easily ascertainable.

This chapter discusses PDE-oriented finite difference methods developed by the author and coworkers [106, 107, 108, 277, 110] for solving the PDE of **stochastic dynamic programming (SDP)** (6.14-6.17), with special emphasis on techniques and convergence conditions. The numerical foundations and complexity of computational stochastic control are discussed in [110].

An alternative method relies on using Markov chain probabilities to construct convergent finite difference approximations that are rigorously convergent in the weak sense and is called the **Markov chain approximation (MCA)** developed by Kushner and coworkers [174, 175, 179].

Some methods use a canonical model formulation whose solution algorithm results in significant reduction in the dimensional complexity, e.g., the **linear-quadratic (LQ) model** for the optimal control of jump-diffusions (LQJD or LQGP) [274] and the **constant relative risk aversion (CRRA) utility model** for the optimal portfolios in finance [122, 123, 129, 291]. In addition, special integration methods for jump integrals and a least squares approximation for forming simpler LQJD problems are also discussed [277]. The LQJD canonical model dimensional reduction algorithm is covered in Section 6.4 on page 182 in Chapter 6 while the deterministic LQ and variants are covered in Section A.3 on page A23 in Chapter A.

Another canonical model dimensional reduction algorithm is treated in Sections 10.4 on page 326 and 10.5 on page 337 in Chapter 10 for two different optimal portfolio and consumption applications.

For a more historical introduction to computational methods in control, see Larson [182], Polak [227] and Dyer and McReynolds [76].

8.1 Finite Difference PDE Methods of SDP

A decade ago, the author contributed an invited chapter on *Computational Stochastic Dynamic Programming* [108] in a *Control and Dynamic Systems* volume discussing the use of finite difference methods of solution. This section is based on his past experience with large scale stochastic control applications using many of the largest vector and parallel computers available academically from national centers such as Argonne National Laboratory, Los Alamos National Laboratory, National Center for Supercomputing Applications, San Diego Supercomputing Center and the Pittsburgh Supercomputing Center. An updated version of the techniques involved is given but simplified to one state dimension initially for convenience.

Consider the jump-diffusion SDE for state $X(t)$ and control $U(t)$,

$$\begin{aligned} dX(t) \stackrel{\text{sym}}{=} \quad & f(X(t),U(t),t)dt + g(X(t),U(t),t)dW(t) \\ & +h(X(t),U(t),t,Q)dP(t;Q,X(t),U(t),t) \ , \end{aligned} \tag{8.1}$$

where $dP(t;Q,X(t),U(t),t)$ and $dW(t)$ are the stochastic differentials of the jump-

diffusion process including the compound Poisson mark Q with jump-rate $\lambda(t;x,u,t)$. The SDE coefficients, $(f(x,u,t), g(x,u,t), h(x,u,t,q))$, are assumed to be bounded or at least integrable in their arguments, so as not to over-restrict the problem. Let the objective be the minimum of the expected cumulative running costs $C(x,u,t)$ and terminal cost $S(x_f,t_f)$,

$$\begin{aligned} v^*(x,t) \equiv \quad \min_{U[t,t_f)} \Bigg[\underset{(dW,dP)[t,tf)}{\mathrm{E}} \Big[\textstyle\int_t^{t_f} C(X(s),U(s),s)ds + S(X(t_f),t_f) \\ \Big| X(t)=x, U(t)=u \Big] \Bigg] , \end{aligned} \tag{8.2}$$

for $t_0 \le t < t_f$.

The application of **Bellman's Principle of Optimality** and the stochastic chain rule along with the infinitesimal moments $\mathrm{E}[dW(t)]=0$, $\mathrm{Var}[dW(t)]=dt$ and $\mathrm{E}[dP(t;Q,X(t),U(t),t)|X(t)=x,U(t)=u] = \lambda(t;x,u,t)dt$ leads to the stochastic dynamic programming PIDE using only order dt terms,

$$\begin{aligned} 0 = \quad & v_t^*(x,t) + \min_u \left[\mathcal{H}(x,u,t) \right] \\ \equiv \quad & v_t^*(x,t) + \min_u \Big[C(x,u,t) + f(x,u,t) v_x^*(x,t) + \tfrac{1}{2} g^2(x,u,t) v_{xx}^*(x,t) \\ & + \lambda(t;x,u,t) \textstyle\int_{\mathcal{Q}} \left(v^*(x+h(x,u,t,q),t) - v^*(x,t) \right) \phi_Q(q;x,u,t) \Big] \\ = \quad & v_t^*(x,t) + \mathcal{H}^*(x,t) \; . \end{aligned} \tag{8.3}$$

If the regular or unconstrained optimal control exists and is unique, then

$$u^{(\mathrm{reg})}(x,t) = \quad \underset{u}{\mathrm{argmin}} \left[\mathcal{H}(x,u,t) \right] , \tag{8.4}$$

but, in general, the optimal control, $u^*(x,t)$, is subject to any control constraints. The final condition from the minimal conditional expected cost objective (8.2) is

$$v^*(x,t) = S(x,t_f). \tag{8.5}$$

However, the boundary conditions in general are model and domain dependent.

8.1.1 Linear Control Dynamics and Quadratic Control Costs

In order, to keep the focus on basic computations, it will be assumed that the drift of the state dynamics is linear in the control and that the running costs are quadratic in the control, i.e, the **LQJD problem in control only (LQJD/U)** discussed in Subsection 6.4.1. These assumptions are more general than the LQJD problem, but are sufficient to determine optimal control clearly in terms of (x,t). Hence, let

$$\begin{aligned} f(x,u,t) &= f_0(x,t) + f_1(x,t)u, \\ g(x,u,t) &= g_0(x,t), \quad h(x,u,t,q) = h_0(x,t,q), \\ \lambda(t;x,u,t) &= \lambda_0(t;x,t), \quad \phi_Q(q;x,u,t) = \phi_Q(q), \\ C(x,u,t) &= c_0(x,t) + c_1(t;x,t)u + c_2(x,t)u^2, \\ \mathcal{H}(x,u,t) &= \mathcal{H}_0(x,t) + \mathcal{H}_1(x,t)u + \tfrac{1}{2}\mathcal{H}_2(x,t)u^2 \; . \end{aligned} \tag{8.6}$$

Thus, the PDE of stochastic dynamic programming in Hamilton-Jacobi form using (6.19) with the current assumptions,

$$\begin{aligned} 0 = \quad & v_t^*(x,t) + \mathcal{H}^*(x,t) \\ = \quad & v_t^*(x,t) + C_0(x,t) + C_1(x,t)u^* + \tfrac{1}{2}C_2(x,t)(u^*)^2 \\ & +(f_0(x,t) + f_1(x,t)u^*)v_x^*(x,t) + \tfrac{1}{2}g_0^2(x,t)v_{xx}^*(x,t) \\ & +\lambda_0(t;x,t)\int_{\mathcal{Q}}\left(v^*(x+h_0(x,t,q),t) - v^*(x,t)\right)\phi_Q(q)dq \ , \end{aligned} \tag{8.7}$$

and the regular control is from (6.31) after simplifications for the current one state dimension form,

$$u^{(\mathrm{reg})}(x,t) = \ -\left(c_1(x,t) + f_1(x,t)v_x^*(x,t)\right)/c_2(x,t) \ , \tag{8.8}$$

provided $c_2(x,t) > 0$, i.e., positive definite, for a minimum. Since real problems have contraints, let $U^{(\min)} \leq u(x,t) \leq U^{(\max)}$. Then the optimal control law can be written

$$\begin{aligned} u^*(x,t) = \quad & \min(U^{(\max)}, \max(U^{(\min)}, u^{(\mathrm{reg})}(x,t))) \\ = \quad & \left\{ \begin{array}{ll} U^{(\min)}, & u^{(\mathrm{reg})}(x,t) \leq U^{(\min)} \\ u^{(\mathrm{reg})}(x,t), & U^{(\min)} \leq u^{(\mathrm{reg})}(x,t) \leq U^{(\max)} \\ U^{(\max)}, & U^{(\max)} \leq u^{(\mathrm{reg})}(x,t) \end{array} \right\} . \end{aligned} \tag{8.9}$$

For multidimensional state space problems see the stochastic dynamic programming Chapter 6 here or Hanson's computational stochastic dynamic programming chapter in [108].

8.1.2 Crank-Nicolson, Extrapolation-Predictor-Corrector Finite Difference Algorithm for SDP

The numerical algorithm used here is basically a modification of the work of Douglas and Dupont [72, 73] on nonlinear parabolic equations modified for stochastic dynamic programming and the PIDE for jump-diffusions.

First the problem is discretized in backward time since stochastic dynamic programming is a backward problem but the state space is discretized in a regular grid, with N_t nodes in t on $[t_0, t_f]$ and N_x nodes in x on $[x_0, x_{\max}]$,

$$\begin{aligned} & t \rightarrow \ T_k = t_f - (k-1)\cdot\Delta t, \ \text{for } k = 1{:}N_t, \ \Delta t = (t_f - t_0)/(N_t - 1) \ , \\ & x \rightarrow \ X_j = x_0 + (j-1)\cdot\Delta X, \ \text{for } j = 1{:}N_x, \ \Delta X = (x_{\max} - x_0)/(N_x - 1) \ . \end{aligned} \tag{8.10}$$

This grid leads to a corresponding discretization of the dependent variables that follow using a **second order central finite difference (CFD)** for the time derivative, evaluating at the mid-time point, and second order CFDs for the state derivatives when $j = 1{:}N_x$ for each $k = 1{:}N_t$ corresponding to the backward time count

with $T_1 = t_f$:

$$\begin{aligned}
v^*(X_j, T_k) &\to V_{j,k} \ , \\
v_t^*(X_j, T_{k+0.5}) &\to (V_{j,k+1} - V_{j,k})/(-\Delta t) \ , \\
v_x^*(X_j, T_k) &\to \mathrm{DV}_{j,k} = 0.5(V_{j+1,k} - V_{j-1,k})/\Delta X \ , \\
v_{xx}^*(X_j, T_k) &\to \mathrm{DDV}_{j,k} = (V_{j+1,k} - 2V_{j,k} + V_{j-1,k})/(\Delta X)^2 \ , \\
u^{(\mathrm{reg})}(X_j, T_k) &\to \mathrm{UR}_{j,k} = -\left(C_{1,j,k} + F_{1,j,k}\mathrm{DV}_{j,k}\right)/C_{2,j,k} \ , \\
u^*(X_j, T_k) &\to \mathrm{US}_{j,k} = \min(\mathrm{UMAX}, \max(\mathrm{UMIN}, \mathrm{UR}_{j,k})) \ , \\
v^*(X_j + h_0(X_j, T_k, q), T_k) &\to \mathrm{VH}_{j,k}(q),
\end{aligned} \tag{8.11}$$

where $F_{i,j,k} = f_i(X_j, T_k)$ for $i = 0:1$, $C_{i,j,k} = c_i(X_j, T_k)$ for $i = 0:2$, $\mathrm{UMIN} = U^{(\min)}$ and $\mathrm{UMAX} = U^{(\max)}$.

The **Crank-Nicolson Implicit (CNI)** method provides central differencing in state and time, so is second order accurate in both independent variables, i.e., $\mathrm{O}^2(\Delta X) + \mathrm{O}^2(\Delta t)$, and the implicitness provides stability over all positive steps in time, Δt. However, for general problems, such as those that are multi-dimensional or are nonlinear, the implicit and tridiagonal properties are no longer valid, unless CNI can be extended by **alternating directions implicit (ADI)** through known splittings of the spatial operators. However, for nonlinear problems, recalling from Chapter 6 that the PDE of stochastic dynamic programming is nonlinear, ADI is not useful and predictor-corrector methods can be used to preserve the second order accuracy in several dimensions and for nonlinear problems. For these more general applications, the basic structure of the CNI method upon dissection consists of a midpoint integral approximation and an averaging to convert the time-midpoint to integral grid point values. Thus, symbolically using the PDE of stochastic dynamic programming in Hamilton-Jacobi form, $0 = v_t^*(x,t) + \mathcal{H}^*(x,t)$, using (8.7), the **midpoint rule approximation** is then

$$\begin{aligned}
V_{j,k+1} - V_{j,k} = &\ \int_{T_k}^{T_k - \Delta t} v_t^*(X_j, t)dt = -\int_{T_k}^{T_k - \Delta t} \mathcal{H}(X_j, t)dt \\
\simeq &\ +\Delta t \cdot \mathcal{H}(X_j, T_{k+0.5}) = +\Delta t \cdot \mathcal{H}_{j,k+0.5} \ ,
\end{aligned} \tag{8.12}$$

which is finally followed by a **second order accuracy preserving averaging step**,

$$V_{j,k+1} \simeq V_{j,k} + 0.5 \cdot \Delta t \cdot (\mathcal{H}_{j,k} + \mathcal{H}_{j,k+1}) \ , \tag{8.13}$$

where the midpoint (mid-time) value of the objective has been replaced by targeted values at given time nodes. While this last step may look like a linear assumption, in most cases this can be extended by quasi-linearization, e.g., the average for a power can be approximated by $(V_{j,k+0.5})^{n+1} \simeq 0.5(V_{j,k})^n(V_{j,k} + V_{j,k+1})$ in the zeroth correction with further refinement in subsequent corrections, always keeping the newest update of $V_{j,k+1}$ as a linear term. The reader can show that under second order differentiability the averaging step is second order accurate in time ($\mathrm{O}^2(\Delta t)$) at the midpoint, it being well-known that the midpoint rule used here is second order accurate in time. It is the midpoint rule evaluation that makes the seemingly

first order approximation for $v_t^*(x,t)$ in (8.11) accurate to $\mathrm{O}^2(\Delta t)$ rather than to $\mathrm{O}(\Delta t)$.

Integration and Interpolation for Jump Integrals

Another modification is needed for handling the jump integrals. One procedure is the use of **Gauss-statistics rules** introduced by Westman and Hanson in [277] as a generalization of the Gaussian quadrature rules, but customized for the given mark density $\phi_Q(q)$ in the application. These rules use N_q points Q_i and N_q weights w_i and have a polynomial precision of degree $n_q = N_q - 1$. The weights and nodes satisfy the $2 \cdot N_q$ nonlinear equations,

$$\sum_{i=1}^{N_q} w_i \cdot Q_i^j = \mathrm{E}_Q[Q^j] = \int_{\mathcal{Q}} q^j \phi_Q(q) dq \ , \tag{8.14}$$

for $j = 0 : 2N_q - 1$. This leads to the Gauss-statistics approximation for the jump integral:

$$\begin{aligned} \mathrm{IVH}_{j,k} \equiv & \ \int_{\mathcal{Q}} \mathrm{VH}_{j,k}(q)\phi_Q(q)dq \simeq \textstyle\sum_{i=1}^{N_q} w_i \mathrm{VH}_{j,k}(Q_i) \\ = & \ \textstyle\sum_{i=1}^{N_q} w_i v^*(X_j + h_0(X_j, T_k, Q_i), T_k) \ . \end{aligned} \tag{8.15}$$

In general, the $\mathrm{VH}_{j,k}(Q_i)$ will be implicit values that are not necessarily at specified state nodes $j' = 1 : N_t$ in $V_{j',k}$. Just as in Crank-Nicolson averaging, $\mathrm{O}^2(\Delta X)$ interpolation is needed relative to the nearest neighbor state nodes. Let the ith state argument be

$$X_j + h_0(X_j, T_k, Q_i) = X_{j+\ell_i} + \epsilon_i \Delta X \ ,$$

where the floor integer is

$$\ell_i = \ell_{i,j,k} = \lfloor h_0(X_j, T_k, Q_i)/\Delta X \rfloor$$

and fraction

$$\epsilon_i = \epsilon_{i,j,k} = h_0(X_j, T_k, Q_i)/\Delta X - \ell_i.$$

Thus, the $\mathrm{O}^2(\Delta X)$ interpolation is

$$\mathrm{VH}_{j,k}(Q_i) \simeq (1-\epsilon_i) \cdot V_{j+\ell_i,k} + \epsilon_i \cdot V_{j+\ell_i+1,k} \ , \tag{8.16}$$

assuming the jumps are not out of range of the state space or are handled by proper boundary conditions. Thus,

$$\mathrm{IVH}_{j,k} \simeq \sum_{i=1}^{N_q} w_i \left((1-\epsilon_i) \cdot V_{j+\ell_i,k} + \epsilon_i \cdot V_{j+\ell_i+1,k} \right) \ . \tag{8.17}$$

Example 8.1. ***Gauss-Statistics Quadrature for Log-Uniform Jump-Amplitudes:***

For example, in the case that $\phi_Q(q)$ is the density of the uniform distribution on $[a,b]$, then

$$\text{for } N_q = 1,\ n_q = 1,\quad w_1 = 1,\quad Q_1 = 0.5(a+b)\ ;$$

or

$$\text{for } N_q = 2,\ n_q = 3,\quad w_1 = 0.5,\ w_2 = 0.5,$$
$$Q_1 = 0.5(a+b) - 0.5(b-a)/\sqrt{3},\ Q_2 = 0.5(a+b) + 0.5(b-a)/\sqrt{3}\ .$$

For higher precision on finite mark domains $[a,b]$, piecewise applications of these rules can be made on subdivisions $[q_i, q_{i+1}]$ where $q_i = a + (i-1)\Delta q$ for $i = 1\!:\!M_q$ nodes with $\Delta q = (b-a)/(M_q - 1)$. See Westman and Hanson [277] for more information.

In the case that there is a special q-dependence of the jump-amplitude coefficient $h_0(x,t,q)$ for which the moments can be easily or conveniently calculated, then it may be possible to use just the interpolation of $\mathrm{VH}_{j,k}(q)$ without Gauss-statistics quadrature in q.

Example 8.2. ***Geometric Jump-Diffusion with Log-Uniform Jump-Amplitudes Jump-Integral Approximation:***
In the financial geometric jump-diffusion with log-uniform jump-amplitude distribution (10.119), the distribution of q is uniform with respect to the log-return $\ln(x)$, but in the original return values the jump in the return is $h(x,t,q) = x \cdot (e^q - 1)$ by Itô's chain rule. For the financial market q is very small, then so is $e^q - 1$, while a is small and negative with b small and positive. Provided $|\epsilon| \le 1$ where $\epsilon = X_j(e^q - 1)/\Delta X$, then the appropriate piece-wise linear interpolation using the explicit node set $\{V_{j-1,k}, V_{j,k}, V_{j+1,k}\}$ is

$$\mathrm{VH}_{j,k}(q) \simeq \left\{ \begin{array}{ll} (1-\epsilon)V_{j,k} + \epsilon V_{j+1,k}, & q \ge 0, \epsilon \ge 0 \\ -\epsilon V_{j-1,k} + (1+\epsilon)V_{j,k}, & q \le 0, \epsilon \le 0 \end{array} \right\}\ . \tag{8.18}$$

Since the factor $(e^q - 1)$ is now explicit, it can be integrated directly without Gaussian quadrature to produce,

$$\begin{array}{c} \int_a^b \mathrm{VH}_{j,k}(q)\phi_Q(q)dq \simeq V_{j,k} + \frac{X_j}{\Delta X}(V_{j,k} - V_{j-1,k})\frac{1+a-e^a}{b-a} \\ + \frac{X_j}{\Delta X}(V_{j+1,k} - V_{j-,k})\frac{e^b - 1 - b}{b-a}\ . \end{array} \tag{8.19}$$

Extrapolation, Prediction and Correction

Summarizing the above CNI discretizations, the PIDE of stochastic dynamic programming of (8.7) can be put in the preliminary form

$$\begin{array}{rcl} V_{j,k+1} = & V_{j,k} + \Delta t \cdot \mathcal{H}_{j,k+0.5} & \\ = & V_{j,k} + \Delta t\,(C_{j,k+0.5} + F_{j,k+0.5} \cdot \mathrm{DV}_{j,k+0.5} & \\ & +0.5 \cdot G^2_{0,j,k+0.5} \cdot \mathrm{DDV}_{j,k+0.5} + \Lambda_k \cdot (\mathrm{IVH}_{j,k+0.5} - V_{j,k+0.5})\Big)\ , & \end{array} \tag{8.20}$$

where $C_{j,k} = C_{0,j,k} + C_{1,j,k}\mathrm{US}_{j,k} + 0.5 \cdot C_{2,j,k} \cdot \mathrm{US}_{j,k}^2$, $F_{j,k} = F_{0,j,k} + F_{1,j,k}\mathrm{US}_{j,k}$, $G_{0,j,k} = g_0(X_j, T_k)$, $\Lambda_k = \lambda_0(T_k)$, $\mathrm{US}_{j,k} = \min(\mathrm{UMAX}, \max(\mathrm{UMIN}, \mathrm{UR}_{j,k}))$ and $\mathrm{UR}_{j,k} = -\left(C_{1,j,k} + F_{1,j,k} \cdot \mathrm{DV}_{j,k}\right)/C_{2,j,k}$, using (8.11).

Once there are two prior values $V_{j,k-1}$ and $V_{j,k}$ which happens when $k \geq 2$, linear extrapolation (ex) can be used to accelerate the SDP corrections. The first step from the final condition at $k = 1$ to $k = 2$ takes the most corrections since no trend is available, only $V_{j,1}$. Otherwise the extrapolation (ex) step for the time-midpoint is used for $k \geq 2$ rather than the initial prediction at $k = 1$,

$$V_{j,k+0.5}^{(ex)} = \left\{ \begin{array}{ll} V_{j,k}, & k = 1 \\ 0.5(3V_{j,k} - V_{j,k-1}), & k \geq 2 \end{array} \right\} , \tag{8.21}$$

which is used to update the derivative $\mathrm{DV}_{j,k+0.5}$, 2nd derivative $\mathrm{DDV}_{j,k+0.5}$, regular control $\mathrm{UR}_{j,k+0.5}$, optimal control $\mathrm{UR}_{j,k+0.5}$ and jump functions $\mathrm{VH}_{j,k+0.5}(q)$ in the list (8.11) for the pseudo-Hamiltonian $\Delta t \cdot \mathcal{H}_{j,k+0.5}^{(ex)}$ in (8.12, 8.20) using quasi-linearization for nonlinear terms. The resulting update of the value is called the predictor or 1st correction step $(c, 1)$,

$$V_{j,k+1}^{(c,1)} = V_{j,k} + \Delta t \cdot \mathcal{H}_{j,k+0.5}^{(ex)} , \tag{8.22}$$

for all j, as long as $k \geq 2$. Otherwise the predicted step uses the current value or $V_{j,k+1}^{(c,1)} = V_{j,k} + \Delta t \cdot \mathcal{H}_{j,k}$ using (8.20). The evaluation step uses the updated average,

$$V_{j,k+0.5}^{(c,1)} = 0.5(V_{j,k+1}^{(c,1)} + V_{j,k}) , \tag{8.23}$$

which is used to update all the needed values in (8.11) and finally in all the next correction $(c, 2)$,

$$V_{j,k+1}^{(c,2)} = V_{j,k} + \Delta t \cdot \mathcal{H}_{j,k+0.5}^{(c,1)} . \tag{8.24}$$

The γth correction loop given $V_{j,k+1}^{(c,\gamma)}$ will contain

$$V_{j,k+0.5}^{(c,\gamma)} = 0.5(V_{j,k+1}^{(c,\gamma)} + V_{j,k}) , \tag{8.25}$$

plus the corresponding evaluations of $\mathrm{DV}_{j,k+0.5}^{(c,\gamma)}$, $\mathrm{DDV}_{j,k+0.5}^{(c,\gamma)}$, $\mathrm{UR}_{j,k+0.5}^{(c,\gamma)}$, $\mathrm{UR}_{j,k+0.5}^{(c,\gamma)}$, $\mathrm{VH}_{j,k+0.5}^{(c,\gamma)}(q)$ including integration, and $\mathcal{H}_{j,k+0.5}^{(c,\gamma)}$. Then

$$V_{j,k+1}^{(c,\gamma+1)} = V_{j,k} + \Delta t \cdot \mathcal{H}_{j,k+0.5}^{(c,\gamma)} . \tag{8.26}$$

The corrections continue until the stopping criterion is reached, for instance, the relative criteria given tolerance tol_v,

$$\left\| V_{j,k+1}^{(c,\gamma+1)} - V_{j,k+1}^{(c,\gamma)} \right\|_1 < \mathrm{tol}_v \left\| V_{j,k+1}^{(c,\gamma)} \right\|_1 , \tag{8.27}$$

for each k, continuing corrections if not satisfied, otherwise stopping the corrections setting $\gamma_{\max} = \gamma + 1$ and setting the final $(k+1)$st value at

$$V_{j,k+1} = V_{j,k+1}^{(c,\gamma_{\max})} . \tag{8.28}$$

In (8.27), $\| * \|_1$ denotes the one-norm with respect to the state index j for current time index k, but other norms could be used with the one-norm being less computationally costly.

Stability criteria is another matter due to the complexity of the PIDE of SDP in terms of multi-state systems, jump integrals, nonlinear terms and optimization terms. A rough criterion focuses on the diffusion term $G^2_{0,j,k+0.5}\mathrm{DDV}_{j,k+0.5}$ in (8.20), which can be expanded by substituting the CFD form (8.11) for $\mathrm{DV}_{j,k+0.5}$ and $\mathrm{DDV}_{j,k+0.5}$ into (8.20) and produces

$$\begin{aligned} V_{j,k+1} = \quad & \left(1 - \frac{\Delta t}{\Delta X^2} G^2_{0,j,k+0.5}\right) V_{j,k+0.5} \\ & + 0.5 \frac{\Delta t}{\Delta X^2} \left(G^2_{0,j,k+0.5} + F_{j,k+0.5}\Delta X\right) V_{j+1,k+0.5} \\ & + 0.5 \frac{\Delta t}{\Delta X^2} \left(G^2_{0,j,k+0.5} - F_{j,k+0.5}\Delta X\right) V_{j-1,k+0.5} \\ & + \Delta t C_{j,k+0.5} + \Lambda_k \Delta t \cdot (\mathrm{IVH}_{j,k+0.5} - V_{j,k+0.5}) \,, \end{aligned} \tag{8.29}$$

where $C_{j,k} = C_{0,j,k} + C_{1,j,k}\mathrm{US}_{j,k} + 0.5 \cdot C_{2,j,k} \cdot \mathrm{US}^2_{j,k}$ and $F_{j,k} = F_{0,j,k} + F_{1,j,k}\mathrm{US}_{j,k}$.

Following Kushner and Dupuis [179] and ignoring the jump and cost terms, the positivity of the diffusion with drift terms leads to a **parabolic mesh ratio**

$$\max_{j,k} \left(G^2_{0,j,k+0.5}\right) \frac{\Delta t}{(\Delta X)^2} < 1, \tag{8.30}$$

or so, but certainly should be less than one. This assumes that the PIDE is **diffusion-dominated** and accounts for the drift as well as other terms in (8.3). The discrete HJB equation is said to be **diffusion-dominated**, modified for current form from a relation in [179], if

$$\min_{j,k} \left(G^2_{0,j,k} - |F_{j,k}|\,\Delta X\right) \geq 0, \tag{8.31}$$

where $F_{j,k} = F_{0,j,k} + F_{1,j,k}\mathrm{US}_{j,k}$, so that the coefficients of the non-diagonal terms, $V_{j+1,k+0.5}$ and $V_{j-1,k+0.5}$ are also positive. Otherwise the discrete problem is either mixed domination or **drift-dominated**, ignoring the jump cost terms. The technique is to decrease Δt and/or increase ΔX if spurious oscillations appear. Not that the diffusion-dominated condition (8.31) is satisfied for sufficiently small state stepsize ΔX as long as the diffusion coefficient $G^2_{0,j,k+0.5}$ is not also sufficiently small. For more information on linear and multi-state models, see Hanson [108], [216] and [111] or see Kushner and Dupuis [179].

The central finite differences for state derivatives work quite well in the diffusion-dominated regime, but are not useful for specified derivative boundary conditions, such as the convection boundary condition and the no-flux or reflecting boundary condition (7.43), e.g., $v^*_x(x_0, t) = 0$ on the left boundary or $v^*_x(x_{\max}, t) = 0$ on the right boundary, respectively, assuming the diffusion coefficient $g^2_0(x, t)/2 > 0$ for a well defined flux and nonsingular boundary condition. Using second order forward and backward finite differences, respectively, to maintain consistency in numerical accuracy with the central differences in the interior of $[x_0, x_{\max}]$, the derivatives at

the boundaries are

$$\begin{aligned} v_x^*(x_0, T_k) &\simeq \mathrm{DV}_{1,k} = -0.5(V_{3,k} - 4V_{2,k} + 3V_{1,k})/\Delta x, \\ v_x^*(x_{\max}, T_k) &\simeq \mathrm{DV}_{N_x,k} = +0.5(V_{N_x-2,k} - 4V_{N_x-1,k} + 3V_{N_x,k})/\Delta x. \end{aligned} \tag{8.32}$$

Now, these signs of these terms are not a problem for stability since these conditions are used as eliminants for $V_{1,k}$ for left boundary values and $V_{N_x,k}$ for right boundary values rather than a replacements for the discrete HJB equations (8.29). An alternate derivative boundary condition implementation is to add artificial boundary to the domain, but this author has found better performance using only the domain with the derivative boundary values like (8.32).

For **finite element** versions see Chung, Hanson and Xu [54] or Hanson[108]. Although not on SDP, the work of Chakrabarty and Hanson [49] uses the CNI-predictor-corrector methods discussed here with finite elements for a large scale distributed parameter or PDE-driven system. Finite element methods are better for presenting multidimensional systems and systems on irregular domains.

8.1.3 Upwinding Finite Differences If Not Diffusion-Dominated

When the diffusion-dominated condition (8.31) is no longer valid then the drift term becomes important or the system (8.3) becomes drift dominant and the coefficients of the non-diagonal terms, $V_{j+1,k}$ and $V_{j-1,k}$ are no longer guaranteed to be positive. In this case the system takes upon more hyperbolic PDE characteristics since the drift terms are of hyperbolic type as are first order PDEs. In the case of drift-dominance or near-drift-dominance, following Kushner [179] and others, the finite difference to the first state partial of the optimal value function $v_x^*(X_j, T_k)$ in (8.11) should be changed from second-order CFD to first-order **upwinded finite differences (UFD)** which uses forward or backward finite differences (FFDs or BFDs) to coincide with the sign of the drift coefficient, respectively, i.e.,

$$\mathrm{DV}_{j,k} = \left\{ \begin{array}{ll} (V_{j+1,k} - V_{j,k})/\Delta x, & F_{j,k} \geq 0 \\ (V_{j,k} - V_{j-1,k})/\Delta x, & F_{j,k} < 0 \end{array} \right\}, \tag{8.33}$$

where again $F_{j,k} = F_{0,j,k} + F_{1,j,k}\mathrm{US}_{j,k}$. Thus, upwind is in the direction of the drift. However, upwinding requires a sacrifice of numerical accuracy consistency, going from $\mathrm{O}(\Delta X^2)$ CFD to $\mathrm{O}(\Delta X)$ UFD for the first state partial, in favor of more stable numerical calculations. Substituting the UFD form (8.33) for $\mathrm{DV}_{j,k}$ in (8.20) produces

$$\begin{aligned} V_{j,k+1} = \quad & \left(1 - \frac{\Delta t}{\Delta X^2}\left(G_{0,j,k+0.5}^2 + 0.5|F_{j,k+0.5}|\Delta X\right)\right) V_{j,k+0.5} \\ & +0.5\frac{\Delta t}{\Delta X^2}\left(G_{0,j,k+0.5}^2 + [F_{j,k+0.5}]_+\Delta X\right) V_{j+1,k+0.5} \\ & +0.5\frac{\Delta t}{\Delta X^2}\left(G_{0,j,k+0.5}^2 + [F_{j,k+0.5}]_-\Delta X\right) V_{j-1,k+0.5} \\ & +\Delta t C_{j,k+0.5} + \Lambda_k \Delta t \cdot (\mathrm{IVH}_{j,k+0.5} - V_{j,k+0.5}) \,, \end{aligned} \tag{8.34}$$

where $[f]_{\pm} \equiv \max[\pm f] \geq 0$, such that $[f]_+ + [f]_- = |f|$ and $[f]_+ - [f]_- = f$. Hence, for the diffusion terms, all coefficients are positive provided the **drift-adjusted parabolic mesh ratio** condition is satisfied,

$$\max_{j,k} \left(G^2_{0,j,k+0.5} + 0.5|F_{j,k+0.5}| \right) \frac{\Delta t}{(\Delta X)^2} < 1, \tag{8.35}$$

without the extra diffusion-dominated condition in (8.31) being needed.

8.1.4 Multi-state Systems and Bellman's Curse of Dimensionality

Generalization to multi-dimensional state spaces can lead to very large scale computational problems, since the size of the computational problem grows with the number of dimensions multiplied by the number of nodes per dimension.

Starting with a version of the PDE of SDP in (6.19) modified for the LQJD/U form in (6.21-6.25) and no diffusion process correlations ($R' = I_{n_w \times n_w}$),

$$\begin{aligned} 0 = {} & v^*_t(\mathbf{x},t) + C_0(\mathbf{x},t) + \mathbf{C}_1^\top(\mathbf{x},t)\mathbf{u}^* + \frac{1}{2}(\mathbf{u}^*)^\top C_2(\mathbf{x},t)\mathbf{u}^* \\ & + \nabla_{\mathbf{x}}^\top [v^*](\mathbf{x},t) \cdot (\mathbf{f}_0(\mathbf{x},t) + f_1(\mathbf{x},t)\mathbf{u}^*) \\ & + \frac{1}{2}\left(g_0 g_0^\top\right)(\mathbf{x},t) : \nabla_{\mathbf{x}} \left[\nabla_{\mathbf{x}}^\top [v^*]\right](\mathbf{x},t) \\ & + \sum_{\ell=1}^{n_p} \lambda_\ell(t) \int_{\mathcal{Q}_\ell} \left(v^* \left(\mathbf{x} + \widehat{\boldsymbol{h}}_{0,\ell}(\mathbf{x},t,q_\ell), t \right) - v^*(\mathbf{x},t) \right) \phi_{Q_\ell}(q_\ell) dq_\ell \,, \end{aligned} \tag{8.36}$$

where the double-dot product (:) is defined as a trace in (5.99) and the ℓth jump-amplitude vector is $\widehat{\boldsymbol{h}}_{0,\ell}(\mathbf{x},t,q_\ell) \equiv [h_{0,i,\ell}(\mathbf{x},t,q_\ell)]_{n_x \times 1}$ for $\ell = 1 : n_p$.

Let the state dimension be n_x and realized state vector be given by $\mathbf{x} = [x_i]_{n_x \times 1}$. In discrete form, the state vector with a common N_x nodes per dimension becomes $\mathbf{x} = [x_i]_{n_x \times 1} \rightarrow \mathbf{X_j} = [X_{i,j_i}]_{n_x \times 1}$, representing a single point in state space, given one j_i for each state i from the range $j_i = 1 : N_x$ for $i = 1 : n_x$ with $X_{i,j_i} = x_{i,0} + (j_i - 1)\Delta X_i$ and $\Delta X_i = (x_{i,\max} - x_{i,0})/(N_x - 1)$. The entire set of points in state space can be represented by $\boldsymbol{\mathcal{X}} = [X_{i,j}]_{n_x \times N_x}$ with corresponding vector index $J = [J_{i,j}]_{nx \times N_x}$. This representation leads to a large scale expansion of the independent variables of SDP from that in (8.37) for each current $k = 1 : N_t$,

using CFD for each state component of state partial derivatives:

$$\begin{aligned}
v^*(\mathbf{X_j},T_k) \rightarrow\ & V_{J,k} \equiv [V_{j_1,j_2,\ldots,j_{n_x},k}]_{N_x\times N_x\times\cdots\times N_x},\\
v_t^*(\mathbf{X_j},T_k) \rightarrow\ & (V_{J,k+1}-V_{J,k})/(-\Delta t),\\
\nabla_x[v^*](\mathbf{X_j},T_k) \rightarrow\ & \mathbf{DV}_{J,k} \equiv [\mathrm{DV}_{i,j_1,\ldots,j_{n_x},k}]_{n_x\times N_x\times\cdots\times N_x}\\
& = \left[\left(V_{j_1+\delta_{i,1},\ldots,j_{n_x}+\delta_{i,n_x},k}\right.\right.\\
& \left.\left.-V_{j_1-\delta_{i,1},\ldots,j_{n_x}-\delta_{i,n_x},k}\right)/\Delta X_i\right]_{n_x\times N_x\times\cdots\times N_x},\\
\nabla_{\mathbf{x}}\left[\nabla_{\mathbf{x}}^{\top}[v^*]\right](\mathbf{X_j},T_k) \rightarrow\ & \mathrm{DDV}_{J,k} \equiv [\mathrm{DDV}_{i,j,j_1,\ldots,j_{n_x},k}]_{n_x\times n_x\times N_x\times\cdots\times N_x},\\
u^{(\mathrm{reg})}(\mathbf{X_j},T_k) \rightarrow\ & \mathbf{UR}_{J,k} \equiv [\mathrm{UR}_{i,j_1,\ldots,j_{n_x},k}]_{n_x\times N_x\times\cdots\times N_x}\\
& = -\left(C_{1,J,k}+F_{1,J,k}\mathrm{DV}_{j,k}\right).//,C_{2,J,k},\\
u^*(\mathbf{X_j},T_k) \rightarrow\ & \mathbf{US}_{J,k} \equiv [\mathrm{US}_{i,j_1,\ldots,j_{n_x},k}]_{n_x\times N_x\times\cdots\times N_x}\\
& = [\min(\mathrm{UMAX}_i,\max(\mathrm{UMIN}_i\\
& ,\mathrm{UR}_{i,j_1,\ldots,j_{n_x},k}))]_{n_x\times N_x\times\cdots\times N_x},\\
v^*(\mathbf{X_j}+\widehat{\boldsymbol{h}}_{0,\ell}(\mathbf{X_j},T_k,q_\ell),T_k) \rightarrow\ & \mathrm{VH}_{J,k}(q_\ell).
\end{aligned} \tag{8.37}$$

where $\delta_{i,j}$ is the Kronecker delta, $F_{i,J,k} = f_i(X_J,T_k)$ for $i = 0:1$, $C_{i,J,k} = c_{i-}(X_J,T_k)$ for $i = 0\,{:}\,2$, the symbol "./" denotes element-wise division, $\mathrm{UMIN}_i = U_i^{(\min)}$ for $i = 1\,{:}\,n_x$ and $\mathrm{UMAX}_i = U_i^{(\max)}$ for $i = 1\,{:}\,n_x$. The hypercube form of the control constraints is used here only for a concrete example, and can be replaced for what is appropriate in the application of interest.

The Hessian matrix is not necessarily diagonal and is only so if the diffusion coefficient $0.5(g_0g_0^{\top})(\mathbf{x},t)$ is diagonal, so the full, asymmetric Hessian is given here:

$$\begin{aligned}
\mathrm{DDV}_{J,k} \equiv\ & \left[\mathrm{DDV}_{i,j,j_1,\ldots,j_{n_x},k}\right]_{n_x\times n_x\times N_x\times\cdots\times N_x}\\
=\ & \left[\left(V_{j_1+\delta_{i,1},\ldots,j_{n_x}+\delta_{i,n_x},k}-2V_{j_1,\ldots,j_{n_x},k}+V_{j_1-\delta_{i,1},\ldots,j_{n_x}-\delta_{i,n_x},k}\right)\delta_{i,j}/\Delta X_i^2\right.\\
& +0.25\left(V_{j_1+\delta_{i,1}+\delta_{j,1},\ldots,j_{n_x}+\delta_{i,n_x}+\delta_{j,n_x},k}\right.\\
& -V_{j_1-\delta_{i,1}+\delta_{j,1},\ldots,j_{n_x}-\delta_{i,n_x}1+\delta_{j,n_x},k}-V_{j_1+\delta_{i,1}-\delta_{j,1},\ldots,j_{n_x}+\delta_{i,n_x}-\delta_{j,n_x},k}\\
& \left.+V_{j_1-\delta_{i,1}-\delta_{j,1},\ldots,j_{n_x}-\delta_{i,n_x}-\delta_{j,n_x},k}\right)\\
& \left.\cdot(1-\delta_{i,j})/(\Delta X_i\Delta X_j)\right]_{n_x\times n_x\times N_x\times\cdots\times N_x},
\end{aligned} \tag{8.38}$$

in the second order accuracy, central finite difference form. If the Hessian is diagonal, then only the second line of (8.38) is needed. The off-diagonal terms, i.e., when $i \neq j$, are conveniently calculated as the operator product of two central finite differences for the two independent partials. In the case where the off-diagonal terms are significant enough that they can affect stability and convergence, Kushner and Dupuis [179] recommend a better form form than that given in (8.38) for the cross term in $\mathrm{DDV}_{J,k}$.

These are the basic numerical ingredients for converting the one-state problem **Crank-Nicolson Extrapolator-Predictor-Corrector method** in Subsection 8.1.2 to the multi-state problem.

Curse of Dimensionality

In the full Hessian case, the Hessian is the largest array that will be needed in the computation and will basically determine the order of both computing and memory

demands for the solution of the PDE of SDP. In this full case the demands per time-step k will then be roughly proportional to the order of the $\mathrm{DDV}_{\mathbf{j},k}$ count or

$$\mathrm{O}\left(N_{\mathrm{DDV}}\right) = \mathrm{O}\left(n_x^2 \cdot \prod_{i=1}^{n_x} N_x\right) = \mathrm{O}\left(n_x^2 \cdot N_x^{n_x}\right) = \mathrm{O}\left(n_x^2 \cdot e^{n_x \ln(N_x)}\right) \ , \quad (8.39)$$

which is n_x times the size of the vector functions like $\mathrm{DV}_{\mathbf{j},k}$ and will grow exponentially with state dimension times the logarithm of the common number of nodes per dimension. If the number of nodes per dimension varies, i.e., N_i nodes in dimension i, then the geometric mean $N_x = \left(\sum_{i=1}^{n_x} N_i\right)^{1/n_x}$ can be used in place of the common value N_x in the above exponential estimate. This exponential growth in demands quantifies the exponential complexity in solving the PIDE of SDP and is called **Bellman's curse of dimensionality**. However, the very same exponential complexity (8.39) is found in high dimensional, second order PDEs. If there are $n_x = 6$ states and there are $N_x = 64$ nodes per state using 8-byte (8B) or double words, then the order of the amount of storage required is $N_{\mathrm{DDV}} = 8 \cdot 6^2 \cdot 64^6\mathrm{B} = 18\mathrm{GB}$, where 1GB is a gigabyte or a computer billion bytes or 1024^4 bytes.

If the discrete Hessian is diagonal, then the amount of storage needed is reduced to some multiple of

$$N_{\mathrm{DV}} = 8 \cdot n_x \cdot N_x^{n_x}\mathrm{B},$$

using 8 byte (8B) words, DDV that has the same size as DV, so in the example with $n_x = 6$ and $N_x = 64$, $N_{\mathrm{DV}} = 8 \cdot 6 \cdot 64^6\mathrm{B} = 3\mathrm{GB}$, a more reasonable size for a large scale problem capable computer. If the number of nodes per dimension is reduced to 32 instead of 64, then the amount of storage needed is some multiple of $8 \cdot 6 \cdot 32^6\mathrm{B} = 49,152\mathrm{MB} = 0.0469\mathrm{GB}$, approaching PC desktop capability (1MB being a megabyte or or 1024^2 bytes). The growth of the curse of dimensionality in the logarithm to the base 2 scale is illustrated in Fig. 8.1 for the diagonal Hessian size case N_{DV}. Note the top scale in the figure is about 60 log(B) and $2^{60}\mathrm{B} = 1024^6\mathrm{B}$ is 1 terabytes (1TB) or $1024^2 GB$ (1GB = 2^{40}B, while 1MB = 2^{20}B) and that is well within the capabilities of our current largest scale computers.

For parallel processing techniques in computational stochastic programming refer to Hanson's 1996 chapter [108]. See also [109] for more general supercomputing techniques that were developed originally solving computational control application problems.

8.2 Markov Chain Approximation for SDP

Another method for numerically solving stochastic dynamic programming problems in continuous time is Kushner's **Markov chain approximation** (MCA) [174, 175] that implicitly provides good convergence properties by normalizing the corresponding finite differences as proper Markov chains. In addition, MCA facilitates the proof of weak convergence using probabilistic arguments. Kushner and Dupuis's [179] method of using an auxiliary stochastic process, so that the composite stochastic process properly satisfies boundary conditions, is also treated. The summary here is in the spirit of this applied text to make the Markov chain approximation method

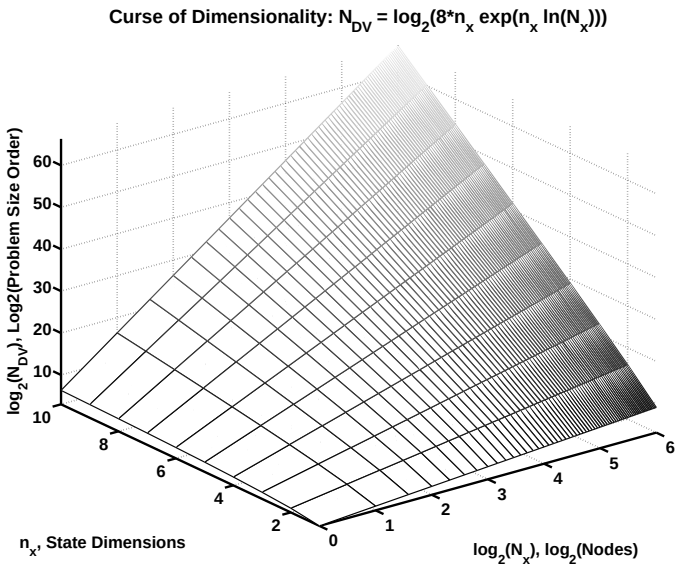


Figure 8.1. *Estimate of the logarithm to the base 2 of the order of the growth of memory and computing demands using 8 byte words to illustrate the curse of dimensionality in the diagonal Hessian case for $n_x = 1:10$ dimensions and $N_x = 1{:}64 = 1{:}2^6$ nodes per dimension. Note that 1KB or one kilobyte has a base 2 exponent of $10 = log_2(2^{10})$, while the base 2 exponent is 20 for 1MB, 40 for 1GB and is 60 for 1TB.*

more accessible, concentrating on the techniques, rather than the problems and formal definitions.

8.2.1 The MCA Formulation for Stochastic Diffusions

Although MCA is valid for jump-diffusions, only diffusions will be considered here to keep the complexity manageable and the reader can consult [179] for a more complete treatment of MCA. Let the diffusion satisfy the SDE,

$$dX(t) \stackrel{\text{sym}}{=} f(X(t), U(t), t)dt + g(X(t), t)dW(t) \,, \tag{8.40}$$

where the notation otherwise is the same as in (8.1) of the prior section, with f and g being bounded, continuous and Lipshitz continuous in X, while f has the same properties in U, but uniformly. For later reference, the following conditional infinitesimal moments are

$$\begin{aligned} \mathrm{E}[dX(t)|X(t) = x, U(t) = u] = &\quad f(x, u, t)dt, \\ \mathrm{Var}[dX(t)|X(t) = x, U(t) = u] = &\quad g^2(x, t)dt. \end{aligned} \tag{8.41}$$

Let the minimal, expected costs be defined as

$$v^*(x,t) \equiv \min_{U[t,t_f)} \left[\underset{(dW,dP)[t,tf)}{\mathrm{E}} \left[\int_t^{t_f} C(X(s),U(s),s)ds + S(X(t_f),t_f) \Big| X(t)=x, U(t)=u \right] \right], \tag{8.42}$$

for $t_0 \le t < t_f$. The corresponding PDE of stochastic dynamic programming is

$$\begin{aligned} 0 = \;& v_t^*(x,t) + \min_u \left[\mathcal{H}(x,u,t)\right] \\ \equiv \;& v_t^*(x,t) + \min_u \left[C(x,u,t) + f(x,u,t)v_x^*(x,t) + \tfrac{1}{2}g^2(x,t)v_{xx}^*(x,t)\right] \\ = \;& v_t^*(x,t) + \mathcal{H}^*(x,t)\ . \end{aligned} \tag{8.43}$$

The first step of the numerical part of the MCA procedure is to approximate the backward PDE (8.43) by a backward Euler method in time for simplicity. Then using the kth time step at t_k with optimal value $v_k(x) \simeq v^*(x,t_k)$, the next value is

$$v_{k-1}(x) = \; v_k(x) + \Delta t_{k-1} \min_u \left[C_k(x,u) + f_k(x,u)v_k'(x) + \tfrac{1}{2}g_k^2(x)v_k''(x)\right]\ , \tag{8.44}$$

for forward index $k = 1:N_t$, $t_k \equiv t_{k-1} + \Delta t_{k-1}$, $t_{N_t} = t_f$, $C_k(x,u) = C(x,u,t_k)$, $f_k(x,u,t_k)$ and $g_k(x) = g(x,t_k)$. The final condition is $v_{N_t}(x) = S(x,t_f)$. The time step Δt_{k-1} is called the MCA **interpolation time increment** and is selected to help form a proper Markov chain for convergence, so the increments are not necessarily constant. Though motivated by an approximation in time, time has been removed from the problem, i.e., the current problem is actually time-independent. Finite differences in the state come after specifying diffusion consistency conditions.

8.2.2 MCA Local Diffusion Consistency Conditions

Let ξ_k for $k \ge 0$ denote a Markov chain of discrete stages, intended as a discrete model for the state x, whose spacing is the order of some state mesh measure ΔX, i.e., $|\Delta\xi_k| = \mathrm{O}(\Delta X)$ where $\Delta\xi_k \equiv \xi_{k+1} - \xi_k$. Let the Markov chain **transition probability** for diffusions (D) be defined by

$$p^{(D)}(x,y|u) \equiv \mathrm{Prob}[\xi_{k+1} = y | \xi_j, u_j, j<k, \xi_k = x, u_k = u] \tag{8.45}$$

for transitions from current stage $\xi_k = x$ to the next stage $\xi_{k+1} = y$ using control policy $u_k = u$. (The term stage is used to denote a discrete state.) These transitions must satisfy the probability rules of non-negativity $p^{(D)}(x,y|u) \ge 0$ and probability conservation for transitions, $\sum_\ell p^{(D)}(x,X_\ell|u) = 1$, under current control u and over probable state transitions $y = X_\ell$. The increments $\Delta\xi_k$ must satisfy the MCA **local diffusion consistency** conditions:

$$\begin{aligned} \mathrm{E}[\Delta\xi_k|x,u] \equiv \;& \textstyle\sum_\ell (X_\ell - x)\cdot p^{(D)}(x,X_\ell|u) = \Delta t_{k-1}\cdot(f_k(x,u) + \mathrm{o}(1)); \\ \mathrm{Var}[\Delta\xi_k|x,u] \equiv \;& \textstyle\sum_\ell (X_\ell - x - \mathrm{E}[\Delta\xi_k|x,u])^2 \cdot p^{(D)}(x,X_\ell|u) \\ = \;& \Delta t_{k-1}\cdot(g_k^2(x) + \mathrm{o}(1)), \end{aligned} \tag{8.46}$$

with $\Delta\xi_k \to 0^+$ as $\Delta X \to 0^+$, for $k = 0 : N_t - 1$. The conditions are consistent with the first two conditional infinitesimal moments (8.41) of a stochastic diffusion approximation corresponding to the SDE (8.40), so they are neccesary preconditions for convergence of the Markov chain to the diffusion SDE (8.40).

See Sect. 7.8 on p. 216 or Feller, vol. II [84]) for more information. Also, see Kloeden and Platen [165] for stricter definitions of diffusion consistency conditions. The generalization of these diffusion consistency conditions to jump-diffusions is much more complicated, but is treated in Subsect. 8.2.4.

The discrete process can be used to construct a **piece-wise constant** ($pw\not{c}$) **interpolation** of the state and control processes in continuous time, i.e.,

$$(X^{(\mathrm{pw}\not{c})}(t), U^{(\mathrm{pw}\not{c})}(t) = \{(\xi_k, u_k), \quad t_{k-1} \le t < t_{k-1} + \Delta t_{k-1} = t_k, \quad \text{for } k \ge 1\}, \tag{8.47}$$

with the relationship between the interpolation times t_k and interpolation time increments Δt_{k-1} being $t_{k+1} = \sum_{j=0}^{k} \Delta t_j$. In general, the time increments will depend on ξ_k and u_k, which also depends on the order of state mesh ΔX, so $\Delta t_{k-1} = \Delta t_{k-1}(\xi_k, u_k; \Delta X)$. As the state mesh goes ot zero, it is required that the maximal state mesh go to zero, i.e., $\max_{u,x}[\Delta t_{k-1}(x, u; \Delta X)] \to 0^+$.

8.2.3 MCA Numerical Finite Differences for State Derivatives and Construction of Transition Probabilities

Construction of the Markov chain transition probabilities is found by finite differencing the state derivative. The state derivative is upwinded by first order forward or backward differences (UFD) for greater stability depending on the sign of the drift coefficient $f_k(x, u, t)$ as in (8.33),

$$v_k'(x) \simeq \left\{ \begin{array}{ll} (\frac{v_k(x + \Delta X) - v_k(x)}{\Delta X}, & f_k(x, u) \ge 0 \\ (\frac{v_k(x) - v_k(x - \Delta X)}{\Delta X}, & f_k(x, u) < 0 \end{array} \right\} \tag{8.48}$$

and central finite differences (CFDs) of second order accuracy are used for the second state partial

$$v_k''(x) \simeq \frac{v_k(x + \Delta X) - 2v_k(x) + v_k(x - \Delta X)}{\Delta X^2}. \tag{8.49}$$

Alternately, second order upwinding can be used for the state first derivative so that the accuracy is consistent with $\mathrm{O}(\Delta X^2)$ accuracy of the second derivative used above, but this leads to a double jump in the state by $2 \pm \Delta X$ so this complication will not be introduced here although the larger $\mathrm{O}(\Delta X)$ error **numerically pollutes** the smaller $\mathrm{O}(\Delta X^2)$ error for small ΔX. Using the $\mathrm{O}(\Delta X^2)$ forward and backward finite differences of the form used for derivative boundary conditions in (8.32) would not be useful since the alternating signs would lead to improper, negative transition probabilities for a least one double step transition.

Substituting into Eq. (8.44) for $v_{k-1}(x)$ and then collecting the coefficients in terms of transition probabilities,

$$\begin{aligned} v_{k-1}(x) = \quad & \min_{u_k-1}\Big[\Delta t_{k-1}\cdot C_k(x,u_{k-1}) + p_k^{(D)}(x,x|u_{k-1})\cdot v_k(x) \\ & + p_k^{(D)}(x,x+\Delta X|u_{k-1})\cdot v_k(x+\Delta X) \\ & + p_k^{(D)}(x,x-\Delta X|u_{k-1})\cdot v_k(x-\Delta X)\Big], \end{aligned} \tag{8.50}$$

the transition probabilities are found to be

$$p_k^{(D)}(x,x|u_{k-1}) = 1 - \frac{\Delta t_{k-1}}{\Delta X^2}\cdot\left(g_k^2(x) + \Delta X|f_k(x,u_{k-1})|\right), \tag{8.51}$$

$$p_k^{(D)}(x,x+\Delta X|u_{k-1}) = \frac{\Delta t_{k-1}}{\Delta X^2}\cdot\left(0.5g_k^2(x) + \Delta X[f_k(x,u_{k-1})]_+\right), \tag{8.52}$$

$$p_k^{(D)}(x,x-\Delta X|u_{k-1}) = \frac{\Delta t_{k-1}}{\Delta X^2}\cdot\left(0.5g_k^2(x) + \Delta X[f_k(x,u_{k-1})]_-\right), \tag{8.53}$$

where $[f]_\pm \equiv \max[\pm f] \geq 0$. Upwinding ensures that all terms in the coefficients of Δt_{k-1} are non-negative, so that the up and down transition probabilities, $p_k^{(D)}(x,x+\Delta X|u_{k-1})$ and $p_k^{(D)}(x,x-\Delta X|u_{k-1})$ are nonnegative. Note that on the right-hand-side of the conservation law (8.50) for the transition probabilities to get the value function for the past time t_{k-1}, the value function is evaluated at the current time t_k, but the control is for the past time t_{k-1} which makes it seem like the control is implicit. However, u_{k-1} is thought to be the control to get the state x from time t_{k-1} to time t_k and the optimization over u_{k-1} will determine u_{k-1} in terms of values at t_k anyway, so is not really an implicit term. Genuine implicit methods are discussed in Kushner and Dupuis [179].

It is clear that Δt_{k-1} must be sufficiently small so that the state self-transition probability $p_k^{(D)}(x,x|u_{k-1})$ is non-negative, i.e., is a proper probability. This implies the following convergence criteria

$$\frac{\Delta t_{k-1}}{\Delta X^2} \leq \frac{1}{\gamma_k^2(x) + \Delta X|f_k(x,u_{k-1})|} \tag{8.54}$$

or in terms of a generalization of the parabolic mesh ratio condition

$$\left(g_k^2(x) + \Delta X|f_k(x,u_{k-1})|\right)\cdot\frac{\Delta t_{k-1}}{(\Delta X)^2} \leq 1, \tag{8.55}$$

including both the diffusion coefficient and the upwinded drift term in the scaling of $\Delta t_{k-1}/(\Delta X)^2$. Since (8.54) should hold for all discrete time steps k, then we should have

$$\max_{x,u,k}\left[(g_k^2(x) + \Delta X|f_k(x,u)|)\frac{\Delta t_{k-1}}{\Delta X^2}\right] \leq 1. \tag{8.56}$$

The diffusion consistency conditions (8.46) can be confirmed in this three local state case directly,

$$\begin{aligned}
\mathrm{E}[\Delta\xi_k|x,u_{k-1}] = &\ p_k^{(D)}(x,x|u_{k-1})\cdot 0 + p_k^{(D)}(x,x+\Delta X|u_{k-1})\cdot(+\Delta X)\\
&+p_k^{(D)}(x,x-\Delta X|u_{k-1})\cdot(-\Delta X)\\
= &\ \Delta t_{k-1}\cdot([f_k(x,u_{k-1})]_+ - [f_k(x,u_{k-1})]_-)\\
\equiv &\ \Delta t_{k-1}\cdot f_k(x,u_{k-1}),
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}[\Delta\xi_k|x,u_{k-1}] = &\ p_k^{(D)}(x,x|u_{k-1})\cdot(\Delta t_{k-1}f_k(x,u_{k-1}))^2\\
&+p_k^{(D)}(x,x+\Delta X|u_{k-1})\cdot(\Delta X-\Delta t_{k-1}f_k(x,u_{k-1}))^2\\
&+p_k^{(D)}(x,x-\Delta X|u_{k-1})\cdot(-\Delta X-\Delta t_{k-1}f_k(x,u_{k-1}))^2\\
= &\ \Delta t_{k-1}\cdot\big(g_k^2+|f_k(x,u_{k-1})|\Delta X-2\Delta t_{k-1}f_k^2(x,u_{k-1})\big)\\
= &\ \Delta t_{k-1}\cdot\big(g_k^2+\mathrm{o}(1)\big)
\end{aligned}$$

as $\Delta X\to 0^+$ and consequently $\Delta t_{k-1}\to 0^+$.

Upon proper choice of the time and state grids satisfying (8.56), for example in the case of regular grids as used in the previous section in (8.10) with N_t nodes in t on $[t_0,t_f]$ and N_x nodes in x on $[x_0,x_{\max}]$, $T_k=t_f-(k-1)\Delta t$ for $k=1:N_t$, $\Delta t_{k-1}=\Delta t=(t_f-t_0)/(N_t-1)$ and $X_j=x_0+(j-1)\Delta X$ for $j=1:N_x$, $\Delta X=(x_{\max}-x_0)/(N_x-1)$, then

$$\begin{aligned}
V_{j,k-1} \equiv &\ v_{k-1}(X_j)\\
= &\ \Delta t\cdot C_k(X_j,U_{j,k-1}) + p_k^{(D)}(X_j,X_j|U_{j,k-1})\cdot V_{j,k}\\
&+p_k^{(D)}(X_j,X_{j+1}|U_{j,k-1})\cdot V_{j+1,k}\\
&+p_k^{(D)}(X_j,X_{j-1}|U_{j,k-1})\cdot V_{j-1,k},
\end{aligned}\tag{8.57}$$

when the optimal control is

$$\begin{aligned}
U_{j,k-1} = &\ \mathrm{argmin}_{u_{k-1}}\Big[\Delta t_{k-1}\cdot C_k(X_j,u_{k-1}) + p_k^{(D)}(X_j,X_j|u_{k-1})\cdot V_{j,k}\\
&+p_k^{(D)}(X_j,X_{j+1}|u_{k-1})\cdot V_{j+1,k}\\
&+p_k^{(D)}(X_j,X_{j-1}|u_{k-1})\cdot V_{j-1,k}\Big],
\end{aligned}\tag{8.58}$$

for $j=1:N_x$ for each stage $k=N_t:-1:2$ in backward order. Note that in [179], Kushner and Dupuis suggest a preference for selecting the interpolation time-step Δt_{k-1} so that the self-transition probability $p^{(D)}(x,x|u)$ vanishes leading to a renormalization of the non-self-transition probabilities like $p^{(D)}(x,x\pm\Delta X|u)$.

In this section, the **Markov chain approximation** has only been summarized to convey the main ideas, but for those interested in the weak convergence proofs and related theory they should consult [176, 179] and additional references therein.

8.2.4 MCA Extensions to Include Jump Processes

In [179, Sect. 5.6, pp. 127-133], Kushner and Dupuis briefly present the extensions of the **Markov chain approximation** for diffusions to that for jump-diffusions. Earlier Kushner and DiMasi [178] made contributions to the jump-diffusion optimal control problem, while Kushner [177] more recently gave further results on existence and numerical methods for the problem.

The main idea is based upon the facts that the Poisson process is instantaneous compared to the continuity of the diffusion process and that the Poisson process during short time intervals Δt can be asymptotically treated as a zero-one Bernoulli process as mentioned in prior chapters. Starting with the jump-diffusion SDE extension of (8.40),

$$\begin{aligned} dX(t) \stackrel{\text{sym}}{=} \quad & f(X(t),U(t),t)dt + g(X(t),t)dW(t)\ , \\ & +h(X(t),U(t),t,Q)dP(t;Q,X(t),U(t),t)\ , \end{aligned} \tag{8.59}$$

where $dP(t;Q,X(t),U(t),t)$ is the differential Poisson process with rate $\lambda(t;x,u,t)$, $h(x,u,t,q)$ is the state jump-amplitude and generalized probability density $\phi_Q(q)$. The conditional infinitesimal moments are given by

$$\begin{aligned} \mathrm{E}[dX(t)|X(t)=x,U(t)=u] = \quad & f(x,u,t)dt + \mathrm{E}_Q[h(x,u,t,Q)]\lambda(t;x,u,t)dt, \\ \mathrm{Var}[dX(t)|X(t)=x,U(t)=u] = \quad & g^2(x,t)dt + \mathrm{E}_Q[h^2(x,u,t,Q)]\lambda(t;x,u,t)dt. \end{aligned} \tag{8.60}$$

By separability of the diffusion and the jumps for sufficiently small time-steps Δt_{k-1}, the diffusion transition probabilities are unchanged, $p_k^{(D)}(x,y|u)$ for stage k. The probability of zero or one Poisson jump in time-steps of Δt_{k-1} can be written

$$p_{j,k}^{(J)} = \left\{ \begin{array}{ll} 1-\lambda\Delta t_{k-1}+\mathrm{o}(\Delta t_{k-1}), & j=0 \text{ jumps} \\ \lambda\Delta t_{k-1}+\mathrm{o}(\Delta t_{k-1}), & j=1 \text{ jump} \\ \mathrm{o}(\Delta t_{k-1}), & j\geq 2 \text{ jumps} \end{array} \right\}, \tag{8.61}$$

as $\Delta t_{k-1} \to 0^+$.

For the discretization jump-amplitude function $h(x,t,q)$ of the corresponding compound Poisson process, a concrete rather than the abstract formulation of Kushner and Dupuis [179] will be given so that the transition of a piece-wise-constant pre-jump stage $x = X_j$ for some j to a piece-wise-constant post-jump stage $y = X_\ell$ for some ℓ, where $X_{j+1} = X_j + \Delta X_j$ for $j = 1 : N_x - 1$, $X_1 = x_0$, $X_{N_x} = x_{\max}$ and the mesh is given by $\Delta X = \max_j(\Delta X_j) \to 0^+$. However, the treatment of jumps is much more complicated than that for diffusion whose dependence is only local, depending on only nearest neighbor or similarly close nodes, but jump behavior is globally dependent on nodes that may be remote from the current node X_j. Also, the connection of the jump-amplitude function to the jump-amplitude random mark variable q will be clarified. The jump-amplitude may be continuously distributed due to a continuous mark density $\phi_Q(q)$. It is assumed that post-jump stage $y = x + h(x,t,q)$ is uniquely invertible with q as a function of y given x, but it is necessary to have a set target $\mathcal{S}(X_\ell)$ rather than a point target $y = X_\ell$ so a

corresponding set $\mathcal{Q}_{j,\ell}(t)$ of positive probability measure can be found. Let $\mathcal{S}(X_\ell)$ be a partition of the state domain $[X_1, X_{N_x}]$ such that

$$\sum_{\ell=1}^{N_x} \mathcal{S}(X_\ell) = [X_1, X_{N_x}].$$

The $\mathcal{S}(X_\ell)$ will usually depend on the application due particular boundary conditions, singular points or related zero points, which could lead to forward or backward shifted intervals or intervals centered about X_ℓ as with rounding. The discretized, here piece-wise-continuous (*pwc*), instead of the prior piece-wise-constant (*pwc*) step functions, jump-amplitude $H_{j,\ell}^{(\text{pwc})}(t)$ given the stage set $\mathcal{S}(X_\ell)$ is

$$H_{j,\ell}^{(\text{pwc})}(t) = h(X_j, t, \mathcal{Q}_{j,\ell}(t)) = \mathcal{S}(X_\ell) - X_j, \tag{8.62}$$

implicitly defining the mark set $\mathcal{Q}_{j,\ell}(t)$ for $1 \le j < \infty$ and $1 \le \ell < \infty$. This ensures that a jump takes a proper (*pwc*) stage X_j to a proper (*pwc*) stage X_ℓ defined by the set $\mathcal{S}(X_\ell)$. Given a jump it is also necessary to know the corresponding probability of the transition referenced by (8.62) , i.e.,

$$\begin{aligned} &\text{Prob}\left[y = x + h(x,t,q) \in \mathcal{S}(X_\ell) \mid x = X_j,\ y \in \mathcal{S}(X_\ell)\right] \\ &\qquad = \overline{\Phi}(X_j, X_\ell, t) \equiv \textstyle\int_{\mathcal{Q}_{j,\ell}(t)} \phi_Q(q) dq, \end{aligned} \tag{8.63}$$

where $\phi_Q(q)$ is the generalized mark density with corresponding distribution $\Phi_Q(q)$, except that when $h(X_{\hat{j}}, t, q) = 0$ for some $\hat{j}$, i.e., there is a **zero jump** and $y \in \mathcal{S}(X_\ell)$ is not achievable for general ℓ, then $\overline{\Phi}_{\hat{j},\ell}(t) \equiv 0$. In the case that $\overline{\Phi}(X_j, X_\ell, t)$ leads to a probabilistically deficient distribution, in general the renormalized form is

$$\widehat{\Phi}(X_j, X_\ell, t) = \overline{\Phi}(X_j, X_\ell, t) \Big/ \overline{\overline{\Phi}}(X_j, t)\ , \tag{8.64}$$

where

$$\overline{\overline{\Phi}}(X_j, t) \equiv \sum_{\ell=1}^{N_x} \overline{\Phi}(X_j, X_\ell, t) = \sum_{\ell=1}^{N_x} \int_{\mathcal{Q}_{j,\ell}(t)} \phi_Q(q) dq.$$

Example 8.3. ***Geometric Jump-Diffusion Target Mark Set Calculations:*** *For the geometric jump-diffusion used in finance, with linear jump-amplitude*

$$h(x,t,q) = xJ(q,t),$$

it is convenient to choose the log-return jump as the mark, i.e.,

$$q = [\ln(X)](t) = \ln\left(\left(X(t^-) + X(t^-)J(q,t^-)\right)/X(t^-)\right) = \ln(1 + J(q,t^-)),$$

so $h(x,t,q) = x(\exp(q) - 1)$. *Hence,* $X_1 = x_0 = 0$ *is a* ***zero point*** *needing special treatment since there can be no target stage except for* $[X_1, X_1] = \{0\}$, *so that a proper partition of* $[X_1, X_{N_x}]$ *would be* $\mathcal{S}(X_1) = \{0\}$ *and* $\mathcal{S}(X_\ell) = (X_{\ell-1}, X_{\ell-2}]$ *for*

$\ell = 2{:}N_x$. The discrete jump-amplitude $H^{(\mathrm{pwc})}_{1,\ell}(t) \equiv 0$ for definiteness when $X_1 = 0$ and

$$H^{(\mathrm{pwc})}_{j,\ell}(t) \equiv X_\ell - X_j$$

for $\ell = 2{:}N_x$. The target mark set is

$$\mathcal{Q}_{j,\ell}(t) = (\ln(X_{\ell-1}/X_j), \ln(X_\ell/X_j)]$$

for $\ell = 2 : N_x$ when $j > 1$. Given a mark density, then a renormalized target distribution $\widehat{\Phi}(X_j, X_\ell, t)$ can be calculated.

The Markov chain approximation $\xi_k(\Delta X)$ is **locally jump-diffusion consistent** if there is an **interpolation time interval** $\Delta t_{k-1} = \Delta t(x, u; \Delta X) \to 0^+$ uniformly in $(x, u, \Delta X)$ as the mesh gauge $\Delta X \to 0^+$ and so that

1. Along with $\Delta t(x, u; \Delta X)$, there is a locally diffusion consistent transition probability $p^{(D)}(x, y \mid u; \Delta X)$ satisfying the conditions in (8.46);
2. The **jump-diffusion transition probabilities** $p^{(JD)}(x, y \mid u; \lambda, \Delta X)$ must conserve probability over the post-jump values $y = X_\ell$ from any given pre-jump value $x = X_j$, i.e.,
$$\sum_\ell p^{(JD)}(X_j, X_\ell \mid u; \lambda, \Delta X) = 1.$$
3. Markov chain increments $\Delta\xi_k$ satisfy the MCA **jump-diffusion local consistency** conditions consistent with the jump-diffusion conditional infinitesimal moments (8.60), with replacements $f(x, u, t) \to f_k, (x, u)$, $g(x, t) \to g_k(x)$, $h(x, t, q) \to h_k(x, q)$, $H^{(\mathrm{pwc})}_{j,\ell}(t) \to H^{(\mathrm{pwc})}_{j,\ell,k}$ $\widehat{\Phi}(X_j, X_\ell, t) \to \widehat{\Phi}_k(X_j, X_\ell)$, under current control u and over probable state transitions
$$\begin{aligned} \mathrm{E}[\Delta\xi_k \mid X_j, u_{k-1}] \equiv & \quad \textstyle\sum_\ell (X_\ell - X_j) \cdot p^{(JD)}(X_j, X_\ell \mid u_{k-1}; \lambda, \Delta X) \\ = & \quad \Delta t_{k-1} \cdot (f_k(X_j, u_{k-1}) + \lambda \mathrm{E}_Q[h_k(X_j, Q)] + \mathrm{o}(1)) ; \\ \mathrm{Var}[\Delta\xi_k \mid X_j, u_{k-1}] \equiv & \quad \textstyle\sum_\ell (X_\ell - X_j - \mathrm{E}[\Delta\xi_k \mid X_j, u_{k-1}])^2 \\ & \quad \cdot p^{(JD)}(x, X_\ell \mid u_{k-1}; \lambda, \Delta X) \\ = & \quad \Delta t_{k-1} \cdot \left(g_k^2(x) + \lambda \mathrm{E}_Q[h_k^2(X_j, Q)] + \mathrm{o}(1)\right), \end{aligned} \tag{8.65}$$
with $\Delta\xi_k \to 0^+$ as $\Delta X \to 0^+$, for $k = 0{:}N_t - 1$.
4. There is a small error factor $\varepsilon(s, u; \Delta X) = \mathrm{o}(\Delta t(x, u; \Delta X))$ that can be used to construct (Kushner and Dupuis [179], modified for clarification here) the **jump-diffusion transition probability** $p^{(JD)}(x, y \mid u; \lambda, \Delta X)$ and is of the form
$$\begin{aligned} & p^{(JD)}(X_j, X_\ell \mid u; \lambda, \Delta X) \\ & = (1 - \lambda\Delta t(X_j, u; \Delta X) - \varepsilon(X_j, u; \Delta X)) \cdot p^{(D)}(X_j, X_\ell \mid u; \Delta X) \\ & + (\lambda\Delta t(X_j, u; \Delta X) + \varepsilon(X_j, u; \Delta X)) \cdot \widehat{\Phi}_k(X_j, X_\ell) \mathbf{1}_{X_\ell \in X_j + H^{(\mathrm{pwc})}_{j,\ell,k}}, \end{aligned} \tag{8.66}$$

for $1 \leq j < \infty$ and $1 \leq \ell < \infty$, where $\mathbf{1}_{\mathcal{S}}$ is the indicator function for set $\mathcal{S} = \{X_\ell \in X_j + H^{(\text{pwc})}_{j,\ell,k}\}$ and is used so the term it multiplies is only used for a jump.

By using the conservation laws

$$\sum_{\ell=1}^{N_x} p^{(D)}(X_j, X_\ell \mid u; \Delta X) = 1$$

and

$$\sum_{\ell=1}^{N_x} \widehat{\Phi}(X_j, X_\ell, t) = 1,$$

it is easy to show the constructed jump-diffusion transition probability in (8.66) is conserved, i.e.,

$$\sum_{\ell=1}^{N_x} p^{(JD)}(X_j, X_\ell \mid u; \lambda, \Delta X) = 1.$$

The error factor $\varepsilon(s, u; \Delta X)$ reflects the the asymptotically small error terms $\mathrm{o}(\Delta t_{k-1})$ in the Poisson counting process definition (8.61), but is selected so the conservation is exact.

Using the first moment diffusion local consistency condition in (8.46) and a mark density weighted rectangular integration rule,

$$\mathrm{E}_Q[h_k(x, Q)] \simeq \sum_{\ell=1}^{N_x} H^{(\text{pwc})}_{j,\ell,k} \widehat{\Phi}_k(X_j, X_\ell).$$

Then,

$$\begin{aligned} \mathrm{E}[\Delta\xi_k \mid X_j, u] \simeq & \quad \Delta t_{k-1}(X_j, u; \Delta X) \cdot (f_k(X_j, u) + \mathrm{E}_Q[h_k(x, Q)] + \mathrm{o}(1)) \\ = & \quad \overline{X}^{(D)} + \overline{X}^{(J)}, \end{aligned}$$

splitting the diffusion and jump parts. Similarly, for the second moment jump-diffusion consistency condition, except with more algebra with the above splitting and more small time asymptotics in absorbing all quadratic and smaller time increments into $\Delta t_{k-1} \cdot \mathrm{o}(1)$, it can be demonstrated that

$$\mathrm{Var}[\Delta\xi_k \mid X_j, u] \simeq \Delta t_{k-1}(X_j, u; \Delta X) \cdot \left(g_k^2(X_j) + \mathrm{E}_Q\big[h_k^2(x, Q)\big] + \mathrm{o}(1)\right).$$

Further evaluations require knowledge of the mark density, the jump-diffusion coefficients (f, g, h) and the boundary condition on the state domain. Due to the global nature of the compound jump process with jump beyond the local nodes needed by the diffusion component process, the diffusion mesh ratio criteria (8.56) (or (8.30) in case the central finite differences are usable) will have to suffice for practical reasons. See Kushner and Dupius [179] for information on reflected boundary conditions and other techniques for handling boundary conditions when there are jumps.

Suggested References for Further Reading

- Chung, Hanson and Xu, 1992 [54].
- Douglas and Dupont, 1970 [72].
- Douglas, 1979 [73] .
- Dyer and McReynolds, 1979 [76].
- Gunzburger, 2003 [101].
- Hanson, 1989 [106], 1991 [107], 1996 [108] and 2003 [109, 110].
- Hanson and Naimipour, 1993 [111].
- Kushner, 1976 [174], 1990 [175], 2000a [176] and 2000b [177].
- Kushner amd DiMasi, 1978 [178].
- Kushner and Dupuis, 2001 [179].
- Kushner and Yin, 1997 [181].
- Larson, 1967 [182].
- Naimipour and Hanson, 1993 [216].
- Polak, 1973 [227].
- Press et al., 2002 [230].
- Westman and Hanson, 1997 [274] and 2000 [277].
- Zhu and Hanson, 2006 [291].

Chapter 9
Stochastic Simulations

Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin.
—John von Neumann (1903-1957),
apparently meant just as a caution,
`http://en.wikiquote.org/wiki/John_von_Neumann` .

Fast cars, fast women, fast algorithms...
what more could a man want?
—Joe Mattis at `http://www.xs4all.nl/~jcdverha/scijokes/1_5.html#subindex`.

Methods are considered that treat stochastic dynamics, such as direct simulations of SDEs [166, 165] with many numerical techniques offering improvements over the elementary integration methods beyond stochastic versions of Euler's method.

Monte Carlo methods simulate solutions to higher level applications, which include many improvements to increase the probable accuracy in order to reduce the need of large scale sample sizes, many of the techniques involve variance reduction and generation sample variates for nonuniform distributions [96, 150, 290].

9.1 SDE Simulation Methods

Simulation methods for the dynamics of stochastic differential equations are discussed. Basic simulation procedures have been introduced in Chapters 2-5, but here the simulation of diffusion and jump-diffusion simulation is discussed and explored much further. Primary references are by Kloeden et al. [166], Cyganowski et al. [65, 64, 66], the compact review by D. Higham [139] and D. Higham and Kloeden [143, 144]. Many of these references deal almost entirely with diffusions and the most comprehensive, theoretically and numerically, on diffusions is the mono-

graph of Kloeden and Platen [165]. The references of Maghsoodi [191], Cyganowski et al. [65, 64] and D. Higham and Kloeden [144] treat jump-diffusions in a serious way. However, random simulations to solve stochastic optimal control problems are not too useful due to the additional complexity involved in the optimization step, while optimal control problems can be reduced to deterministic ODE or PDE formulations which can be solved more systematically.

9.1.1 Convergence and Stability for Stochastic Problems and Simulations

Consider the jump-diffusion stochastic differential equation.

$$dX(t) = f(X(t),t)dt + g(X(t),t)dW(t) + h(X(t),t)dP(t), \tag{9.1}$$

$X(0) = x_0$ with probability one and $0 \le t \le t_f$, where the coefficient functions $f(X(t),t)$, $g(X(t),t)$ and $h(X(t),t)$ are continuously differentiable (see [165] for tighter conditions; $h(X(t),t)$ could also depend on random marks Q).

In Section 4.3.3, the main concern was formal SDE simulations, but here there will be more attention on convergence of the simulations. Let t_k denote a discrete time such that $t_{k+1} = t_k + \Delta t$ for $k = 0\!:\!N_t - 1$, so $t_{N_t} = t_f$ and $\Delta t = t_f/N_t$. For the state, let X_k denote the discrete approximation at time t_k to the exact value $X(t_k)$, i.e., $X_k \simeq X(t_k)$.

Definition 9.1. *The approximation X_k is said to* ***converge*** *to the exact value $X(t_k)$,*

- *in the* ***strong mean absolute error sense*** *if the conditional expectation,*

$$\mathrm{E}\left[|X_k - X(t_k)| \mid X(0) = x_0\right] \to 0^+ \quad \text{as} \quad \Delta t \to 0^+, \tag{9.2}$$

for fixed time $t_k = k\Delta t$, e.g., $t_f = t_{N_t}$;

the strong convergence in the mean absolute error is said to be ***order*** *or with* ***log-rate*** *$\gamma_s > 0$ in mean absolute error if*

$$\mathrm{E}\left[|X_k - X(t_k)| \mid X_0 = x_0\right] \le C_s \cdot (\Delta t)^{\gamma_s}, \tag{9.3}$$

for sufficiently small Δt, for fixed time $t_k = k\Delta t$, e.g., $t_f = t_{N_t}$ and constant $C_s > 0$.

- *in the* ***weak sense*** *if the difference in conditional expectations,*

$$\left|\mathrm{E}\left[X_k \mid X_0 = x_0\right] - \mathrm{E}\left[X(t_k) \mid X(0) = x_0\right]\right| \to 0^+ \quad \text{as} \quad \Delta t \to 0^+, \tag{9.4}$$

for fixed time $t_k = k\Delta t$, e.g., $t_f = t_{N_t}$;

the weak convergence is said to be ***order*** *or with* ***log-rate*** *$\gamma_w > 0$ in mean error if*

$$\left|\mathrm{E}\left[X_k \mid X_0 = x_0\right] - \mathrm{E}\left[X(t_k) \mid X(0) = x_0\right]\right| \le C_w \cdot (\Delta t)^{\gamma_w}, \tag{9.5}$$

for sufficiently small Δt, for fixed time $t_k = k\Delta t$, e.g., $t_f = t_{N_t}$ and constant $C_w > 0$.

- *Alternately,* ***strong convergence in mean square error (mse)****, instead of mean error, can be defined (Maghsoodi [191]),*

$$\sup_k \left(\mathrm{E}\left[(X_k - X(t_k))^2 \middle| X_0 = x_0 \right] \right) \leq C_s^{(\mathrm{mse})} \cdot (\Delta t)^{\gamma_s^{(\mathrm{mse})}}, \tag{9.6}$$

for sufficiently small Δt and constant $C_s^{(\mathrm{mse})} > 0$; thus the maximal root mean square error rate is

$$\mathrm{O}\left((\Delta t)^{\gamma_s^{(\mathrm{mse})}/2} \right),$$

so it is fair to compare the mean absolute error rate γ_s with the root mean square error rate $\gamma_s^{(\mathrm{mse})}/2$.

For ordinary differential equations, a solution $X(t)$ is **asymptotic stable** as $t \to +\infty$ if

$$\lim_{t \to +\infty} |X(t)| = 0,$$

in the continuous time case and in the discrete time case the approximation X_k is **asymptotic stable** as $k \to +\infty$ if

$$\lim_{k \to +\infty} |X_k| = 0.$$

However, such a definition is not applicable even if the coefficient functions are bounded and otherwise nicely behaved, since for diffusions the range of the random process $W(t)$ is infinite. Thus, the notion of stochastic asymptotic stability has to be modified for stochastic processes.

Definition 9.2.

- *For continuous time, the real stochastic solution $X(t)$ is said to be* ***asymptotically mean square stable*** *if*

$$\lim_{t \to +\infty} \mathrm{E}\left[X^2(t) \;\middle|\; X(0) = x_0 \right] = 0. \tag{9.7}$$

Alternately, $X(t)$ is ***asymptotically stable in probability*** *if*

$$\mathrm{Prob}\left[\lim_{t \to +\infty} |X(t)| = 0 \;\middle|\; X(0) = x_0 \right]. \tag{9.8}$$

- *For discrete time, the real stochastic approximation X_k is said to* ***asymptotically mean square stable***

$$\lim_{k \to +\infty} \mathrm{E}\left[X_k^2 \;\middle|\; X_0 = x_0 \right] = 0. \tag{9.9}$$

Alternately, X_k is ***asymptotically stable in probability*** *if*

$$\mathrm{Prob}\left[\lim_{k \to +\infty} |X_k| = 0 \;\middle|\; X_0 = x_0 \right]. \tag{9.10}$$

As a continuous-time example, consider the linear, constant coefficient SDE, letting $(f(x,t), g(x,t), h(x,t)) = (\mu_0, \sigma_0, \nu_0)$ in (9.1),

$$dX(t) = X(t)(\mu_0 dt + \sigma_0 dW(t) + \nu_0 dP(t)),$$

where μ_0, σ_0, ν_0 and λ_0 are constants and where $\mathrm{E}[dP(t)] = \lambda_0 dt$. From (4.80), the exact solution is

$$X(t) = x_0 \exp((\mu_0 - \sigma_0^2/2)t + \sigma_0 W(t))(1+\nu_0)^{P(t)}. \tag{9.11}$$

Using the independent increment techniques for the expectation in (4.81), the mean square is

$$\mathrm{E}\left[X^2(t_f) \,\middle|\, X(0) = x_0\right] = x_0^2 e^{(2(\mu_0+\lambda_0\nu_0)+\sigma_0^2+\lambda_0\nu_0^2)t_f}.$$

Thus, $X(t_f)$ is asymptotically mean square stable if the exponential is decaying as $t_f \to +\infty$, so

$$2(\mu_0 + \lambda_0\nu_0) + \sigma_0^2 + \lambda_0\nu_0^2 < 0, \tag{9.12}$$

which, in qualitative terms of the relative conditional infinitesimal moments, can be put in the form:

$$\mathrm{E}[dX(t)/X(t) \mid X(t)] < -0.5\mathrm{Var}[dX(t)/X(t) \mid X(t)],$$

assuming $x_0 > 0$ so $X(t) > 0$. Hence, the combined jump-diffusion relative infinitesimal mean has to be less than minus one-half of the relative infinitesimal variance.

9.1.2 Stochastic Diffusion Euler Simulations

The simplest simulation model using Euler's Method for SDEs is more properly called the **Euler-Maruyama (EM) method** to distinguish it from the deterministic Euler method for DEs and this was used in Section 4.3.3 in this text and has the stochastic difference form

$$X_{k+1} = X_k + F_k \Delta t + G_k \Delta W_k, \tag{9.13}$$

for $k = 0 : N_t - 1$, where $F_k \equiv f(X_k, t_k)$, $G_k \equiv g(X_k, t_k)$ and $\Delta W_k \equiv W(t_{k+1}) - W(t_k)$. For instance, in MATLAB™, a fragment of the code for the discrete diffusion approximation for a linear would be like that given in Fig. 9.1. Recall that MATLAB™ is unit based, .i.e., array subscripts start at one. In this example, the drift coefficient rate is time-dependent with $f(x,t) = \mu(t)x$ where $\mu(t) = 1/(1+0.5t)^2$, but the $dW(t)$-coefficient is time-independent with $g(x,t) = \sigma(t)x$ where $\sigma(t) = \sigma_0$ where σ_0 is a constant, i.e.,

$$dX(t) = X(t)(\mu(t)dt + \sigma(t)dW(t)). \tag{9.14}$$

In this case the log-transformation $Y(t) = \ln(X(t))$ by the Itô stochastic chain rule leads to a state-independent SDE, $Y(t) = (\mu(t) - \sigma^2(t)/2)dt + \sigma(t)dW(t)$ and a

```
function sdeeulersim
% Euler-Maruyama Simulation Test: Linear SDE:
%   dX(t) = X(t)(mu(t)dt+sigma(t)dW(t)),
% Given  Initial data: x0, t0, tf, Nt; functions: f, g, xexact
clc
%
randn('state',8); % Set random state or seed;
x0 = 1; t0 = 0; tf = 5; Nt = 2^14; DT = tf/Nt; sqrtt = sqrt(DT);
X(1) = x0; Xexact(1) = x0;  t = [t0:DT:tf];
DW = randn(1,Nt)*sqrtt; % Simulate DW as sqrt(DT)*randn;
W = cumsum(DW); % Omits initial zero value;
for k = 1:Nt % Exact formula to fine precision}
    Xexact(k+1) = xexact(x0,t(k+1),W(k)); % Calls subfunction;
end
L = 2^3; NL = Nt/L; KL = [0:L:Nt]; DTL = L*DT; tL = [t0:DTL:tf];
for k = 1:NL % Euler formula to lumped, coarse precision:
    DWL = sum(DW(1,KL(k)+1:KL(k+1)));
    Xeul(k+1)=Xeul(k)+f(Xeul(k),tL(k))*DTL+g(Xeul(k),tL(k))*DWL;
    Xdiff(k+1) = Xeul(k+1) - Xexact(KL(k+1));
end
plot(tL,Xeul,'k--','linewidth',3); hold on
plot(t,Xexact,'k-','linewidth',3); hold off
title('SDE Euler-Maruyama and Exact Linear SDE  Simulations');
xlabel('t, Time'); ylabel('X(t), State');
legend('X(t): Euler','Xexact: Exact','Location','Best');
%
function y = f(x,t)
    mu = 1/(1+0.5*t)^2; % Change with application;
    y = mu*x;
%
function y = g(x,t)
    sig = 0.5; % Change with application;
    y = sig*x;
%
function y = xexact(x0,t,w)
% exact solution if available for general linear SDE:
    mubar = 2-2/(1+0.5*t); sig = 0.5; sig2bar = sig^2*t/2;
    y = x0*exp(mubar-sig2bar + sig*w);
%End Code
```

Figure 9.1. *Code: Euler SDE simulations.*

simple integration followed by a transformation inversion leads to the general exact stochastic solution

$$X^{(\text{exact})}(t) = x_0 \exp(\overline{\mu}(t) - \overline{\sigma^2}(t)/2 + \overline{(\sigma * W)}(t)), \tag{9.15}$$

where $\overline{\mu}(t) = \int_0^t \mu(s)ds$, $\overline{\sigma^2}(t) = \int_0^t \sigma^2(s)ds$ and $\overline{(\sigma * W)}(t) = \int_0^t \sigma(s)dW(s)$, which in the simpler case here reduces the integral to $\overline{(\sigma * W)}(t) = \sigma_0 W(t)$, so that an

approximation of this diffusion integral in not necessary. Equation (9.15) is an exact formula, but comparison of the Euler-Maruyama approximation to that of the exact requires an approximate simulation of $W(t)$ in $\overline{(\sigma * W)}(t)$. Following D. Higham's [139] lead, a fine grid of N_t sample points is used for the exact formula and a lumped, coarse grid with $N_t/8$ points is taken from the set for the exact case. This makes for a more accurate comparison. The comparison between the coarse Euler-Maruyama approximation and the fine exact approximation $X^{(\text{exact})}(t)$ in (9.15), is illustrated in Fig. 9.2. The error between the Euler-Mayuyama approximate path

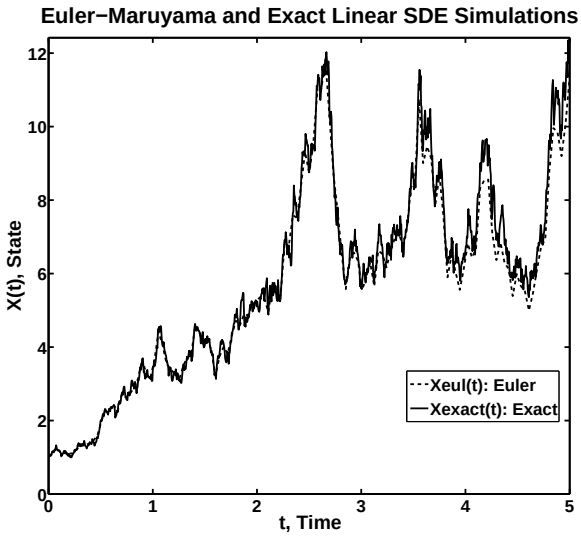


Figure 9.2. *Comparison of coarse Euler-Maruyama and fine exact paths, simulated using MATLAB with* $N_t = 1024$ *fine sample points for the exact path (9.15) and* $N_t/8 = 128$ *coarse points for the Euler path (9.13), initial time* $t_0 = 0$*, final time* $t_f = 5$ *and initial state* $x_0 = 1.0$*. Time-dependent parameter values are* $\mu(t) = 0.5/(1 + 0.5t)^2$ *and* $\sigma(t) = 0.5$*.*

and the exact path at the coarse time points is presented in Fig. 9.3. For further computer experiments verifying convergence using paths averages, see D. Higham [139]. For the complete sample code used to generate these Euler-Maruyama figures, see Sect. C.16 of Appendix C.

Kloeden and Platen [165, Section 10.2] show for the Euler-Maruyama simulation method, using a level of analysis beyond the scope of this text, that the log-rate of convergence in the strong sense is $\gamma_s = 0.5$, while in the weak sense the rate is $\gamma_w = 1$. Thus, the log-rate for convergence in the weak sense is the same as that for the traditional Euler's method applied to deterministic DEs in the strong or weak sense, i.e., $\gamma = 1$ for the deterministic case, since the expectation operator plays no role.

For convergence in the weak sense, the Euler-Maruyama method and the lin-

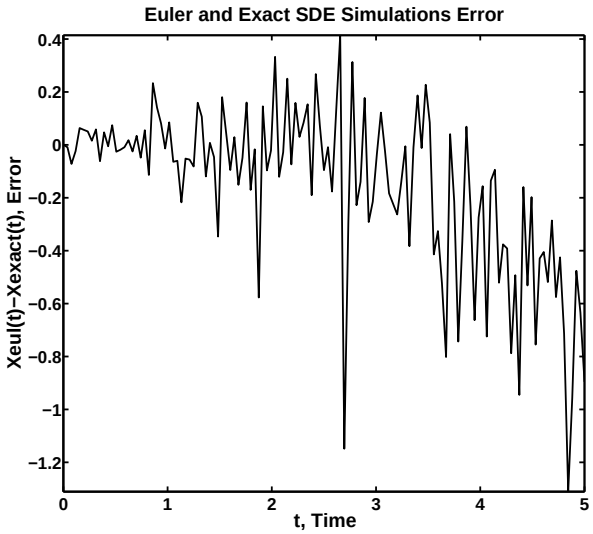


Figure 9.3. *Error in coarse Euler-Maruyama and fine exact paths using the coarse discrete time points. The simulations use MATLAB with the same values and time-dependent coefficients as in Fig. 9.2. The Euler maximal-absolute error for this example is* $1.3 \simeq 34\Delta t/8$*, while for* $N_t = 4096$ *the maximal error is better at* $0.28 \simeq 29\Delta t/8$*.*

ear, constant rate SDE,

$$dX(t) = \mu_0 X(t)dt + \sigma_0 X(t)dW(t),$$

where μ_0 and σ_0 are constants, the log-rate result can be shown with a reasonable effort. From (9.15) or (9.11) with $\nu_0 = 0$, the exact solution is

$$X^{(\text{exact})}(t) = x_0 \exp((\mu_0 - \sigma_0^2/2)t + \sigma_0 W(t)).$$

In this case, the EM approximation from (9.13) has the form of a stochastic difference equation (SΔE),

$$X_k = X_{k-1} \cdot (1 + \mu_0 \Delta t + \sigma_0 \Delta W_{k-1}), \tag{9.16}$$

for $k = 1{:}N_t$, and the expectation of X_k conditioned on the past value X_{k-1} is

$$\mathrm{E}[X_k \mid X_{k-1}] = X_{k-1} \cdot (1 + \mu_0 \Delta t),$$

so by iterated expectations

$$\mathrm{E}[X_k \mid X(0) = x_0] = (1 + \mu_0 \Delta t)\mathrm{E}[X_{k-1} \mid X_j, j = 0{:}k-2] = (1 + \mu_0 \Delta t)^k x_0$$

and finally $\mathrm{E}[X_{N_t} \mid X(0) = x_0] = x_0(1 + \mu_0\Delta t)^{N_t}$ at $t_{N_t} = t_f$. From (4.81), for jump-diffusions but ignoring the jumps, the expectation of the exact solution at the final fixed time is

$$\mathrm{E}\left[X^{(\mathrm{exact})}(t_f) \;\middle|\; X(0) = x_0 \right] = x_0 e^{\mu_0 t_f}.$$

The asymptotic evaluation, for sufficiently small Δt, of weak convergence criteria is then

$$\begin{aligned} \left|\mathrm{E}[X_{N_t} \mid X_0 = x_0] - \mathrm{E}[X^{(\mathrm{exact})}(t_f) \mid X(0) = x_0]\right| &= |x_0| \cdot \left|(1 + \mu_0\Delta t)^{N_t} - e^{\mu_0 t_f}\right| \\ &= |x_0| \cdot \left|e^{N_t \ln(1+\mu_0\Delta t)} - e^{\mu_0 N_t \Delta t}\right| \\ &\sim |x_0| e^{\mu_0 t_f} \cdot \left|e^{-0.5\mu_0^2 t_f \Delta t} - 1\right| \\ &\sim |x_0| e^{\mu_0 t_f} \cdot 0.5\mu_0^2 t_f \Delta t = \widetilde{C}_w \Delta t, \end{aligned}$$

so $\gamma_w = 1$, as advertised, with $\widetilde{C}_w = 0.5\mu_0^2 |x_0| \exp(\mu_0 t_f)$, for both linear deterministic and stochastic Euler's method, although only in the weak sense in the linear stochastic case.

Finally, consider the mean square stability of the EM approximation X_k. Recasting the EM SΔE (9.16) to the recursion form $X_k = A_{k-1} \cdot X_{k-1}$, where $A_k \equiv (1 + \mu_0\Delta t + \sigma_0\Delta W_k)$, so that the solution can be written

$$X_k = x_0 \prod_{\ell=0}^{k-1} A_\ell.$$

Next, considering the mean square,

$$\begin{aligned} \mathrm{E}[X_k^2 \mid X_0 = x_0] &= x_0^2 \mathrm{E}\left[\left(\textstyle\prod_{\ell=0}^{k-1} A_\ell \right)^2 \right] = x_0^2 \mathrm{E}\left[\textstyle\prod_{\ell=0}^{k-1} A_\ell^2 \right] = x_0^2 \textstyle\prod_{\ell=0}^{k-1} \mathrm{E}\left[A_\ell^2\right] \\ &= x_0^2 \textstyle\prod_{\ell=0}^{k-1} ((1+\mu_0\Delta t)^2 + \sigma_0^2\Delta t) = x_0^2 \left((1+\mu_0\Delta t)^2 + \sigma_0^2\Delta t\right)^k \qquad (9.17) \\ &= x_0^2 \left(1 + 2\mu_0\Delta t + (\mu_0\Delta t)^2 + \sigma_0^2\Delta t\right)^k, \end{aligned}$$

by interchanging the power and product operators, interchanging the product and expectation operators due to the independent increments property of the ΔW_k, using $\mathrm{E}[\Delta W_\ell] = 0$ and $\mathrm{E}[\Delta W_\ell^2] = \Delta t$, and the final fact that $\prod_{\ell=0}^{k-1} \theta = \theta^k$. Since as $k \to \infty$, $\theta^k \to 0$ if and only if $\theta < 1$ and in this case obviously $\theta > 0$, so asymptotic mean square stability of the X_k requires that

$$2\mu_0 + \sigma_0^2 + \mu_0^2\Delta t < 0. \qquad (9.18)$$

Note from (9.12) with $\nu_0 = 0$, the corresponding critical stability condition for the exact solution is $2\mu_0 + \sigma_0^2 < 0$ or that $\mu_0 < -0.5\sigma_0^2$ and that μ_0 must be sufficiently negative, but (9.18) for EM is much more restrictive requiring

$$\mu_0 < -0.5(\sigma_0^2 + \mu_0^2\Delta t),$$

since the discrete term $\mu_0^2\Delta t$ has been retained because Δt may not be so small to be negligible, although $\mu_0^2 dt$ would be negligible compared to one in the dt-precision

used in the exact, continuous time case. For numerical consideration, (9.18) could be interpreted as a constraint on the discrete time-step, i.e.,

$$\Delta t < 2 \left|\mu_0 + 0.5\sigma_0^2\right| / \mu_0^2 \,,$$

valid only if μ_0 is selected to be in the asymptotically mean square stable range, $\mu_0 < -0.5\sigma_0^2$, of the exact solution. For more elaborate discussion of asymptotic stability, see D. Higham [139] for diffusions or D. Higham and Kloeden [145] for jump-diffusions.

9.1.3 Milstein's Higher Order Stochastic Diffusion Simulations

It is difficult to see how to improve on the Euler-Maruyama method (9.13) since it is perfectly consistent with Itô's formulation of forward integration of the diffusion stochastic integral equation

$$X(t) = X(0) + \int_0^t \left(f(X(s), s)ds + g(X(s), s)dW(s)\right), \tag{9.19}$$

corresponding to the diffusion SDE (9.1). Here, only a formal applied mathematical derivation is given, since comprehensive details fill the large volume of Kloeden and Platen [165]. Clues about where to start are the fact that Euler's method has a theoretical log-rate of $\gamma_s = 0.5$ for strong convergence ([165]) and that the same power obtained for just the expectation of absolute value of the standard diffusion process, $\mathrm{E}[|\Delta W_k|] = \mathrm{O}(\sqrt{\mathrm{E}[\Delta W_k^2]}) = \mathrm{O}(\sqrt{\Delta t})$ as given in Table 1.1 on page 7 of Chapter 1. The main idea of expanding the simulation approximation is to expand the coefficient $g(x,t)$ of the term whose expected absolute value would give rise to the $\mathrm{O}(\sqrt{\Delta t})$ convergence. A way to do this is to apply iterations with Itô's stochastic chain rule in integral of $g(X(t),t)$ on $[t_k, t]$, $t \geq t_k$,

$$\begin{array}{rl} g(X(t),t) = & g(X_k,t_k) + \int_{t_k}^t \left(\left(g_t + fg_x + 0.5g^2 g_{xx}\right)(X(s),s)ds \right. \\ & \left. + (gg_x)(X(s),s)dW(s)\right), \end{array} \tag{9.20}$$

loosely upgrading the $g(x,t)$ requirements needed to twice continuously differential and where **wholesale arguments** have been used, e.g., $(gg_x)(x,t) = g(x,t)g_x(x,t)$.

This stochastic Taylor technique is also called an **Itô-Taylor expansion**. It can be used recursively to obtain very high order approximations, but here just (9.20) is substituted into a version of (9.19) rewritten for $[t_k, t_{k+1}]$,

$$\begin{array}{rcl} X_{k+1} = & X_k + \int_{t_k}^{t_{k+1}} \left(f(X(t),t)dt + g(X(t),t)dW(t)\right) & \\ = & X_k + \int_{t_k}^{t_{k+1}} \left(f(X(t),t)dt + (g(X_k,t_k)\right. & \\ & + \int_{t_k}^t \left(\left(g_t + fg_x + 0.5gg_{xx}\right)(X(s),s)ds\right. & \\ & \left.\left. + (gg_x)(X(s),s)dW(s)\right)\right) dW(t)) & \\ \simeq & X_k + F_k\Delta t + G_k\Delta W_k + G_kGX_k \int_{t_k}^{t_{k+1}} \int_{t_k}^t dW(s)dW(t), & \end{array} \tag{9.21}$$

where $GX_k \equiv g_x(X_k, t_k)$. Next, using the Itô forward integration approximation on coefficient terms and the negligibility of the residual double integral,

$$\int_{t_k}^{t_{k+1}} \int_{t_k}^{t} ds dW(t) = \int_{t_k}^{t_{k+1}} (t - t_k) dW(t) = \int_0^{\Delta t} t dW(t) \stackrel{\text{dt}}{=} 0,$$

by Itô mean square rules in dt-precision, which justifies dropping the corresponding terms. The retained double integral is just another form of Itô's fundamental Theorem 2.30 on page 41,

$$\begin{aligned} \textstyle\int_{t_k}^{t_{k+1}} \int_{t_k}^{t} dW(s) dW(t) = & \int_{t_k}^{t_{k+1}} (W(t) - W_k) dW(t) \\ = & \left(\textstyle\int_0^{\Delta t} W(t) dW(t) \right)_k \stackrel{\text{dt}}{=} 0.5 \cdot (\Delta w_k^2 - \Delta t). \end{aligned} \tag{9.22}$$

Thus, **Milstein's approximate method** is the stochastic difference equation (SΔE),

$$X_{k+1} = X_k + F_k \Delta t + G_k \Delta W_k + 0.5 G_k G X_k \cdot (\Delta W_k^2 - \Delta t), \tag{9.23}$$

for the SDE (9.1) and $k = 0 : N_t - 1$, where $F_k \equiv f(X_k, t_k)$, $G_k \equiv g(X_k, t_k)$, $GX_k \equiv g_x(X_k, t_k)$ and $\Delta W_k \equiv W(t_{k+1}) - W(t_k)$. Using the linear, time-dependent SDE model (9.14) as in Fig. 9.2 and the same fine-coarse grid numerical procedure, the Milstein and exact simulations are displayed in Fig. 9.4. The difference is very slight and hardly noticeable and the error between the Milstein approximate path and the exact path at the coarse time points is presented in Fig. 9.5. Finally, Fig. 9.6 illustrates the direct difference between the Milstein and Euler-Maruyama approximations. (For the sample code used to generate these Milstein figures, see Sect. C.17 of the Appendix.)

The Milstein algorithm converges strongly with log-rate $\gamma_s = 1$, but for the proof and computational justification see Kloeden and Platen [165, Sections 10.3 and 10.6]. Also see D. Higham's very accessible tutorial review [139] for computational justification and a nice Milstein-strong MATLAB™ code. Maple™ and MATLAB™ codes for diffusion SDEs for finance can be given in D. Higham and Kloeden [143] along with higher order approximations. Other diffusion Maple™ codes are found in Cyganowski, Kloeden and Ombach [66]. *Mathematica*™ diffusion SDE codes are presented in Stojanovic [259].

However, note that the diffusion factor $0.5(\Delta W_k^2 - \Delta t)$ in the Milstein approximation has the mean $\mathrm{E}[0.5(\Delta W_k^2 - \Delta t)] = 0$ and variance

$$\mathrm{Var}[0.5(\Delta W_k^2 - \Delta t)] = 0.25(\mathrm{E}[\Delta W_k^4] - (\Delta t)^2) = 0.5(\Delta t)^2,$$

which normally would be negligible in dt-precision. Using Table 1.1 on page 7, indicates limited correction possibilities.

9.1.4 Convergence and Stability of Jump-Diffusion Euler Simulations

The stochastic Euler's method for jump-diffusions governed by the SDE (9.1) with discrete Poisson jumps at mark-independent amplitudes $h(x, t)$, i.e.,

$$dX(t) = f(X(t), t)dt + g(X(t), t)dW(t) + h(X(t), t)dP(t),$$

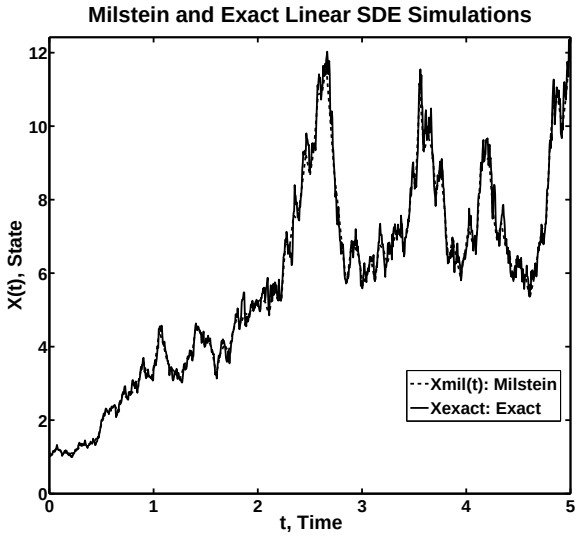


Figure 9.4. *Comparison of coarse Milstein and fine exact paths, simulated using MATLAB with* $N_t = 1024$ *fine sample points for the exact path (9.15) and* $N_t/8 = 128$ *coarse points for the Milstein path (9.23), initial time* $t_0 = 0$*, final time* $t_f = 5$ *and initial state* $x_0 = 1.0$ *as in Fig. 9.2. Time-dependent parameter values are* $\mu(t) = 0.5/(1 + 0.5t)^2$ *and* $\sigma(t) = 0.5$*.*

in its simplest form using the forward integral approximation of Itô for fixed Δt is

$$X_{k+1} = X_k + F_k \Delta t + G_k \Delta t + H_k \Delta P_k, \tag{9.24}$$

where $(F_k, G_k, H_k) = (f(X_k, t_k), g(X_k, t_k), h(X_k, t_k))$, $\Delta t = t_{k+1} - t_k$, $\Delta W_k = W_{k+1} - W_k$, and $\Delta P_k = P_{k+1} - P_k$, for $k = 0 : N_t$. Maghsoodi [191] and also Maghsoodi and Harris [192] derived most of the theory behind this method and derived numerous Milstein-like higher order approximations, so sometimes (9.24) is called the **Euler-Maghsoodi method**.

Linear Jump-Diffusion Euler Method Convergence

Following the stochastic diffusion Euler analysis for the linear, constant coefficient case,

$$dX(t) = X(t)(\mu_0 dt + \sigma_0 dW(t) + \nu_0 dP(t)),$$

the discrete Euler is written

$$X_k = B_{k-1} \cdot X_{k-1}; \quad B_k \equiv (1 + (\mu_0 + \lambda_0 \nu_0)\Delta t + \sigma_0 \Delta W_k + \nu_0(\Delta P_k - \lambda_0 \Delta t)),$$

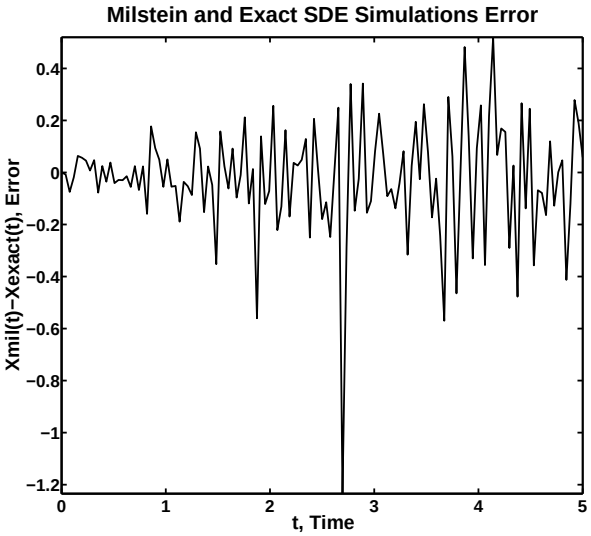


Figure 9.5. *Error in coarse Milstein and fine exact paths using the coarse discrete time points. The simulations use MATLAB with the same values and time-dependent coefficients as in Fig. 9.2. The Milstein maximal-absolute error for this example is* 1.2*, while for* $N_t = 4096$ *the maximal error is better at* 0.95*.*

where the discrete Poisson process is written in mean-zero (i.e., martingale) independent increment form for convenience, so that

$$X_k = x_0 \prod_{\ell=0}^{k-1} B_\ell,$$

and by independent increments as well as independent jump-diffusion processes,

$$\mathrm{E}[X_k \mid X_0 = x_0] = x_0 \prod_{\ell=0}^{k-1} \mathrm{E}[B_\ell] = x_0 \prod_{\ell=0}^{k-1} (1 + (\mu_0 + \lambda_0\nu_0)\Delta t) = x_0(1 + (\mu_0 + \lambda_0\nu_0)\Delta t)^k.$$

From the exact solution (9.11) using the expectation in (4.81), the final mean square at $t_f = N_t\Delta t$ is

$$\mathrm{E}\left[X(t_f) \mid X(0) = x_0\right] = x_0 e^{(\mu_0 + \lambda_0\nu_0)t_f}.$$

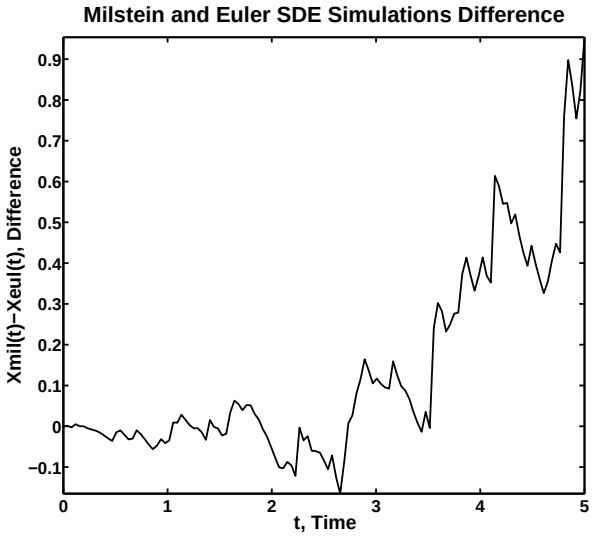


Figure 9.6. *Difference in coarse Milstein and Euler paths using the coarse discrete time points. The simulations use MATLAB with the same values and time-dependent coefficients as in Fig. 9.2. The Milstein-Euler maximal-absolute difference for this example is* 0.19*, while for* $N_t = 4096$ *the maximal difference is comparable at* 0.24*.*

Next, computing the convergence criteria in the weak sense asymptotically,

$$\begin{aligned}
&|\mathrm{E}[X_{Nt} \mid X_0 = x_0] - \mathrm{E}[X(t_f) \mid X_0 = x_0]| \\
&\qquad = |x_0| \left| (1 + (\mu_0 + \lambda_0\nu_0)\Delta t)^{N_t} - e^{(\mu_0 + \lambda_0\nu_0)N_t\Delta t} \right| \\
&\qquad = |x_0| e^{(\mu_0 + \lambda_0\nu_0)t_f} \left| e^{N_t \ln(\mu_0+\lambda_0\nu_0)\Delta t) - (\mu_0+\lambda_0\nu_0)N_t\Delta t} - 1 \right| \\
&\qquad \sim |x_0| e^{(\mu_0 + \lambda_0\nu_0)t_f} \left| e^{-0.5N_t(\mu_0+\lambda_0\nu_0)^2\Delta t^2} - 1 \right| \\
&\qquad \sim C_w \Delta t,
\end{aligned}$$

where $C_w = |x_0|(\mu_0 + \lambda_0\nu_0)^2 t_f \exp(\mu_0 + \lambda_0\nu_0)t_f)$ and the convergence in the weak sense is order one in Δt with $\gamma_w = 1$.

The distributed jump case is somewhat similar, except the marks introduce much more complications. Let the linear distributed jump-diffusion SDE have constant coefficients except that the relative jump amplitude depends on the random mark Q and the symbolic product $\nu(Q)dP(t;Q)$ is replaced by the proper jump

sum. So

$$dX(t) = X(t)\left(\mu_0 dt + \sigma_0 dW(t) + \sum_{\ell=1}^{dP(t;Q)} \nu(Q_\ell)\right),$$

and the discrete Euler processes are written in zero mean form,

$$\begin{aligned} X_k =&\ \beta_{k-1} \cdot X_{k-1}; \\ \beta_k \equiv&\ 1 + (\mu_0 + \lambda_0 \mathrm{E}[\nu(Q)])\Delta t + \sigma_0 \Delta W_k + \mathrm{E}[\nu(Q)](\Delta P_k - \lambda_0 \Delta t) \\ &+ \textstyle\sum_{\ell=1}^{\Delta P_k} (\nu(Q_\ell) - \mathrm{E}[\nu(Q)]) \end{aligned}$$

The exact solution at node t_k upon using the stochastic chain rule and integrating is

$$X(t_k) = x_0 \exp\left((\mu_0 - \sigma_0^2/2)t_k + \sigma_0 W_k + \sum_{\ell=1}^{P_k} Q_\ell\right),$$

where we have again set $Q \equiv \ln(1+\nu(Q))$ or $\nu(Q) = \exp(Q) - 1$ for convenience of setting the mark distribution appropriate for the log-process. Using the iterated expectations technique to nest the Poisson and jump mark expectations, the expectations are

$$\begin{aligned} \mathrm{E}[X_k \mid X_0 = x_0] =&\ x_0 \mathrm{E}\left[\textstyle\prod_{j=0}^{k-1} \beta_j\right] = x_0 \textstyle\prod_{j=0}^{k-1} \mathrm{E}[\beta_j] \\ =&\ x_0 \textstyle\prod_{j=0}^{k-1} (1 + (\mu_0 + \lambda_0(\mathrm{E}[\exp(Q)] - 1))\Delta t) \\ =&\ x_0 (1 + (\mu_0 + \lambda_0(\mathrm{E}[\exp(Q)] - 1))\Delta t)^k \end{aligned}$$

for the approximation and

$$\begin{aligned} \mathrm{E}[X(t_k) \mid X(0) = x_0] =&\ x_0 \exp((\mu_0 - \sigma_0^2/2)t_k)\mathrm{E}_{W_k}[\exp(\sigma_0 W_k)] \\ &\cdot \mathrm{E}_{P_k}\Big[\mathrm{E}_Q\Big[\exp\left(\textstyle\sum_{\ell=1}^{P_k} Q_\ell\right)\Big| P_k\Big]\Big] \\ =&\ x_0 \exp((\mu_0 + \lambda_0(\mathrm{E}[\exp(Q)] - 1))t_k), \end{aligned}$$

for the exact. Again, asymptotic results are derived for weak absolute mean error as $\Delta t \to 0^+$ for fixed t_k,

$$\begin{aligned} &|\mathrm{E}[X_k \mid X_0 = x_0] - \mathrm{E}[X(t_k) \mid X(0) = x_0]| \\ &\quad = |x_0| \cdot \left| e^{-k \ln(1 + (\mu_0 + \lambda_0(\mathrm{E}[\exp(Q)] - 1))\Delta t} - e^{(\mu_0 + \lambda_0(\mathrm{E}[\exp(Q)] - 1))t_k} \right| \\ &\quad \sim |x_0| e^{(\mu_0 + \lambda_0(\mathrm{E}[\exp(Q)] - 1))t_k} \left| e^{-0.5k(\mu_0 + \lambda_0(\mathrm{E}[\exp(Q)] - 1))^2 \Delta t^2} - 1 \right| \sim C_w \Delta t, \end{aligned}$$

where

$$C_w = 0.5|x_0| t_k (\mu_0 + \lambda_0(\mathrm{E}[\exp(Q)] - 1))^2 e^{(\mu_0 + \lambda_0(\mathrm{E}[\exp(Q)] - 1))t_k}.$$

Again the weak convergence rate is linear with $\gamma_w = 1$.

Maghsoodi [191] shows that the strong mean square error convergence rate (9.6) is $\mathrm{O}(\Delta t)$ for the jump-diffusion Euler method for nonlinear coefficients subject to linear Lipschitz bounds, which translates into a strong root mean square rate of

$$\mathrm{O}\left(\sqrt{\Delta t}\right).$$

A similar result was shown by D. Higham and Kloeden [145] for the implicit jump-diffusion or **stochastic theta method (STM)** with the mean square error based upon piece-wise-constant interpolation functions rather than the discrete approximate and exact values themselves ($\theta = 0$ is the explicit, stochastic Euler method, while the theta method is implicit for $0 < \theta \leq 1$). For the jump-diffusion problem, the theta method only applies to the drift term in (9.24),

$$X_{k+1} = X_k + ((1-\theta)F_k + \theta F_{k+1}) + G_k \Delta W_k + H_k \Delta P_k, \tag{9.25}$$

in order to preserve stochastic consistency with jump-diffusion with the jump-diffusion conditional infinitesimal moments (8.46,8.65), by avoiding implicit, backward steps in the diffusion and jump terms. The technical details of STM are beyond the scope of this chapter.

Euler Mean Square Linear Asymptotic Stability for Jump-Diffusions

For the mean square asymptotic stability of the jump-diffusion Euler method, the procedure leading up to the corresponding diffusion critical condition (9.18) is used. Starting with the jump-diffusion linear system recursive form,

$$X_k = B_{k-1} \cdot X_{k-1},$$

then the mean square is

$$\begin{aligned}
\mathrm{E}[X_k^2 | X_0 = x_0] = \; & x_0^2 \mathrm{E}\left[\left(\textstyle\prod_{\ell=0}^{k-1} B_\ell\right)^2\right] = x_0^2 \textstyle\prod_{\ell=0}^{k-1} \mathrm{E}\left[B_\ell^2\right] \\
= \; & x_0^2 \textstyle\prod_{\ell=0}^{k-1} \left((1+(\mu_0 + \lambda_0\nu_0)\Delta t)^2 + (\sigma_0^2 + \lambda_0\nu_0^2)\Delta t\right) \\
= \; & x_0^2 \left((1+(\mu_0 + \lambda_0\nu_0)\Delta t)^2 + (\sigma_0^2 + \lambda_0\nu_0^2)\Delta t\right)^k.
\end{aligned}$$

Again, as $k \to \infty$, the base of the power k must be less than one since the base is non-negative, so the mean square asymptotic stability criterion for the linear, constant coefficient, jump-diffusion Euler approximation is

$$2(\mu_0 + \lambda_0\nu_0) + \sigma_0^2 + \lambda_0\nu_0^2 + (\mu_0 + \lambda_0\nu_0)^2 \Delta t < 0, \tag{9.26}$$

which means that $\mu_0 + \lambda_0\nu_0$ needs to be sufficiently negative (note that $\lambda_0 > 0$ if the jump process is to be genuine),

$$\mu_0 + \lambda_0\nu_0 < -0.5(\sigma_0^2 + \lambda_0\nu_0^2 + (\mu_0 + \lambda_0\nu_0)^2 \Delta t)$$

and when interpreted in terms of the first and second relative conditional infinitesimal moments is

$$\mathrm{E}[\Delta X_k / X_k \mid X_k \neq 0] < -0.5\mathrm{E}[(\Delta X_k / X_k)^2 \mid X_k \neq 0].$$

If we restrict our attention to when the exact solution is mean square stable, i.e., $2(\mu_0 + \lambda_0\nu_0) + \sigma_0^2 + \lambda_0\nu_0^2 < 0$ from (9.12), then (9.26) can be used to construct a constraint on the discrete time step,

$$\Delta t < 2\left|\mu_0 + \lambda_0\nu_0 + 0.5(\sigma_0^2 + \lambda_0\nu_0^2)\right| / (\mu_0 + \lambda_0\nu_0)^2 \ .$$

9.1.5 Jump-Diffusion Euler Simulation Procedures

A simple numerical procedure is given in Subsection 4.3.3 on page 115 for the linear system with discrete jump of size ν_0,

$$dX(t) = X(t)(\mu_0 dt + \sigma_0 dW(t) + \nu_0 dP(t)),$$

using MATLAB™'s normal random number generator `randn` and a small time-step zero-one Poisson-Bernoulli jump law using the acceptance-rejection method. Since this zero-one jump law uses the Δt-order asymptotic precision definition of the Poisson process there is a restriction that $\lambda\Delta t < 1$ so that the one-jump probability is positive. See Program C.14 in the Appendix C for the MATLAB code used.

However, this $\lambda\Delta t < 1$ condition can be easily rectified by just renormalizing Poisson distribution, $p_k(\lambda\Delta t) = \exp(-\lambda\Delta t)(\lambda\Delta t)^k/k!$, for a finite number of jumps $k \le j$ without expanding the $\exp(-\lambda\Delta t)$ factor in the numerator, so

$$p_k^{(j)}(\lambda\Delta t) \equiv \frac{(\lambda\Delta t)^k/k!}{\sum_{\ell=0}^{j}(\lambda\Delta t)^\ell/\ell!} \tag{9.27}$$

is valid as long as $\lambda\Delta t > 0$ and conserves probability. This is the same as if the original normalization $\exp(+\lambda\Delta t)$ were expanded by $\lambda\Delta t$ in the denominator to $j+1$ terms and the result called a Padé approximation or rational function. Form (9.27) also exactly preserves the ordering of the Poisson jump probabilities, i.e.,

$$\frac{p_{k+1}^{(j)}(\lambda\Delta t)}{p_k^{(j)}(\lambda\Delta t)} = \frac{\lambda\Delta t}{k+1}$$

as long as $k = 0\!:\!j-1$. This form can be used with the acceptance-rejection method as long as the unit interval $[0,1]$ is partitioned into $j+1$ subintervals of length $p_k^{(j)}(\lambda\Delta t)$ for $k = 0:j$ such that a random number generator like MATLAB™'s `rand` is used and if the number generated lands in the subinterval corresponding to $p_k^{(j)}(\lambda\Delta t)$, then the realized number of jumps is k. Computer experiment experience indicates that it is best not to put the small subintervals adjacent to the endpoints of $[0,1]$ due to the open interval $(0,1)$ bias of computer random generators.

Distributed Jump Linear Jump-Diffusion Euler Method

In Fig. 5.1 on page 161, the simulations for uniformly distributed marks Q on $(a,b) = (-2,+1)$ and time-dependent linear or geometric jump-diffusion SDE,

$$dX(t) = X(t)(\mu_d(t)dt + \sigma(t)dW(t) + \nu(Q)dP(t;Q)).$$

However, it is more convenient to work with the exponent of the exact solution derived by the stochastic chain rule to obtain the SDE,

$$dY(t) = d\ln(X(t)) = (\mu(t) - \sigma^2(t)/2)dt + \sigma(t)dW(t) + QdP(t;Q),$$

where the mark has been selected as $Q \equiv \ln(1+\nu(Q))$ for convenience (this would seem to preclude time-dependence in the jump amplitude $\nu(Q)$, but time can be included in the mark range $[a,b]$ or the mark density $\phi_Q(q)$. The MATLAB code C.15 is a modification of the linear jump-diffusion SDE simulator code C.14 illustrated in Fig. 4.3 for constant coefficients and discrete mark-independent jumps. The state exponent $Y(t)$ is simulated as

$$YS(k+1) = YS(k) + (\mu(k) - \sigma^2(k)/2) * DT + \sigma(k) * DW(k) + Q(k) * DP(k)\,,$$

with $t(k+1) = t0+k*DT$ for $k = 0{:}NI-1$ with $NI = 1,000$, $t0 = 0$ and $0 \le t(k) \le 2$. The incremental Poisson jump term $DP(k) = P(t(k)+DT;Q) - P(t(k);Q)$ is simulated by the MATLAB™ uniform random number generator `rand` on $(0,1)$ using the acceptance-rejection technique [230, 96] (see also Subsect. 9.2.3 on p. 278) to implement the zero-one jump law to obtain the probability of $\lambda(i)Dt$ that a jump is accepted, else a jump is rejected. The same random state (seed), but a different set of generated random samples, is used to obtain the simulations of the uniformly distributed Q on (a,b). i.e., $Q = a + (b-a) * \texttt{rand}(1,NI)$, that is used only if there is a jump event. Finally, the state itself is computed by a simple exponential inversion of the log-process as

$$X(k+1) = x0 * \exp(Y(k+1)),$$

and should be highly accurate for sufficiently small DT since this procedure based upon the exact exponent is the same procedure that is used for producing the exact simulation, say by Maghsoodi [191]. Clearly, if one has a linear SDE with constant parameter coefficients for an application, then the best strategy is to simulate the exact solution since it is available. However, if the object is just to use the linear SDE for testing a method on more general nonlinear SDEs, related perhaps by similar Lipschitz linear bounds, then simulation of the original linear SDE for $X(t)$ is recommended.

Many deterministic numerical methods are difficult to translate directly into numerical methods of diffusions or jump-diffusions due to the non-smooth or discontinuous nature of the diffusion process $W(t)$ or the jump process $P(t;Q)$, respectively. Hence, implicit methods or multistep methods (many of these are designed to reduce or eliminate the implicitness of implicit methods) have to be modified to separate the treatment of the deterministic term $(f(x,t)\Delta t)$ from that of the diffusion term $(g(x,t)\Delta W(t))$ or that of the jump term $(h(x,t)\Delta P(t;Q)$ or $h(x,t,q)\Delta P(t;Q))$. It is necessary to preserve stochastic approximation consistency with respect to the jump-diffusion conditional infinitesimal moments (8.46,8.65).

Stochastic Split-Step Backward Euler Method

One such method is a stochastic modification of the **deterministic backward Euler (DBE)** method $(X_{k+1} = X_k + f(X_{k+1}, t_{k+1})\Delta t)$ which for the jump-diffusion

problem is split into two stages by Cyganowski and Kloeden [65] and more recently by D. Higham and Kloeden [144], the first stage is just a backward Euler step, $X_{k+1}^{(\mathrm{dbe})}$, only improved by the deterministic drift and a second stage that adds diffusion and jump term improvement,

$$\begin{aligned} X_{k+1}^{(\mathrm{dbe})} &= X_k + f\left(X_{k+1}^{(\mathrm{dbe})}, t_{k+1}\right)\Delta t, \\ X_{k+1}^{(\mathrm{ssbe})} &= X_k + g\left(X_{k+1}^{(\mathrm{dbe})}, t_{k+1}\right)\Delta W_k + h\left(X_{k+1}^{(\mathrm{dbe})}, t_{k+1}\right)\Delta P_k, \end{aligned} \tag{9.28}$$

which they call a **split-step backward Euler (SSBE)**. The first stage is implicit in $X_{k+1}^{(\mathrm{dbe})}$, so enhances the stability and convergence, for which some results are given in [65, 144], but no rates of convergence. The coefficients in [144] are autonomous, but time-dependence is added here for generality. An improved refinement is also included in [65, 144] and that is using the compensated or zero-mean Poisson $\Delta P_k - \lambda_k \Delta t$, a martingale, to obtain the **compensated split-step backward Euler (CSSBE)**,

$$\begin{aligned} X_{k+1}^{(\mathrm{dbe})} &= X_k + \left(f\left(X_{k+1}^{(\mathrm{dbe})}, t_{k+1}\right) + \lambda_k h\left(X_{k+1}^{(\mathrm{dbe})}, t_{k+1}\right)\right)\Delta t, \\ X_{k+1}^{(\mathrm{cssbe})} &= X_k + g\left(X_{k+1}^{(\mathrm{dbe})}, t_{k+1}\right)\Delta W_k + h\left(X_{k+1}^{(\mathrm{dbe})}, t_{k+1}\right)(\Delta P_k - \lambda_k \Delta t), \end{aligned} \tag{9.29}$$

which provides better improvement in the first, deterministic backward Euler, stage. No computational validation is given in [65, 144]. In [146], D. Higham, Mao and Stuart show $\mathrm{O}(\Delta t)$ mean square error convergence rates for SSBE on nonlinear diffusion SDEs with coefficient functions satisfying linear Lipschitz conditions.

Maghsoodi [191] also extended the Milstein algorithm for diffusions to jump-diffusions by expanding the jump coefficient $h(x,t)$ like the diffusion coefficient $g(x,t)$ stochastic Taylor expansion. However, the new and numerous jump terms are much more complicated than the diffusion version and Cyganowski, Kloeden and Ombach [66] demonstate by computer experiment that this method works well for discrete jump problems but not for distributed (mark-dependent) jump problems, so the extension will not be discussed here.

Related convergence and stability results for discrete jump-diffusions are given by D. Higham and Kloeden in [144] for the stochastic theta method as previously mentioned in association with the STM algorithm (9.25).

Jump-Adapted Euler Method

Thus far, methods using constant time-steps $\Delta t = t_{k+1} - t_k$ or a fixed set of variable time-steps $\Delta t_{k-1} t_{k+1} - t_k$ have been discussed, such that the number of jumps of ΔP_k in $[t_k, t_{k+1}]$ have been enumerated and corresponding jump marks, if present, simulated. An alternate numerical approach, suggested by Maghsoodi [191], is to interlace the set of Poisson random jump times, T_j for $j = 1 : N_J$ such that $T_{N_J} \le t_f$, with a fixed set t_ℓ for $\ell = 0 : N_t$ to define a **jump-adaptive (JA) method** grid augmented by initial and final times, such that $\tau_0 \equiv 0 < \tau_k < \tau_{k+1} = \tau_k + \Delta\tau_k < \tau_{N^{(\mathrm{ja})}} = \tau_f$ with subintervals of length $\Delta\tau_k = \tau_{k+1} - \tau_k$ for $k = 0 : N^{(\mathrm{ja})} - 1$. One

restriction is that the mesh measure satisfies $\max_{0\le k\le N^{(\mathrm{ja})}-1}(\Delta\tau_k) \le \overline{\Delta\tau}$ where $\overline{\Delta\tau} \simeq \Delta t$ plus some leeway.

It is well known that it is the Poisson subintervals or the time to the next jump $\Delta T_j = T_{j+1} - T_j$ are independent and identically, exponentially distributed (1.24) with rate λ (unfortunately, the literature on jump-adapted method confuses the IID properties of the inter-jump times and the interdependence of the jump-times themselves). The exponentially distributed Poisson jump-time generation is given on page 14 using the logarithmic transformation of a uniform random number generator and a vector version is

```
% log-uniform exponential density:
DT=-log(rand(1,NJ))/lambda;                    (9.30)
T=cumsum(DT);
```

where `rand(1,NJ)` is MATLAB™'s $1 \times NJ$ vector random generator and `cumsum` is the cumulative sum function, assuming that the total number of jumps is known.

Let the discrete state be denoted as $X_k^{(\mathrm{ja})} \simeq X(\tau_k)$ corresponding to adapted-jump-time τ_k, so the jump-diffusion Euler method for discrete jumps is

$$X_{k+1}^{(\mathrm{ja})} = X_k^{(\mathrm{ja})} + F_k^{(\mathrm{ja})}\Delta\tau_k + G_k^{(\mathrm{ja})}\Delta W_k^{(\mathrm{ja})} + H_k^{(\mathrm{ja})}\Delta P_k^{(\mathrm{ja})}, \tag{9.31}$$

where $\Delta W_k^{(\mathrm{ja})} = W(\tau_{k+1}) - W(\tau_k)$, $\Delta P_k^{(\mathrm{ja})} = P(\tau_{k+1}) - P(\tau_k)$, $F_k^{(\mathrm{ja})} = f(X_k^{(\mathrm{ja})}, \tau_k)$; similarly for $G_k^{(\mathrm{ja})}$ and $H_k^{(\mathrm{ja})}$. Note that if τ_{k+1} coincides with a jump time T_j for some j then $\Delta P_k^{(\mathrm{ja})} = 1$, otherwise $\Delta P_k^{(\mathrm{ja})} = 0$. However, as Maghsoodi [191] warns, when analyzing something like convergence in the mean then it must be recognized that if $\tau_{k+1} = T_j$ then $\Delta W_k^{(\mathrm{ja})} = W(T_j) - W(\tau_k)$ is not statistically independent of $\Delta P_k^{(\mathrm{ja})} = P(T_j) - P(\tau_k)$, if expectations are to be calculated. A sample fragment of the code to compute $\Delta\tau_k$, $\Delta W_k^{(\mathrm{ja})}$ and $\Delta P_k^{(\mathrm{ja})}$ could be as given in Fig. 9.7. This code fragment can be patched together with the given application SDE and chosen base numerical algorithm such as the jump-diffusion Euler or split-step backward Euler, for instance.

9.2 Monte Carlo Methods

The Monte Carlo method started as a statistical sampling procedure at Los Alamos National Laboratory in 1946 from an idea of Ulam in analogy considering the probability of winning the card game of solitaire, from the idea of von Neumann for the programming neutron transport on a newly emerging electronic computer and Metropolis for computer implementation [78, 206, 208]. Without the emergence of electronic computers very few people would attempt to use large scale statistical sampling to solve large problems. One exception was the famous physicist Fermi who could calculate very fast using a mechanical calculator and had time to do big calculations because he often could not sleep, so in fact he was using a smaller scale version of the Monte Carlo method fifteen years before it had a name (for other earlier examples see Hammersley and Handscomb [104] for instance). The method

```
function jumpadapt
% Jump adaptive (JA) code fragment:
%    merged regular and jump times
Nt=10; lambda=9; t0=0; tf=1; Dt = (tf-t0)/Nt;
t = Dt*[0:Nt];  % Regular grid
DT = -log(rand)/lambda; S=DT; j=0;
while S < tf  % Get jump time grid, T(NJ)<tf
    j=j+1;
    NJ=j;
    T(j)=S; DTJ(j)=DT;
    DT = -log(rand)/lambda;  % Exponential density
    S=S+DT;
end
[tau,ktau]=sort([t T]);  % Concatenate and sort times
Dtau=tau(2:Nja)-tau(1:Nja-1);  % Concatenate and sort times
randn('state',10);
RN=randn(1,Nja-1);  % Std. normal density
DP=zeros(1,Nja-1);
for k=2:Nja
    DW(k-1)=sqrt(Dtau(k-1))*RN(k-1);  % Get DW
    if ktau(k)>Nt+1
        DP(k-1)=1;  % Get final DP
    end
end
```

Figure 9.7. *Code: Jump-adapted code fragment.*

was named for an uncle of Ulam's who had a obsession about going to gamble at Monaco, the gambling capital of Europe. In a 1949 paper of Metropolis and Ulam [208] entitled **The Monte Carlo Method**, they spelled out the basic ideas in a more or less essay form: the potential applications, the statistical approach, the independent random sampling, the frequency distributions, the law of large numbers for convergence and the asymptotic theorems for probable errors. Although von Neumann is not an author on this paper, it contains his ideas on techniques of random number generation and a hint of his acceptance and rejection method to handle general shaped domains by rejecting those samples which land outside of the domain.

A more major idea of von Neumann was the logical structure of most modern programmable computers, the **von Neumann computer**. The newly emerging electronic computer mentioned was the ENIAC, a very primitive, nonprogrammable and decidedly non-von computer as non-von Neumann computers are called. Not too long afterward, there was a parallel effort at both Princeton with von Neumann and at Los Alamos with Metropolis to build a von Neumann computer, but Metropolis was able to get the Los Alamos computer named MANIAC working first.

As it is with most computer advances, faster computers do not save the user time because the user will bring a bigger problem that will take about the same amount of time as the previous problem. The user who thought of the larger Monte Carlo problem to bring to the MANIAC was the physicist Teller and the problem was calculating the equation of state of an ideal rigid sphere gas. However, the major contribution of the resulting 1953 paper by Metropolis, the Rosenbluths and the Tellers [207] was the use of **weighted sampling**, now called **importance sampling**, by using the exponential distribution of the energy change as the weight. This version of the Monte Carlo method is called the **Metropolis algorithm** [70] and was selected a one of **ten top algorithms of the century** [68, 23]. This may be confusing, because the basic Monte Carlo algorithm is sometimes called the **Metropolis algorithm** too. The 1953 paper [207] contains significantly more detail than the 1949 paper [208], in both cases Metropolis is the lead author and some would say the lead Monte Carlo computation teacher. The title of the 1953 paper is **Equation of State Calculation by Fast Computing Machines** and the quoted cycle time of the MANIAC translated to 5.6 mHz, i.e., 5.6e-3 cycles per second, which would be extremely slow compared to todays 2GHz to 4GHz PCs or 2.0e+9 to 4.0e+9 cycles per second, not fast at all.

For general references on the Monte Carlo method, see the classic monograph of Hammersley and Handscomb [104] or the more recent book of Kalos and Whitlock, 1986 [157]. Much of the more recent advances have come from applications of the Monte Carlo method to finance, so for general references on Monte Carlo with application to finance see Glasserman [96] and Jäckel [150]. For the pioneering and award winning paper on application of the Monte Carlo method to financial options see Boyle [38] or for a two decade update see Boyle, Broadie and Glasserman [39].

9.2.1 Basic Monte Carlo Simulations

The benefits of Monte Carlo are only realized in high dimensions and for functionals of stochastic processes with simulation complexities beyond direct simulations of SDEs as covered in the previous section or for deterministic problems such as physical diffusions whose solutions can be simulated by Monte Carlo. Many problems can be transformed into an integral form or integral functional such as

$$\mathrm{I}[F] = \int_{\mathcal{V}} F(\mathbf{x}) d\mathbf{x}, \tag{9.32}$$

where $\mathbf{x} = [x_i]_{n_x \times 1}$ is a n_x-dimensional vector on volume $\mathcal{V}$ and $F(\mathbf{x})$ is a bounded, integrable scalar-valued function on $\mathcal{V}$. For instance, if $\mathcal{V}$ is finite then $\mathrm{I}[F]$ could be interpreted in terms if the expectation

$$\mathrm{I}[F] = V \cdot \mathrm{E}_{\mathbf{X}}[F(\mathbf{X})]$$

of F with respect to uniform variates $\mathbf{X}$ such that $V \equiv \int_{\mathcal{V}} d\mathbf{x} < \infty$ with uniform density $\phi_{\mathbf{X}}(\mathbf{x}) = 1/V$ on domain $\mathcal{V}$.

In general (9.32) can be interpreted to include nonuniform distributions by scaling F by a suitable density $\phi_{\mathbf{X}}(\mathbf{x})$ for variates $\mathbf{X}$ on $\mathcal{V}$ so that

$$F(\mathbf{x}) = f(\mathbf{x})\phi_{\mathbf{X}}(\mathbf{x}),$$

$$\int_{\mathcal{V}} \phi_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = 1$$

and

$$\mathrm{I}[F] = \mathrm{E}_{\mathbf{X}}[f(\mathbf{X})] = \int_{\mathcal{V}} f(\mathbf{x}) \phi_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \tag{9.33}$$

The general rule for the selection of the density $\phi_{\mathbf{X}}(\mathbf{x})$ is that it capture important characteristics, such as variability, of the integrand $F(\mathbf{x})$ on domain $\mathcal{V}$ such that the function $f(\mathbf{X})$ is bounded and not very variable. The density $\phi_{\mathbf{X}}(\mathbf{x})$ should be known and the generation of its variates should be computable with reasonable effort. In the uniform case, $\phi_{\mathbf{X}}(\mathbf{x}) = 1/V$ and $f(\mathbf{X}) = V \cdot F(\mathbf{X})$.

Example 9.3. ***Risk-Neutral European Call Option Pricing:***
An example of a complex functional is the risk-neutral European call option pricing model of Zhu and Hanson [290] using a jump-diffusion SDE with log-uniformly distributed jump-amplitude marks,

$$\mathcal{C}(S_0, t_f) = \mathrm{E}_{\widetilde{\mathcal{P}}(t_f)}\Big[\mathcal{C}^{(BS)}\Big(S_0 e^{\widetilde{\mathcal{P}}(t_f) - \lambda \bar{J} t_f}, t_f\Big)\Big] \tag{9.34}$$

where

$$\widetilde{\mathcal{P}}(t_f) = \sum_{i=1}^{P(t_f;Q)} Q_i \tag{9.35}$$

is the compound Poisson jump process cumulative sum at the strike time t_f with uniformly distributed IID random marks Q_i on $[a, b]$, mean jump-amplitude

$$\bar{J} \equiv \mathrm{E}_Q[J(Q)] \equiv \mathrm{E}[\exp(Q) - 1] = (\exp(b) - \exp(a)/(b - a) - 1 \tag{9.36}$$

and Black-Scholes call option price

$$\mathcal{C}^{(BS)}(s, t_f) \equiv s\Phi(d_1(s)) - Ke^{-rt_f}\Phi(d_2(s)), \tag{9.37}$$

with strike price K, interest rate r, diffusive volatility σ, standardized normal distribution function $\Phi(x)$ and Black-Scholes argument functions $d_1(s) \equiv (\ln(s/K) + (r + \sigma^2/2)t_f)/(\sigma\sqrt{t_f})$ and $d_2(s) \equiv d_1(s) - \sigma\sqrt{t_f}$. Refer to [290] for the transformations used to achieve this form, which one would not attempt to evaluate directly but would try to estimate the call option price.

Returning to the general integral functional problem (9.33), an estimate $\widehat{\mathrm{I}}_n$ of the value of the integral $\mathrm{I}[F] = \mathrm{E}_{\mathbf{X}}[f(\mathbf{X})$ is the **sample mean** s_n of n independent, identically distributed sample points $\mathbf{X}$ distributed on $\mathcal{V}$ corresponding to the density $\phi_{\mathbf{X}}(\mathbf{x})$,

$$\widehat{\mathrm{I}}_n = s_n, \tag{9.38}$$

where the sample mean s_n or **Monte Carlo Estimator** $\widehat{\mu}_n = s_n$ is

$$\widehat{\mu}_n \equiv s_n = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{X}_i) \equiv \frac{1}{n}\sum_{i=1}^{n} f_i\,, \tag{9.39}$$

the estimate of the mean of f with respect to $\phi_{\mathbf{X}}(\mathbf{x})$. Obviously, the function $f(\mathbf{x})$ must be bounded for the sample mean to exist. The true mean of f is

$$\mu_f = \mathrm{E}_{\mathbf{X}}[f(\mathbf{X})] = \int_{\mathcal{V}} f(\mathbf{x})\phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x}.$$

Then, the estimate $\widehat{\mu}_n$ is an **unbiased estimate**, since the **bias** of the estimator from the true mean is zero, i.e.,

$$\beta_{\widehat{\mu}_n} \equiv \mathrm{E}_{\mathbf{X}}[\widehat{\mu}_n - \mu_f] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}_{\mathbf{X}}[f(\mathbf{X}_i)] - \mu_f = \mathrm{E}_{\mathbf{X}}[f(\mathbf{X})] - \mu_f = 0, \tag{9.40}$$

using the IID property of the sample points. Further, by the **strong law of large numbers (SLLN)** (B.117),

$$\widehat{\mu}_n \longrightarrow \mu_f \text{ with probability one as } n \to +\infty.$$

The true variance of f is

$$\sigma_f^2 = \mathrm{Var}_{\mathbf{X}}[f(\mathbf{X})] = \int_{\mathcal{V}} (f(\mathbf{x}) - \mu_f)^2\phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$$

and so the unbiased estimate of the sample variance from (B.111) is

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \widehat{\mu}_n)^2. \tag{9.41}$$

Example 9.4. ***Choice of Monte Carlo Sampling Distribution:***
A rule of thumb is that, while many other distributions may work in generating Monte Carlo estimations, the better density captures more variability of $F(\mathbf{x})$ along with the domain $\mathcal{V}$ and leaves a less variable $f(\mathbf{x})$ to simulate. Thus, the better choice will be the the better Monte Carlo results.

It is general numerical practice to choose an integrand weight function that captures most of the variability and can easily be integrated exactly so that the remaining integrand factor can be discretely and well approximated. For example the truncated normal distribution,

$$I = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx, \tag{9.42}$$

can be Monte Carlo estimated using a uniform (u) density $\phi^{(u)}(x) = 1/(b-a)$ on $[a,b]$ with sampled function

$$f^{(u)}(x) = (b-a)\exp(-x^2/2)/\sqrt{2\pi}$$

or a normal (n) *density* $\phi^{(n)}(x) = \exp(-x^2/2)/\sqrt{2\pi}$ *on* $[-\infty, +\infty]$ *and*

$$f^{(n)}(x) = \mathbf{1}_{x\in[a,b]} = \{1, x \in [a,b]; 0, x \notin [a,b]\},$$

is an indicator function. The exact mean is invariant with respect to the density,

$$\mu_f^{(n)} = I = \Phi_n(a,b;0,1) = \mu_f^{(u)},$$

where $\Phi_n(x,y;0,1)$ *is the usual standard normal distribution in this book on* $[x,y]$ *However, it is obvious that the exact variance assuming a normal density factor will be much smaller than the exact variance assuming a uniform density factor and a highly variable* $f(x)$*, if* a *and* b *are not small. In fact,* $\left(\sigma_f^{(n)}\right)^2 = I - I^2$ *for the normal case since* $\mathbf{1}^2_{x\in[a,b]} = \mathbf{1}_{x\in[a,b]}$ *and* $\left(\sigma_f^{(u)}\right)^2 = (b-a)\Phi_n(\sqrt{2}a, \sqrt{2}b; 0, 1)/(2\sqrt{\pi}) - I^2$ *for the uniform by transformation* $\mathrm{E}^{(u)}[(f^{(u)})^2(x)]$ *to the standard normal distribution. As* $a \to -\infty$ *and* $b \to +\infty$*, the standard normal distributions* $\Phi_n \to 1$ *in uniform as well as in normal cases and the difference has the unbounded asymptotic approximation,*

$$\left(\sigma_f^{(u)}\right)^2 - \left(\sigma_f^{(n)}\right)^2 \sim \frac{b-a}{2\sqrt{\pi}} - 1,$$

demonstrating in this extreme case that the choice of the sampling density $\phi_{\mathbf{X}}(\mathbf{x})$ *can make a big difference in the variance* σ_f^2*. A companion computational demonstration code C.18 for this problem when* $[a,b] = [-R,R]$ *is given on p. C29 of Appendix A. Of course, one would not use the uniform distribution on an infinite domain.*

Convergence of Scaled Monte Carlo Esimate Distribution to a Normal Distribution

By the **central limit theorem** (B.118) the **sample mean converges in distribution to a normal distribution**,

$$\mathrm{Prob}\left[\frac{\widehat{\mu}_n - \mu_f}{\sigma_f/\sqrt{n}} \le \xi\right] \longrightarrow \Phi_n(\xi;0,1) \;\text{ as }\; n \to +\infty\,, \tag{9.43}$$

or alternately we say $(\widehat{\mu}_n - \mu_f)/(\sigma_f/\sqrt{n}) \xrightarrow{\text{dist}} \xi$, distributed according to $\Phi_n(\xi;0,1)$, where $\Phi_n(\xi;0,1)$ is the standard normal distribution defined in (B.1.4) and $\sigma_f/\sqrt{n}$ is called the **standard error** or **probable error**. However, this form of the standard error is not too useful since neither σ_f or μ_f are known, else a Monte Carlo approximation would not be neeeded, but $\widehat{\sigma}_n^2$ is an unbiased estimator of σ_f^2 and therefore $\widehat{\sigma}_n^2$ must converge to σ_f^2 in distribution too and thus σ_f will be replaced by $\widehat{\sigma}_n$ relying on continuous extensions of the central limit theorem [150]. However, in general $\widehat{\sigma}_n$ is not necessarily an unbiased estimate of σ_f, since a function of an unbiased estimator of a parameter is not the unbiased estimate of the function of the parameter, as pointed out by Hammersley and Handscomb [104].

Monte Carlo Estimate Confidence Intervals

Following Glasserman's [96] arguments for confidence intervals with variations, the convergence in distribution (9.43) implies as $n \to +\infty$,

$$\text{Prob}\left[\widehat{\mu}_n - \mu_f \leq \frac{\widehat{\sigma}_n}{\sqrt{n}}\xi\right] \sim \Phi_n(\xi; 0, 1),$$

so replacing ξ by $-\xi$,

$$\text{Prob}\left[\widehat{\mu}_n - \mu_f \leq -\frac{\widehat{\sigma}_n}{\sqrt{n}}\xi\right] \sim \Phi_n(-\xi; 0, 1)$$

and consequently we have an asymptotic formula for confidence intervals about the true mean μ_f,

$$\text{Prob}\left[-\frac{\widehat{\sigma}_n}{\sqrt{n}}\xi \leq \widehat{\mu}_n - \mu_f \leq \frac{\widehat{\sigma}_n}{\sqrt{n}}\xi\right] \sim \Phi_n(\xi; 0, 1) - \Phi_n(-\xi; 0, 1) = 2\Phi_n(\xi; 0, 1) - 1.$$

Putting this in a more useful form, let $\delta > 0$ and $\xi = \xi(\delta)$ such that $2\Phi_n(\xi(\delta); 0, 1) - 1 = 1 - \delta$ or

$$\Phi_n(\xi(\delta); 0, 1) = 1 - \delta/2 \tag{9.44}$$

to simplify the inversion. Thus, a practical, asymptotic **confidence level** $1 - \delta$ or $100(1 - \delta)\%$ is given by the probability

$$\text{Prob}\left[\widehat{\mu}_n - \frac{\widehat{\sigma}_n}{\sqrt{n}}\xi(\delta) \leq \mu_f \leq \widehat{\mu}_n + \frac{\widehat{\sigma}_n}{\sqrt{n}}\xi(\delta)\right] \sim 1 - \delta, \tag{9.45}$$

that the true mean μ_f is in the **confidence interval**

$$\left(\widehat{\mu}_n - \frac{\widehat{\sigma}_n\xi(\delta)}{\sqrt{n}}, \widehat{\mu}_n + \frac{\widehat{\sigma}_n\xi(\delta)}{\sqrt{n}}\right).$$

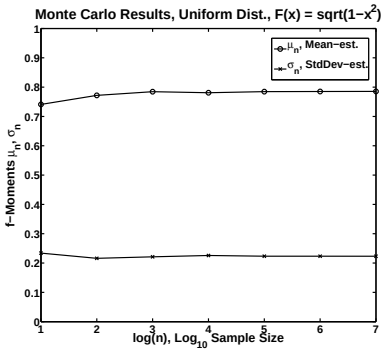
If $\xi(\delta) = 1$, the difference between the true value and the estimate is just $\pm$standard error with a confidence level of 68.27% that the simulation will be in the confidence interval, but 32.63% chance that it will be out of it. If the difference is $\pm 2\cdot$standard error then the level is 95.45%, but only a 4.55% "**lack of confidence**" level. Anyway, it will be assumed that the probable error of the Monte Carlo estimator

$$\widehat{e}_n = |\widehat{\mu}_n - \mu_f| \propto \widehat{\sigma}_n/\sqrt{n}.$$

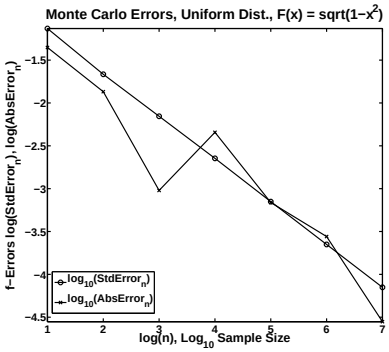
An important observation is that this probable or standard error is independent of the dimension of the volume n_x, as long as the volume is known. However, if it is necessary to approximate the volume due to its complexity, then this approximation will influence the real error.

Example 9.5. ***Convergence and Errors in Monte Carlo Estimators:*** *Monte Carlo simulations are illustrated in Fig. 9.8 using the uniform density* $\phi_X(x) =$

$1/(b-a)$ on $[a,b]$ for the one-dimensional integral of $F(x) = \sqrt{1-x^2}$ on $[a,b]$, $-1 \le a < b \le +1$, so $f(x) = (b-a) \cdot F(x)$. The computational convergence of the mean $\widehat{\mu}_n$ and standard deviation $\widehat{\sigma}_n$ estimations of $f(x)$ versus the logarithm of sample size $log_{10}(n)$ are exhibited in Subfig. 9.8(a), while the logarithm of the standard error $\log_{10}(\widehat{\sigma}_n/\sqrt{n})$ is shown versus the logarithm of the actual absolute error $\log_{10}(|\widehat{\mu}_n - \mu_f|)$, in Subfig. 9.8(a). The computational convergence is somewhat smooth from $n = 10$ to $n = 10,000,000$, but differences in the errors are more dramatic reflecting the slight variability of $\widehat{\sigma}_n/\sqrt{n}$ and the greater variability of $\widehat{\mu}_n$ compared to the constant exact value μ_f on a log-log plot.



(a) Moments of $f(x)$, $\widehat{\mu}_n$ and $\widehat{\sigma}_n/\sqrt{n}$.



(b) Logarithm of errors, $\log_{10}(\widehat{\sigma}_n/\sqrt{n})$ and $\log_{10}(|\widehat{\mu}_n) - \mu_f|)$.

Figure 9.8. *Monte Carlo simulations for testing use of the uniform distribution to approximate the integral of the integrand $F(x) = \sqrt{1-x^2}$ on $(a,b) = (0,1)$ using MATLAB code C.19 on p. C31 for $n = 10^k,\ k = 1{:}7$.*

Finite Difference Comparison

Three important characterisitics of Monte Carlo estimators, from Glasserman [96], are bias, variance and computational effort or time. For computational effort, a primary comparison is with the traditional finite difference methods.

Let the Monte Carlo target integral of (9.32) be over a unit n_x-dimensional hypercube for simplicity, i.e.,

$$\mathcal{V} \equiv [0,1]^{n_x} = [0,1] \times [0,1] \times \cdots \times [0,1];\ V = (1-0)^{n_x} = 1,$$

decomposed into a regular grid of m fixed steps $\Delta X = 1/m$ in each dimension. so that the grid points in the ith dimension are

$$X_{i,j_i} = j_i/m,\ \text{for } j_i = 0{:}m \ \text{ and } \ i = 1{:}n_x.$$

The finite difference approximation will be an expansion of the form,

$$I[F] \simeq I_m^{(\mathrm{fd})} = \sum_{j_1=1}^{m} \cdots \sum_{j_{n_x}=1}^{m} \omega_{j_1} \cdots \omega_{j_{n_x}} \cdot F(j_1/m, \ldots, j_{n_x}/m),$$

where the finite difference method weights are denoted by ω_{j_i} for $j_i = 0 : m$ and $i = 1 : n_x$, but must at least satisfy the volume conservation consistency condition that

$$\prod_{i=1}^{n_x} \sum_{j_i=1}^{m} \omega_{j_i} \cdot 1 = V = 1,$$

and the higher order the method will have even more conditions to be satisfied. There are $m+1$ grid points per dimension, so the total number of grid points will be $n_{\mathrm{fd}} = (m+1)^{n_x}$ or $m = n_{\mathrm{fd}}^{n_x} - 1$. An r**th order finite difference (fd) method** will have the following error estimate

$$e_{\mathrm{fd}} = I_m^{(\mathrm{fd})} - I[F] = \mathrm{O}\left((\Delta X)^r\right) = \mathrm{O}\left(m^{-r}\right) = \mathrm{O}\left((n_{\mathrm{fd}})^{-r/n_x}\right), \tag{9.46}$$

so for n_{fd} and r fixed,

$$e_{\mathrm{fd}} \longrightarrow \mathrm{O}\left((n_{\mathrm{fd}})^{-0}\right) = \mathrm{O}(1), \quad \text{as} \quad n_x \to \infty,$$

i.e., in the limit of high problem dimensions, finite difference methods with fixed step sizes become useless, independent of the order r of the method.

A rough theoretical comparison between the computational effort of the Monte Carlo method and fixed spaced finite difference methods (Newton-Cotes rules) can be made by assuming that the gross computational effort will be the order of the total number of points and they will be the same for both types of methods, i.e., $n_{\mathrm{fd}} = n$. Also, for a fair comparison, assume that these methods have comparable global errors, i.e., $e_{\mathrm{fd}} = \mathrm{O}(\widehat{e}_n)$ or that the orders of the errors are the same,

$$n^{-r/n_x} = 1/\sqrt{n},$$

which implies that the dimension of $\mathcal{V}$ is related to the order of the finite difference method r,

$$n_x = 2r.$$

Since the Monte Carlo method is a global method, r must be taken to be the global order of the finite difference method. For the simplest integration rule, the left or right rectangular rules (Itô's forward integration is the left rectangular rule), the global order is $r = 1$, so Monte Carlo and finite differences are comparable in computational effort and error when $n_x = 2$. For the trapezoidal or midpoint rule, $r = 2$ and $n_x = 4$ when comparable. For Simpson's (1/3) rule, $r = 4$ and $n_x = 8$ when comparable for even spacing, but for uneven grid spacing $r = 3$ since the cubic bonus due to even spacing symmetry is lost and $n_x = 6$ instead (similarly the midpoint rule order is reduced to that of the other rectangular rules). See the comments corresponding to Fig. 9.10 for comparing results from the trapezoidal and Simpson's rules with the Monte Carlo method using the rejection technique.

Monte Carlo Advantages*

- Error is theoretically independent of problem dimension, $n_x = \dim[\mathcal{V}]$.
- So, no curse of dimensionality, but best if $n_x \geq 5$ or so and several random samples are used, i.e.,

$$\left\{ X_{i,j}^{(k)} \middle| \, i = 1{:}n_x, \ j = 1{:}n \text{ sample points}, \ k = 1{:}K \text{ samples} \right\}.$$

- Works for complex integrands and domains.
- Not too sensitive to reasonable sample random number generator.

Monte Carlo Disadvantages*

- Probabilistic error bounds, not strict errors bounds that can not be exceeded, e.g., 32% of samples can exceed standard error, $\sigma_f/\sqrt{n} \simeq \widehat{\sigma}_n/\sqrt{n}$.
- Irregularity of $F(\mathbf{x})$ or $f(\mathbf{x})$ is not considered, so missed spikes or outliers possible.
- Generating many large random sample sets for high accuracy can be costly in computer and user time.
- Interplay of functions and volumes can be very complex.

Monte Carlo Ratios and Efficiencies

Any advantages* and disadvantages* are subject to testing and performance evaluation in each case. When comparing two different Monte Carlo methods, say one with the **basic Monte Carlo method** of Subsect. 9.2.1 with variance σ_1^2 and another with variance reduced to variance σ_2^2, both likely to be estimated valued, then the user should compare the methods with the **variance reduction ratio**, or simply the **variance ratio**, defined [104] as the improvement ratio from method 1 relative to method 2,

$$\mathrm{VRR}_{1,2} = \frac{\sigma_1^2}{\sigma_2^2}, \tag{9.47}$$

that is method 2 is the better variance reducer if $\mathrm{VRR}_{1,2} > 1$ and significantly larger.

However, checking for variance reduction alone is not sufficient since the computational costs of the variance reduction should not be excessive, so the **computational cost ratio**

$$\mathrm{CCR}_{1,2} = \frac{\tau_1}{\tau_2}, \tag{9.48}$$

should also be checked, where τ_1 is the computational cost (e.g., CPU time) of the first method (usually the basic Monte Carlo method) and τ_2 is the computational cost of the second method.

Hammersley and Handscomb [104] combine both the variance and computational cost ratios into the **efficiency** of method 2 relative to method 1 as

$$\text{Eff}_{1,2} = \text{VRR}_{1,2} \cdot \text{CCR}_{1,2} = \frac{\sigma_1^2 \tau_1}{\sigma_2^2 \tau_2}. \tag{9.49}$$

See also Glasserman [96, pp. 9-12] or Glynn and Whitt [97] for a more thorough description of Monte Carlo efficiency. In addition, Glasserman [96, p. 185] has observed that

> *The greatest gains in efficiency from variance reduction techniques result from exploiting specific features of a problem, rather than from generic potential variance reduction.*

In fact, two primary methods of variance reduction rely on the Monte Carlo user choosing a known factor that represents a significant proportion of the variability of the target function $f(\mathbf{x})$ or the associated density $\phi_{\mathbf{X}}(\mathbf{x})$. **Importance sampling** techniques rely on finding a multiplicative factor that is a better density than the one originally proposed. **Control variate** techniques rely on finding an known additive factor so that when the factor is subtracted from the target function the variance is significantly reduced. In any case, the selection usually depends on good user knowledge of the problem or related model problems.

9.2.2 Inverse Method for Generating Non-Uniform Variates

When there is an explicit formula for a distribution of a non-uniform variate in terms of elementary functions, then since the distribution function must lie in $[0, 1]$, an inverse of the distribution function in terms of a uniform variate can transform the non-uniform random variate so that it can be generated by a uniform random variate.

Example 9.6. ***Inversion of Exponential to Uniform Distribution:***
This was illustrated very early in Subsection B.1.7 for the exponential distribution. From (B.40), the exponential distribution for variable $x \geq 0$ and mean $\mu \geq 0$ is

$$\Phi_e(x; \mu) = 1 - \exp(-x/\mu),$$

so equating this to the uniform distribution on $[0, 1]$,

$$\Phi_u(u) = \text{Prob}[0 \leq U \leq u] = u = 1 - \exp(-x/\mu)$$

and inverting yields the inverse relation,

$$x = -\mu \cdot \ln(1 - u).$$

However, some computing effort can be saved by eliminating the floating point subtraction in the log-argument by using the complementary property of $\Phi_u(u)$ that $1 - \Phi_u(u) = 1 - u = \text{Prob}[0 \leq U \leq 1 - u]$ is also a uniform distribution for $(1 - u)$

on $[0,1]$ *(this may seem overly simple, but many students in the sciences without strong statistics background have difficulty accepting this unless it is spelled out). Thus, matching the uniform to the exponential distribution can also be formated as,*

$$\text{Prob}[0 \le U \le 1-u] = 1-u = 1-\exp(-x/\mu),$$

leading to a more efficient form for simulations,

$$x = -\mu \cdot \ln(u), \tag{9.50}$$

especially when there are a large number of simulations, $X_i = -\mu \cdot \ln(U_i)$ *for* $i = 1:n$*, e.g.,* n *=1.e+6.*

In general, if it is necessary to generate random variates from a non-uniform random variate X_i with a known distribution function $\Phi_X(x)$ but without an existing random number generator, then if $\Phi_X(x)$ is strictly increasing, $\Phi'_X(x) > 0$, and so an inverse exists,

$$U_i = \Phi_X(X_i) \iff X_i = \Phi_X^{-1}(U_i). \tag{9.51}$$

Validation that (9.51) is correct follows from the chain of equations,

$$\begin{aligned} \Phi_X(x) \equiv &\ \text{Prob}[X \le x] = \text{Prob}[\Phi_X^{-1}(U) \le x] \\ = &\ \text{Prob}[U \le \Phi_X(x)] = \text{Prob}[U \le u] \equiv \Phi_U(u), \end{aligned}$$

using the definition of a probability distribution, (9.51) for pairs (X, U) and (u, x) and the definition of the inverse. For practical purposes, this would mean that $\Phi_X(x)$ is in the form of elementary functions.

Example 9.7. ***Use of Built-in Inverses:***
In some special cases, efficient numerical inverses are available, such as the inverse for the error function or complementary error function, `erfinv` *or* `erfcinv`*, in MATLAB™, which can be used for inverting the normal distribution (if access to the Statistics Toolbox of MATLAB™ is available, then* `norminv` *builtin function can be used, but the definition* `norminv(x) = -sqrt(2)*erfcinv(2*x)` *is trivial, so the toolbox is not necessary). In Maple™, the general procedure using the* `stats[random]` *statistics subpackage is based upon its uniform random generator function with the specification of* `'inverse'` *option for non-uniform distributions by the inverse cumulative distribution function (*`'icdf'`*) method, unless a builtin function is called by name, e.g.,* `normald` *for normal distribution, or the automatic (*`'auto'`*) builtin option is specified.*

For more general cases, when either (1) the distribution $\Phi_X(X_i)$ has a flat subinterval on the interior of its range, say (c, d), i.e., there is a least one subinterval $c < x_i \le x \le x_{i+1} < d$ where $\Phi'_X(x) = 0$, or (2) the distribution has a jump in the interior of its range, i.e., there is an x_j such that $\Phi_X(x_j^+) > \Phi_X(x_j^-)$. The book of Glasserman [96, Section 2.2.1] is a good reference for these irregular cases and also a good source for many **inverse transform method** examples.

One important example for this book on jump-diffusions is the inversion of the cumulative discrete Poisson distribution with mean Λ to the continuous uniform distribution. The Poisson distribution (B.50) is written as the nth order cumulative distribution is written with a distribution recursion as

$$P(N) = \sum_{k=0}^{N} p_k(\Lambda);\ p_0(\Lambda) = 1;\ p_{k+1}(\Lambda) = \Lambda \cdot p_k(\Lambda)/(k+1).$$

Glasserman's [96] pseudo-code is translated to MATLAB™ code in Fig. 9.9 below.

```
function N = cumpois(Lambda)
% cumpois function turns uniform point into Poisson jump count;
U = rand; % generate 1 uniform random point;
% code can be changed to use vector U if needed;
pk = exp(-Lambda); % initialize Poisson distribution;
P = 0; % initialize cumulative distribution;
N = 0; % initialize cumulative jump counter;
while P < U,  % generate cumulative Poisson count;
    N = N + 1; \% step jump counter if U too small;
    pk = Lambda*pk/N; \% update Poisson distribution;
    P = P + pk; \% update cumulative distribution;
end
% End function cumpois; returns count N at mean rate Lambda;
```

Figure 9.9. *Code: Inverse Poisson method to generate jump counts using the uniform distribution [96, Fig. 3.9].*

A facsimile of the code in Fig 9.9 has been used successfully by Zhu and Hanson [290] in their Monte Carlo simulation of risk-neutral European call option pricing, cited in Example 9.3. Note that since `cumpois` takes the jump count Λ as input, `cumpois` can be used for temporal Poisson processes such as in Properties 1.21 on page 21.

If the components of a vector variate are an independent set of random variates, then it is fairly easy to invert the distribution in favor of a set of independent uniform variates since the joint distribution of independent variates is the product of component marginal distributions (Defn. B.35, p. B25), .i.e., if

$$\Phi_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n_x} \Phi_{X_i}(x_i) = \prod_{i=1}^{n_x} u_i, \tag{9.52}$$

then

$$\mathbf{x} = [x_i]_{n_x \times 1} = \left[\Phi_{X_i}^{-1}(u_i)\right]_{n_x \times 1}, \tag{9.53}$$

using the inversion transform method component by component.

For instance, if $\mathbf{X}_j$ are IID exponentially distributed random vectors with vector mean $\boldsymbol{\mu}$ and $\mathbf{U}_j$ are generated IID uniformly distributed random vectors for

each sample point $j = 1{:}n$, then the $\mathbf{X}_j$ can be generated by

$$\mathbf{X}_j = [X_{i,j}]_{n_x \times 1} = [-\mu_i \ln(U_{i,j})]_{n_x \times 1} = -\boldsymbol{\mu}. * \log(\mathbf{U}_j)\,,$$

for all j, where `.*` is the elements-wise multiplication symbol and `log(U)` is the vectorized natural logarithm function of vector $\mathbf{U}$ as in MATLAB™ (`log10` is the corresponding MATLAB™ base 10 logarithm).

9.2.3 Acceptance and Rejection Method of von Neumann

The method of acceptance and rejection is due to von Neumann [273], one of the earliest techniques introduced into the Monte Carlo method. It can be applied to produce samples for unusual probability distributions as well as for unusual domains since the method uses simpler problems that are easier to draw variates in simpler domains. In two dimensions, it is just a matter to find the proportion of points from the simpler, bounding area which lie in the more complicated, interior area.

Note that knowing the formula for a density function, $\phi_{\mathbf{X}}^{(1)}(\mathbf{x})$ on domain $\mathcal{V}$, does not mean we know how to generate random variates $\mathbf{X}_i$ for it. Let $\phi_{\mathbf{X}}^{(2)}(\mathbf{x})$ be another density function, such as a uniform, normal or exponential density function, which is simpler (else not useful), for which there is a known method for generating the corresponding random variates, $\mathbf{X}_i^{(2)}$ and suppose there is a positive constant c for the relative bound

$$\phi_{\mathbf{X}}^{(1)}(\mathbf{x}) \le c \cdot \phi_{\mathbf{X}}^{(2)}(\mathbf{x}), \tag{9.54}$$

for $\mathbf{x}$ in $\mathcal{V}$. For consistency, the target density $\phi_{\mathbf{X}}^{(1)}(\mathbf{x})$ should have a zero when the known comparison, generating density $\phi_{\mathbf{X}}^{(2)}(\mathbf{x})$ does, so the relative bound can be written

$$\frac{\phi_{\mathbf{X}}^{(1)}(\mathbf{x})}{\left(c\phi_{\mathbf{X}}^{(2)}(\mathbf{x})\right)} \le 1,$$

assuming that $0/0 \le 1$ has been defined. The unit bound indicating that a scalar uniform density will be useful. Since both are densities, the relative bound means that $1 \le c \cdot 1$ upon integrating both sides of (9.54), so $c \ge 1$ is required. The procedure for the acceptance-rejection method or technique on the ith step is

1. Generate a random variate $\mathbf{X}_i^{(2)}$ for the comparison density $\phi_{\mathbf{X}}^{(2)}(\mathbf{x})$ (e.g., this comparison density could also be a uniform density for one-dimension, in which case, `X(i) = rand`, in MATLAB™).

2. Compute the relative ordinate

$$Y_i = \frac{\phi_{\mathbf{X}}^{(1)}(\mathbf{X}_i^{(2)})}{\left(c\phi_{\mathbf{X}}^{(2)}(\mathbf{X}^{(2)})\right)}, \tag{9.55}$$

with generated $\mathbf{X}_i^{(2)}$, assuming the relative bound constant c has already been calculated.

3. Generate a scalar uniform random variate U_i and use it to accept or reject the relative ordinate Y_i, such that

   - If $U_i \leq Y_i$, then **accept** $\mathbf{X}_i^{(1)} = \mathbf{X}_i^{(2)}$ as a variate for the target density $\phi_{\mathbf{X}}^{(1)}$ and get another point $\mathbf{X}_{i+1}$, stepping i, unless $i+1 > n$.
   - Else, if $U_i > Y_i$, then **reject** the current $\mathbf{X}_i^{(2)}$ and try another ith generated variate from comparison density $\phi_{\mathbf{X}}^{(2)}$.

Note that

$$\text{Prob}\left[\mathbf{X}^{(2)} \text{ Rejected}\right] = \frac{\text{TotalArea}\left[c\phi_{\mathbf{X}}^{(2)}(\mathbf{x}) - \phi_{\mathbf{X}}^{(1)}(\mathbf{x})\right]}{\text{TotalArea}\left[c\phi_{\mathbf{X}}^{(2)}(\mathbf{x})\right]} = \frac{c-1}{c} \leq 1,$$

so, in addition, the user wants $(c-1)$ to be small and positive, i.e., c should be a tight bound constant, to reduce the amount of computation to avoid too many rejected attempts and thus increase efficiency. Also, the target distribution $\Phi_{\mathbf{X}}^{(1)}(\mathbf{x})$ for $\mathbf{X} = \mathbf{X}^{(1)}$ (vector inequalities are shorthand notation for a set of component equalities) is

$$\begin{aligned}\text{Prob}\left[\mathbf{X} \leq \mathbf{x}\right] &= \frac{\text{TotalArea}\left[\phi_{\mathbf{X}}^{(1)}(\mathbf{y}) \middle| \mathbf{y} \leq \mathbf{x}\right]}{\text{TotalArea}\left[c\phi_{\mathbf{X}}^{(2)}(\mathbf{x})\right]} + \text{Prob}\left[\mathbf{X}^{(2)} \text{ Rejected}\right] \cdot \text{Prob}\left[\mathbf{X} \leq \mathbf{x}\right] \\ &= \frac{1}{c}\Phi_{\mathbf{X}}^{(1)}(\mathbf{x}) + \frac{c-1}{c}\Phi_{\mathbf{X}}^{(1)}(\mathbf{x}) = \Phi_{\mathbf{X}}^{(1)}(\mathbf{x}),\end{aligned}$$

consistent with the definition of a distribution.

Example 9.8. ***Application of Acceptance-Rejection with Normal Distribution:***
Figure 9.10, a computational application of the acceptance-rejection technique is illustrated for the truncated normal distribution $\Phi_n(a, b; 01)$ defined for a previous uniform-normal comparison in (9.42) of Example 9.4. The computation converges nicely, with standard errors of 2.1e-4 when $n = 10^6$ sample points and 6.59e-5 when $n = 10^7$. However, when these one-dimensional results are compared to standard finite difference methods the results are not so impressive, e.g., the trapezoidal rule has an absolute error of 2.88e-05 using 101 points and Simpson's (1/3) rule has an absolute error of 3.09e-9 using the same 101 points, although, as we have said, the finite difference methods are better for low dimensions.

Example 9.9. ***Multidimensional Application of Acceptance-Rejection Technique:***
Figure 9.11 illustrates the application of Monte Carlo multidimensional simulations with the von Neumann acceptance-rejection technique similar the former $n_x = 1$ truncated normal distribution problem (9.42) in Example 9.4, but here for dimensions $n_x = 2:5$. Subfig. 9.11(a) exhibits the Monte Carlo mean estimates, $\widehat{\mu}_n$,

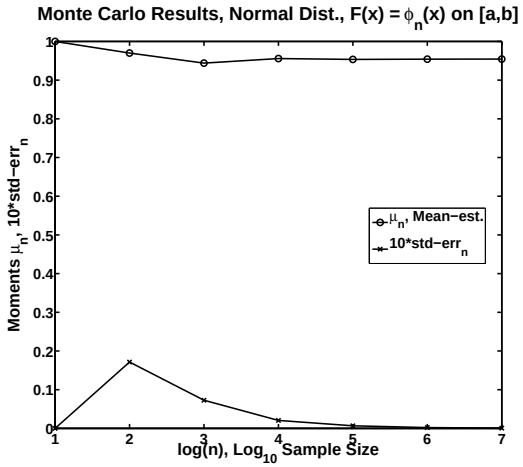


Figure 9.10. *Monte Carlo simulations shown apply the acceptance and rejection technique and the normal distribution to compute the estimates for the mean* $\widehat{\mu}_n$ *and the magnified standard error* $10 \cdot \widehat{\sigma}_n/\sqrt{n}$ *for the integral of the truncated normal distribution with* $F(x) = \phi_n(x)$ *on* $[a, b] = [-2, 2]$ *using MATLAB code C.20 on p. C33 for* $n = 10^k,\ k = 1{:}7$.

which roughly settle down by sample size $n = 10^4$*, but definitely by* $n = 10^5$ *for this problem and sample sets. In Subfig. 9.11(b), the Monte Carlo standard error estimates,* $\widehat{\sigma}_n/\sqrt{n}$ *are displayed, showing a remarkable similar decay in sample size beyond sample size* $n = 10^3$*. Note that since the integrand* $F(\mathbf{x}) = \phi_n(\mathbf{x})$ *is the normal density restricted to the vector interval* $[\mathbf{a}, \mathbf{b}]$*, the normal density scaled integrand is* $f(\mathbf{x}) = \mathbf{1}_{\mathbf{x}\in[\mathbf{a},\mathbf{b}]}$*, an indicator function for the set* $[\mathbf{a}, \mathbf{b}]$*, so* $f^2(\mathbf{x}) = f(\mathbf{x})$ *and the estimate of the standard error,*

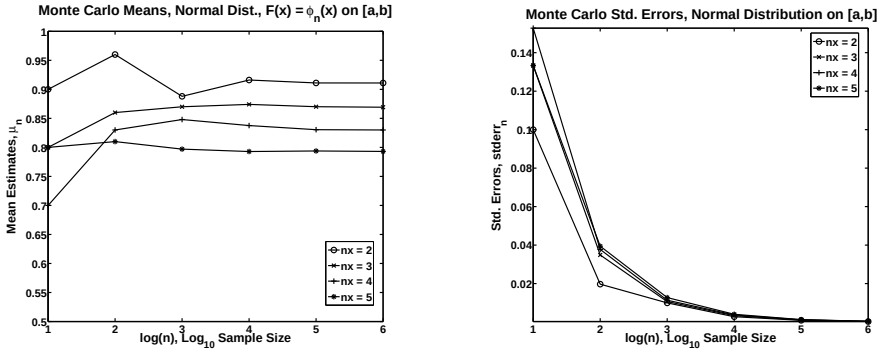
$$\widehat{\sigma}_n/\sqrt{n} = \sqrt{\widehat{\mu}_n(1 - \widehat{\mu}_n)/(n-1)},$$

satisfies the same formula regardless of dimension n_x *as suggested by the Monte Carlo theory.*

Box-Muller Algorithm for Normal Random Variates

Many of the normal random number generators, if not all, use the algorithm of Box and Muller [37] or updates of it [195] (see also [230, 96]). Since the normal distribution is a special function that cannot be put in terms of elementary functions, it is not exactly invertible by the inverse transform method, except numerically or artificially by defining another special function for the inverse. Box and Muller use pairs of uniform variates and polar coordinate to construct their algorithm to compute a pair of normal variates.

Let U_1 and U_2 be two independent uniform variates on $(0, 1)$, use them to construct a pair of polar coordinates (R, T) and then use those to construct two



(a) Mean estimates, $\widehat{\mu}_n$, for $f(\mathbf{x})$.

(b) Logarithm of standard errors, $\log_{10}(\widehat{\sigma}_n/\sqrt{n})$.

Figure 9.11. *Monte Carlo simulations for estimating multi-dimensional integrals for the n_x-dimension normal integrand $F(\mathbf{x}) = \phi_n(\mathbf{x})$ on $[\mathbf{a}, \mathbf{b}] = [-2, 2]^{n_x}$ using MATLAB code C.21 on p. C35 for $n = 10^k$, $k = 1{:}6$. The acceptance-rejection technique is used to handle the finite domain.*

independent normal variates (X_1, X_2),

$$R = \sqrt{-2\ln(U_1)} \quad \text{and} \quad T = 2\pi U_2,$$
$$X_1 = R\cos(T) \quad \text{and} \quad X_2 = R\sin(T),$$

where $0 < R < \infty$ and $0 < T < 2\pi$ since $0 < U_i < 1$ for $i = 1:2$. The inverse transformation is then $\tan(2\pi U_2) = X_2/X_1$ and $-2\ln(U_1) = X_1^2 + X_2^2$ or

$$U_1 = \exp\left(-\left(X_1^2 + X_2^2\right)/2\right) \quad \text{and} \quad U_2 = \tan^{-1}(X_2/X_1)/(2\pi).$$

The Jacobian of the transformation, $(X_1, X_2) \longrightarrow (U_1, U_2)$, is

$$\begin{aligned} J = \ & \tfrac{\partial(U_1,U_2)}{\partial(X_1,X_2)} = \mathrm{Det}\begin{bmatrix} \partial U_1/\partial X_1 & \partial U_1/\partial X_2 \\ \partial U_2/\partial X_1 & \partial U_2/\partial X_2 \end{bmatrix} \\ = \ & -\exp\left(-\left(X_1^2 + X_2^2\right)/2\right)/(2\pi) = -\Phi_n(X_1, X_2; (0,0), (1,1)), \end{aligned} \tag{9.56}$$

i.e., the negative of a standard 2-dimensional normal distribution for two independent, standard normal variates (X_1, X_2), so the objective and only $|J|$ is needed. Conservation of probability consistency is easily verified, since in theory,

$$1 = \int_0^1 \int_0^1 du_1 du_2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |J| dx_1 dx_2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-x_1^2/2} e^{-x_2^2/2} dx_1 dx_2 = 1.$$

Marsaglia and Bray [195] modified the Box-Muller algorithm save computing costs by using the acceptance-rejection technique between a square enclosing the unit circle so that the sine and cosine functions would not be needed. They begin by generating two independent uniform variates on the square $(-1, +1) \times (-1, +1)$

rather than on $(0,1) \times (0,1)$, i.e., keeping U_1 and U_2 as the initial $(0,1)$ uniform variates, $U_3 = 2 \cdot U_1 - 1$ and $U_4 = 2 \cdot U_2 - 1$. Next let the squared radius be $R_2 = U_3^2 + U_4^2$ and while $R_2 > 1$, i.e., out of the unit circle, then reject it and try again, but if $R_2 \leq 1$ then compute the normalized Box-Muller radius $R_3 = \sqrt{-2\ln(R_2)/R_2}$ and finally output the independent, standard normal variate pair,

$$X_3 = R_3 \cdot U_3 \quad \text{and} \quad X_4 = R_3 \cdot U_4.$$

9.2.4 Importance Sampling

There are two principal ways to reduce the standard error and thus improve the likely accuracy of Monte Carlo simulation relative to the basic Monte Carlo simulation (Subsection 9.2.1; Hammersley and Handscomb call the method in their 1964 compact little book [104, Section 5.2] the **crude Monte Carlo method**). One way is to increase the sample size n, but the computational cost is high, e.g., increasing the sample size 100 times is necessary to reduce the standard error by 1/10th of its magnitude due to the weak reciprocal square root order. The other way is to reduce the variance and a way to do that is to pick a better density to draw samples from that more closely matches the integrand $F(\mathbf{x})$. **Importance sampling** methods strive to find the better or practical best distribution. As previously mentioned, importance sampling was introduced into the Monte Carlo method in one of the earliest papers [207] on the subject, sometimes called the Metrhopolis algorithm, in which the desirable sampling distribution was the exponential distribution of energy changes.

Suppose there is an initial density $\phi_{\mathbf{X}}(\mathbf{x})$ with mean of $f(x)$ integral

$$\mu_f = \mathrm{E}_{\mathbf{X}}[f(\mathbf{X})] = \int_{\mathcal{V}} f(\mathbf{x})\phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = \int_{\mathcal{V}} F(\mathbf{x})d\mathbf{x}, \tag{9.57}$$

but we seek a better density $\widetilde{\phi}_{\mathbf{X}}(\mathbf{x})$ that more closely characterizes the original integrand $F(\mathbf{x})$ and leads to the equivalent formulation,

$$\widetilde{\mu}_{\widetilde{f}} = \mathrm{E}_{\widetilde{\phi}}\left[\left(f\phi/\widetilde{\phi}\right)(\mathbf{X})\right] = \int_{\mathcal{V}} \left(f\phi/\widetilde{\phi}\right)(\mathbf{x})\widetilde{\phi}_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = \mu_f, \tag{9.58}$$

where $\widetilde{f}(\mathbf{x}) \equiv \left(f\phi/\widetilde{\phi}\right)(\mathbf{x})$ is a potentially less variable sample target function and $\mathrm{E}_{\widetilde{\phi}}$ denotes an expectation with respect to the new density $\widetilde{\phi}_{\mathbf{X}}(\mathbf{x})$, subject to minimal likelihood properties that

$$\widetilde{\phi}_{\mathbf{X}}(\mathbf{x}) \geq 0 \Longleftrightarrow \phi_{\mathbf{X}}(\mathbf{x}) \geq 0,$$

mainly so that any indeterminant 0/0 form can be defined as 1. The corresponding variance is given by

$$\widetilde{\sigma}_{\widetilde{f}}^2 = \int_{\mathcal{V}} \left(\widetilde{f}(\mathbf{x}) - \widetilde{\mu}_{\widetilde{f}}\right)^2 \widetilde{\phi}_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = \int_{\mathcal{V}} \widetilde{f}^2(\mathbf{x})\widetilde{\phi}_{\mathbf{X}}(\mathbf{x})d\mathbf{x} - \widetilde{\mu}_{\widetilde{f}}^2. \tag{9.59}$$

Since the means are the same, $\widetilde{\mu}_{\widetilde{f}}^2 = \mu_f^2$, reduction of the variance is equivalent to reduction of the new second moment, i.e., making

$$\int_{\mathcal{V}} \widetilde{f}^2(\mathbf{x})\widetilde{\phi}_{\mathbf{X}}(\mathbf{x})d\mathbf{x} < \int_{\mathcal{V}} f^2(\mathbf{x})\phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x}.$$

In importance sampling the goal is to sample at important points of $\widetilde{f}(\mathbf{x})$ such as points of maximum likelihood (see Glasserman [96]).

The Monte Carlo unbiased estimates are the means for n-point samples,

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{X}_i) \;\text{ and }\; \widehat{\widetilde{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n} \widetilde{f}\left(\widetilde{\mathbf{X}}_i\right), \tag{9.60}$$

where the points $\widetilde{\mathbf{X}}_i$ are sampled from the distribution of the $\widetilde{\phi}_{\mathbf{X}}(\mathbf{x})$ density, while the unbiased sample variances are

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} \left(f(\mathbf{X}_i) - \widehat{\mu}_n\right)^2 \;\text{ and }\; \widehat{\widetilde{\sigma}}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} \left(\widetilde{f}\left(\widetilde{\mathbf{X}}_i\right) - \widehat{\widetilde{\mu}}_n\right)^2. \tag{9.61}$$

As with the exact variances or second moments, it is expected that the new second sampled moment is reduced, i.e.,

$$\frac{1}{n}\sum_{i=1}^{n} \widetilde{f}^2\left(\widetilde{\mathbf{X}}_i\right) < \frac{1}{n}\sum_{i=1}^{n} f^2(\mathbf{X}_i),$$

since this is equivalent to $\widehat{\widetilde{\sigma}}_n^2 < \widehat{\sigma}_n^2$ for sample variances or $\widehat{\widetilde{\sigma}}_n/\sqrt{n} < \widehat{\sigma}_n/\sqrt{n}$ for the Monte Carlo standard error estimates.

The best choice of a new density would obviously be the normalized absolute value the full problem,

$$\widetilde{\phi}_{\mathbf{X}}(\mathbf{x}) = |F(\mathbf{x})| \Big/ \int_{\mathcal{V}} |F(\mathbf{y})| d\mathbf{y} \;,$$

but that would be an absurd circular argument since the normalization factor in the denominator would be the integral we are seeking to estimate if it were the case that $F(\mathbf{x}) > 0$. As Glasserman [96, Fig. 4.16] states, importance sampling is the most complex of Monte Carlo techniques for reducing variance, but has the potential whose effectiveness ranges from the best to the worst. See Glasserman's [96] book for a more advanced treatment.

Analogous concepts arose long ago in the statistically related Gaussian quadrature rules [230], i.e., Gauss statistics quadrature [275], of numerical analysis. For instance, the Gauss-Legendre rules correspond to integrals weighted in proportion to a uniform density on $[-1, +1]$, Gauss-Laguerre rules to the exponential or gamma densities on $(0, \infty)$ and Gauss-Hermite to the normal distribution $(-\infty, +\infty)$. The criteria for the numerical weights w_i and nodes x_i is that the Gaussian rules give the **best polynomial precision** for the polynomial approximation to the weighted

function corresponding to the importance sampled $f(x)$. Practical criteria concern matching the rule with the domain, whether finite, semi-infinite or full-infinite, but also matching integrand singularities in the case of certain Gaussian rules not mentioned here.

There is a more advanced code like the adaptive Monte Carlo code called **VEGAS** [183] of Lepage that primarily uses importance sampling, but also uses stratified sampling discussed in the next subsection. The VEGAS algorithm and code is discussed in Numerical Recipes [230].

9.2.5 Stratified Sampling

If the integrands are very variable, then partitioning the domain into disjoint subdomains, computing Monte Carlo estimates on each subdomain and reassembling the estimates to form a global estimate can usually reduce the global estimated variance, sometimes significantly [104, 96, 150, 230].

Consider a partition of the domain $\mathcal{V}$ into np disjoint parts, called **strata**, such that the union

$$\bigcup_{k=1}^{np} \Delta\mathcal{V}_k = \mathcal{V}$$

and the Monte Carlo integral of interest (9.33)

$$\mu_f = \mathrm{E}_{\mathbf{X}}[f(\mathbf{X})] = \sum_{k=1}^{np} \int_{\Delta\mathcal{V}_k} f(\mathbf{x})\phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x} = \sum_{k=1}^{np} p_k \mu_f^{(k)}, \tag{9.62}$$

where the kth stratum probability is

$$p_k = \int_{\Delta\mathcal{V}_k} \phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x} > 0 \quad \ni \quad \sum_{k=1}^{np} p_k = 1,$$

assumed known, and the strata mean

$$\mu_f^{(k)} = \int_{\Delta\mathcal{V}_k} f(\mathbf{x})\phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x}/p_k.$$

Let $\mathbf{X}_i^{(k)}$ be a sample point drawn from the density $\phi_{\mathbf{X}}(\mathbf{x})$ on the kth strata $\Delta\mathcal{V}_k$ for $i = 1 : n_k$ where $n_k > 0$ and $\sum_{k=1}^{np} n_k = n$, the sample size. Also let $f_i^{(k)} \equiv f(\mathbf{X}_i^{(k)})$, so that the kth strata Monte Carlo estimate of the mean is

$$\widehat{\mu}_{n_k}^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} f_i^{(k)} \tag{9.63}$$

and since $\mu_f = \sum_{k=1}^{np} p_k \mu_f^{(k)}$ the total mean estimate is

$$\widehat{\mu}_{n,np} = \sum_{k=1}^{np} p_k \widehat{\mu}_{n_k}^{(k)} = \sum_{k=1}^{np} \frac{p_k}{n_k} \sum_{i=1}^{n_k} f_i^{(k)}. \tag{9.64}$$

This strata sampled estimate is an unbiased estimate since

$$\mu_f^{(k)} = \mathrm{E}^{(k)}\left[f\left(\mathbf{X}_i^{(k)}\right)\right] \equiv \mathrm{E}\left[f\left(\mathbf{X}_i^{(k)}\right)\middle|\mathbf{X}_i^{(k)} \in \Delta\mathcal{V}\right].$$

Then,

$$\mathrm{E}\left[\widehat{\mu}_{n,np}\right] = \sum_{k=1}^{np}\frac{p_k}{n_k}\sum_{i=1}^{n_k}\mathrm{E}^{(k)}\left[f_i^{(k)}\right] = \sum_{k=1}^{np}\frac{p_k}{n_k}\sum_{i=1}^{n_k}\mu_f^{(k)} = \sum_{k=1}^{np}p_k\mu_f^{(k)} = \mu_f,$$

independent of the sample distribution n_k. Note that the order of Monte Carlo estimation and stratification are not generally interchangeable if the unbiased property is to be preserved. For instance, if the original simple Monte Carlo estimate $\widehat{\mu}_n$ (9.39) is directly converted to a stratified sum,

$$\widehat{\mu}_n = \frac{1}{n}\sum_{k=1}^{np}\sum_{i=1}^{n_k}f_i^{(k)}$$

and the expectation is calculated as

$$\mathrm{E}[\widehat{\mu}_n] = \frac{1}{n}\sum_{k=1}^{np}\sum_{i=1}^{n_k}\mathrm{E}^{(k)}\left[f_i^{(k)}\right] = \frac{1}{n}\sum_{k=1}^{np}n_k\mu_f^{(k)},$$

which for general strata means $\mu_f^{(k)}$ this sum will not be μ_f. However, in the special case of **proportional strata sampling** in which $n_k = p_k \cdot n$, then

$$\mathrm{E}[\widehat{\mu}_n] = \sum_{k=1}^{np}p_k\mu_f^{(k)} = \mu_f.$$

Recall that the exact variance of f is

$$\sigma_f^2 = \mathrm{Var}_{\mathbf{X}}[f(\mathbf{X})] = \sum_{k=1}^{np}\int_{\Delta\mathcal{V}_k}(f(\mathbf{x})-\mu_f)^2\phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x},$$

but due to the total mean μ_f the total variance does not easily decompose into the strata variances,

$$\begin{aligned}\left(\sigma_f^{(k)}\right)^2 &= \mathrm{Var}_{\mathbf{X}}^{(k)}[f(\mathbf{X})] = \mathrm{E}^{(k)}\left[\left(f(\mathbf{X})-\mu_f^{(k)}\right)^2\right] \\ &= \int_{\Delta\mathcal{V}_k}\left(f(\mathbf{x})-\mu_f^{(k)}\right)^2\phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x}/p_k.\end{aligned} \tag{9.65}$$

Following Glasserman [96], the variance is written in with the usual second and first moment technique,

$$\begin{aligned}\sigma_f^2 &= \mathrm{E}\left[f^2(\mathbf{X})\right] - \mu_f^2 = \textstyle\sum_{k=1}^{np}p_k\mathrm{E}^{(k)}\left[f^2(\mathbf{X})\right] - \mu_f^2 \\ &= \textstyle\sum_{k=1}^{np}p_k\left(\left(\sigma_f^{(k)}\right)^2 + \left(\mu_f^{(k)}\right)^2\right) - \left(\sum_{k=1}^{np}p_k\mu_f^{(k)}\right)^2.\end{aligned} \tag{9.66}$$

In contrast, the strata Monte Carlo estimate, the variance, using prior definintions and the IID property of the $X_i^{(k)}$, is

$$\begin{aligned}\sigma^2_{\widehat{\mu}_{n,np}} &= \mathrm{Var}\left[\widehat{\mu}_{n,np}\right] = \mathrm{E}\left[\textstyle\sum_{k=1}^{np} p_k \left(\frac{1}{n_k}\sum_{i=1}^{n_k} f_i^{(k)} - \mu_f^{(k)}\right)\right] \\ &= \textstyle\sum_{k=1}^{np}\sum_{\ell=1}^{np}\frac{p_k p_\ell}{n_k n_\ell}\sum_{i=1}^{n_k}\sum_{j=1}^{n_k}\mathrm{E}^{(k)}\left[\left(f_i^{(k)}-\mu_f^{(k)}\right)\left(f_j^{(\ell)}-\mu_f^{(\ell)}\right)\right] \\ &= \textstyle\sum_{k=1}^{np}\frac{p_k^2}{n_k^2}\sum_{i=1}^{n_k}\left(\sigma_f^{(k)}\right)^2 = \sum_{k=1}^{np}\frac{p_k^2}{n_k}\left(\sigma_f^{(k)}\right)^2.\end{aligned} \tag{9.67}$$

Thus, the strata reduction of variance will be

$$\begin{aligned}\sigma_f^2 - \sigma^2_{\widehat{\mu}_{n,np}} &= \textstyle\sum_{k=1}^{np} p_k\left(1-\frac{p_k}{n_k}\right)\left(\sigma_f^{(k)}\right)^2 \\ &\quad + \textstyle\sum_{k=1}^{np} p_k\left(\mu_f^{(k)}\right)^2 - \left(\sum_{k=1}^{np} p_k\mu_f^{(k)}\right)^2 \\ &\geq \textstyle\sum_{k=1}^{np} p_k\left(1-\frac{p_k}{n_k}\right)\left(\sigma_f^{(k)}\right)^2,\end{aligned} \tag{9.68}$$

since the second moment majorizes the square of the first moment, here for $\mu_f^{(k)}$ with probability $p_k = 1$. For strata proportional sampling, $n_k = p_k \cdot n$, then

$$\sigma_f^2 - \sigma^2_{\widehat{\mu}_{n,np}} \geq \frac{n-1}{n}\sum_{k=1}^{np} p_k\left(\sigma_f^{(k)}\right)^2, \tag{9.69}$$

so proportional sampling stratification always reduces the variance.

Another form of strata sampling, $n_k = q_k \cdot n$ so $q_k > 0$ and $\sum_{k=1}^{np} q_k = 1$, but q_k is otherwise arbitrary. This form is called **fractional sampling**. The arbitrariness of the fractions q_k can be used to determine the **optimal sampling allocation** of the stratification with the objective to achieve maximum variance reduction for stratification. Since when $n_k = q_k \cdot n$,

$$\sigma_f^2 - \sigma^2_{\widehat{\mu}_{n,np}} \geq \sum_{k=1}^{np} p_k\left(1-\frac{p_k}{nq_k}\right)\left(\sigma_f^{(k)}\right)^2, \tag{9.70}$$

but instead of maximizing the full right-hand-side of (9.68) for $\sigma_f^2 - \sigma^2_{\widehat{\mu}_{n,np}}$, it is only necessary to minimize the bound in (9.70) containing the variable parameter q_k. This can be done using the **Lagrange multiplier** technique to handle the $\sum_{k=1}^{np} q_k = 1$ constraint with λ as the multiplier by letting

$$S(\mathbf{q},\lambda) = \sum_{k=1}^{np} p_k\left(1-\frac{p_k}{nq_k}\right)\left(\sigma_f^{(k)}\right)^2 + \lambda\left(\sum_{k=1}^{np} q_k - 1\right).$$

The reader can easily verify that the optimal allocation solution for the vector of probabilities $\mathbf{q}$ is

$$\mathbf{q}^* = \frac{\mathbf{p}.*\boldsymbol{\sigma}}{(\mathbf{b}^\top\boldsymbol{\sigma})} = \frac{[p_i\sigma_i]_{np\times 1}}{\sum_{k=1}^{np} p_k\sigma_k}, \tag{9.71}$$

by taking the gradient of the objective $S(\mathbf{q}, \lambda)$ with respect to $\mathbf{q}$ and elimination of the multiplier λ. Hence, the optimal bound on the variance reduction is

$$\sigma_f^2 - \left(\sigma^*_{\hat{\mu}_{n,np}}\right)^2 \geq \sum_{k=1}^{np} p_k \left(\sigma_f^{(k)} - \frac{1}{n}\sum_{\ell=1}^{np} p_\ell \sigma_f^{(\ell)}\right)\sigma_f^{(k)}, \tag{9.72}$$

See Glasserman [96] for a more advanced treatment of stratified sampling and see Numerical Recipes [229, 230] for a discussion and the advanced **recursive stratified sampling** code called **MISER**.

9.2.6 Antithetic Variates

The **antithetic variate technique** of variance reduction reuses a prior draw, called the **thetic (or thesis) variate**, to construct an opposing random variable, called the antithetic variate and is usually a mirror image of the thetic variate with the same mean, that has a negative correlation with the thetic variate. The most common antithetic examples are $U_i^{(a)} = 1 - U_i$ for the standard uniform distribution on $[0, 1]$ and $Z_i^{(a)} = -Z_i$ for the standard normal distribution. Hence, $\mathrm{E}[U_i^{(a)}] = 0.5 = \mathrm{E}[U_i]$, $\mathrm{Var}[U_i^{(a)}] = 1/12 = \mathrm{Var}[U_i]$ and

$$\mathrm{Cov}[U_i, U_i^{(a)}] = -1/12 < 0$$

for the uniform, while $\mathrm{E}[Z_i^{(a)}] = 0 = \mathrm{E}[Z_i]$, $\mathrm{Var}[Z_i^{(a)}] = 1 = \mathrm{Var}[Z_i]$ and

$$\mathrm{Cov}[Z_i, Z_i^{(a)}] = -1 < 0.$$

The analogous properties hold when the uniform and normal distributions are not standard, i.e, $X_i^{(a)} = b + a - X_i$ on $[a, b]$ for the uniform and $X_i^{(a)} = 2\mu - X_i$ for the normal with mean μ and variance σ. For most other continuous distributions, the samples are drawn from these to standard distributions and are converted by transformation to the target distribution. For example, $X_i = -\mu \ln(U_i)$ and

$$X_i^{(a)} = -\mu \ln(1 - U_i) = -\mu \ln(1 - \exp(-X_i/\mu))$$

for the exponential distribution with mean μ, using Example 9.6.

In order to keep this section from being too complicated, it will be assumed that the distribution from which the Monte Carlo random variates will be drawn will be from the general uniform in one-dimension ($n_x = 1$) with density $\phi(x) = 1/(b-a)$ on (a, b),

$$\mu_f = \frac{1}{(b-a)} \int_a^b f(x) dx.$$

Note that the antithetic mean will be the same as the thetic mean,

$$\mu_f^{(a)} = \frac{1}{(b-a)} \int_a^b f(b + a - x) dx = \frac{1}{(b-a)} \int_a^b f(y) dy = \mu_f.$$

For the Monte Carlo estimates,

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i); \quad \text{and} \quad \widehat{\mu}_n^{(a)} = \frac{1}{n}\sum_{i=1}^{n} f\left(X_i^{(a)}\right),$$

both converging to μ_f by the strong law of large numbers. For the antithetic variate (av) technique, define the thetic-antithetic average mean estimate with limit μ_f as

$$\widehat{\mu}_n^{(av)} = \frac{1}{2}\left(\widehat{\mu}_n + \widehat{\mu}_n^{(a)}\right) \tag{9.73}$$

and note that the Monte Carlo sample size has been doubled to $2n$ using only the original IID sample $\{X_i\}$ sample of n points, but at the computational cost of double the number of function evaluations of $f(x)$. However, if the variance can be reduced substantially then the original sample size of n could be reduced to compensate for the additional function evaluations.

The new variance is then

$$\begin{aligned}
\mathrm{Var}\left[\widehat{\mu}_n^{(av)}\right] &= \tfrac{1}{4}\mathrm{Var}\left[\widehat{\mu}_n + \widehat{\mu}_n^{(a)}\right] \\
&= \tfrac{1}{4}\mathrm{E}\left[(\widehat{\mu}_n - \mu_f)^2 + \left(\widehat{\mu}_n^{(a)} - \mu_f\right)^2 + 2(\widehat{\mu}_n - \mu_f)\left(\widehat{\mu}_n^{(a)} - \mu_f\right)\right] \\
&= \tfrac{1}{4}\mathrm{Var}[\widehat{\mu}_n] + \tfrac{1}{4}\mathrm{Var}\left[\widehat{\mu}_n^{(a)}\right] + \tfrac{1}{2}\mathrm{Cov}\left[\widehat{\mu}_n, \widehat{\mu}_n^{(a)}\right].
\end{aligned} \tag{9.74}$$

If the covariance $\mathrm{Cov}[\widehat{\mu}_n, \widehat{\mu}_n^{(a)}]$ is negative, then a variance reduction ratio of no more than one half would be guaranteed, thus paying for the doubled function evaluations in terms of efficiency (9.49). By a result quoted in Boyle et al. [39], if the target function of f is monotonic, then

$$\mathrm{Cov}[\widehat{\mu}_n, \widehat{\mu}_n^{(a)}] < 0,$$

which is likely true in many applications, e.g., positive payoffs, but perhaps difficult to verify. In multidimensions, negativity conditions will likely have to be replaced by negative semi-definite conditions for practical purposes due to independence across dimensions.

Example 9.10. ***Antithetic Variates for Compound Poisson Process:***
In the jump-diffusion European call option pricing problem of Zhu and Hanson [290], it was necessary to draw a sample from the compound Poisson process with rate λ,

$$S_i = \sum_{j=1}^{N_i} Q_{i,j}, \quad \text{for } i = 1{:}n,$$

estimating the Poisson cumulative sum $\widetilde{\mathcal{P}}(t_f)$ in Eq. (9.35) of Example 9.3, where the jump-amplitude marks $Q_{i,j}$ are uniformly distributed on $[a, b]$. First the jump count N_i for $i = 1{:}n$ sample points is computed by the inverse transform method in Example 9.9, then a set of standard uniform variates $U_{i,j}$ for $j = 1{:}N_i$ jumps and

$i = 1{:}n$ points (i.e., $Q_{i,j} = a + (b-a)U_{i,j}$ and $Q^{(a)}_{i,j} = a + (b-a)(1-U_{i,j})$). Next the partial sums are computed,

$$S_i = aN_i + (b-a)\sum_{j=1}^{N_i} U_{i,j} \quad \text{and} \quad S_i^{(a)} = (a+b)N_i - S_i, \tag{9.75}$$

which are then used to compute thetic-antithetic averages of jump-shifted Black-Scholes formulas and associated jump-exponentials.

9.2.7 Control Variates

As in importance sampling (Subsection 9.2.4) with its multiplicative factoring of the density by seeking a better density, the **control variate** technique [104] seeks an additive factor, but a known one, that is representative of the variability in the target integrand. This technique was introduced in general by Hammersley-Handscomb [104] in their little book and later introduced to finance along with the antithetic techniques to finance by Boyle [38] in 1977 with a substantial update by Boyle, Broadie and Glasserman [39] in 1997. See also Glasserman's book [96, Sect. 4.1] for more recent advances in finance.

Again, consider the target integral, returning back to n_x-dimensional space $\mathcal{V}$ with density $\phi_{\mathbf{X}}(\mathbf{x})$,

$$\mu_f = \int_{\mathcal{V}} f(\mathbf{x})\phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$$

and basic Monte Carlo estimate

$$\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{X}_i),$$

converging to μ_f as $n \to \infty$ by the strong law of large numbers, where the set $\{\mathbf{X}_i\}$ of n IID sample points drawn are from the density $\phi_{\mathbf{X}}(\mathbf{x})$.

Next, through knowledge of the problem a simpler function $f^{(c)}(\mathbf{x})$ is found which significantly represents the variability of the target function $f(\mathbf{x})$ and can be used as a control (c) enabler, such that

$$\mu_f^{(c)} = \int_{\mathcal{V}} f^{(c)}(\mathbf{x})\phi_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$$

is known or those value can be accurately approximated. Using the same IID set $\{\mathbf{X}_i\}$ of sample points, the basic Monte Carlo estimate is

$$\widehat{\mu}_n^{(c)} = \frac{1}{n}\sum_{i=1}^{n} f^{(c)}(\mathbf{X}_i),$$

which is convergent to and is an unbiased estimate of $\mu_f^{(c)}$. The error $\left(\widehat{\mu}_n^{(c)} - \mu_f^{(c)}\right)$ will be used as a control variable to control the variance reduction of the basic

unbiased estimate $\widehat{\mu}_n$ of μ_f by constructing a potentially improved control variate (cv) estimate,

$$\widehat{\mu}_n^{(cv)}(\alpha) \equiv \widehat{\mu}_n - \alpha \left(\widehat{\mu}_n^{(c)} - \mu_f^{(c)} \right), \tag{9.76}$$

where α is a control parameter that will be optimized given the knowledge of the control function $f^{(c)}$. In particular, the partly known error $\left(\widehat{\mu}_n^{(c)} - \mu_f^{(c)} \right)$ will be used to control the control variate estimate error $\left(\widehat{\mu}_n^{(cv)}(\alpha) - \mu_f \right)$, noting from (9.76),

$$\mathrm{E}\left[\widehat{\mu}_n^{(cv)}(\alpha) \right] = \mu_f - \alpha \left(\mu_f^{(c)} - \mu_f^{(c)} \right) = \mu_f,$$

that the unbiased estimation of μ_f is unchanged.

Upon examining the variance of the control variate estimate in terms of α following [96],

$$\begin{aligned} \mathrm{Var}\left[\widehat{\mu}_n^{(cv)}(\alpha) \right] &= \mathrm{Var}\left[\widehat{\mu}_n - \alpha \left(\widehat{\mu}_n^{(c)} - \mu_f^{(c)} \right) \right] \\ &= \mathrm{E}\left[\left((\widehat{\mu}_n - \mu_f) - \alpha \left(\widehat{\mu}_n^{(c)} - \mu_f^{(c)} \right) \right)^2 \right] \\ &= \mathrm{Var}[\widehat{\mu}_n] - 2\alpha \mathrm{Cov}\left[\widehat{\mu}_n, \widehat{\mu}_n^{(c)} \right] + \alpha^2 \mathrm{Var}\left[\widehat{\mu}_n^{(c)} \right], \end{aligned} \tag{9.77}$$

a simple quadratic optimization in α produces an optimal control parameter,

$$\alpha^* = \frac{\mathrm{Cov}\left[\widehat{\mu}_n, \widehat{\mu}_n^{(c)} \right]}{\mathrm{Var}\left[\widehat{\mu}_n^{(c)} \right]} = \frac{\rho_{\widehat{\mu}_n, \widehat{\mu}_n^{(c)}} \sqrt{\mathrm{Var}[\widehat{\mu}_n]}}{\sqrt{\mathrm{Var}\left[\widehat{\mu}_n^{(c)} \right]}}, \tag{9.78}$$

where the correlation function is

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}}.$$

Thus, the optimal control variate variance is

$$\mathrm{Var}\left[\widehat{\mu}_n^{(cv)}(\alpha^*) \right] = \mathrm{Var}\left[\widehat{\mu}_n \right] - \frac{\left(\mathrm{Cov}\left[\widehat{\mu}_n, \widehat{\mu}_n^{(c)} \right] \right)^2}{\mathrm{Var}\left[\widehat{\mu}_n^{(c)} \right]} = \left(1 - \left(\rho_{\widehat{\mu}_n, \widehat{\mu}_n^{(c)}} \right)^2 \right) \mathrm{Var}\left[\widehat{\mu}_n \right], \tag{9.79}$$

so the absolute value of the correlation $|\rho_{\widehat{\mu}_n, \widehat{\mu}_n^{(c)}}|$ must be less than one for variance reduction. Note that Hammersley and Handscomb in their 1964 book [104, Sect. 5.5] do not use a control parameter (i.e., $\alpha \equiv 1$) and so require from (9.77) with $\alpha = 1$ that

$$2\mathrm{Cov}\left[\widehat{\mu}_n, \widehat{\mu}_n^{(c)} \right] > \mathrm{Var}\left[\widehat{\mu}_n^{(c)} \right],$$

i.e., the covariance must be sufficiently positive, unlike (9.79). In fact, the optimal variance reduction ratio, from (9.79) and from the definition of VRR (9.47), is

$$\mathrm{VRR}^*_{\widehat{\mu}_n, \widehat{\mu}_n^{(c)}} \equiv \frac{\mathrm{Var}[\widehat{\mu}_n]}{\mathrm{Var}\left[\widehat{\mu}_n^{(cv)}(\alpha^*) \right]} = \frac{1}{\left(1 - \left(\rho_{\widehat{\mu}_n, \widehat{\mu}_n^{(c)}} \right)^2 \right)}, \tag{9.80}$$

so the absolute value of the correlation $|\rho_{\widehat{\mu}_n,\widehat{\mu}_n^{(c)}}|$ should not only be less than one, but should be sufficiently close to one for significant variance reduction, in theory.

However, the exact statistics represented in the optimal parameter α^* in (9.78) and particularly the related optimal correlation $\rho_{\widehat{\mu}_n,\widehat{\mu}_n^{(c)}}$ are unknown. Hence, in practice, an estimate of α^* is needed, leading to the sample control parameter estimate of α^*,

$$\begin{aligned}\widehat{\alpha}_n &= \frac{\widehat{c}_n^{(c)}}{\left(\widehat{\sigma}_n^{(c)}\right)^2} \equiv \frac{\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(f_i-\mu_f)\left(f_i^{(c)}-\mu_f^{(c)}\right)}{\dfrac{1}{n-1}\displaystyle\sum_{j=1}^{n}\left(f_j^{(c)}-\mu_f^{(c)}\right)^2} \\ &= \frac{\displaystyle\sum_{i=1}^{n}(f_i-\mu_f)\left(f_i^{(c)}-\mu_f^{(c)}\right)}{\displaystyle\sum_{j=1}^{n}\left(f_j^{(c)}-\mu_f^{(c)}\right)^2}\end{aligned} \tag{9.81}$$

and the corresponding estimated control variate Monte Carlo estimate

$$\widehat{\mu}_n^{(cv)}(\widehat{\alpha}_n) = \widehat{\mu}_n - \widehat{\alpha}_n\left(\widehat{\mu}_n^{(c)} - \mu_f^{(c)}\right), \tag{9.82}$$

but introducing some bias particularly due to the approximate covariance $\widehat{c}_n^{(c)}$ in (9.81). The bias (9.40) is given by

$$\beta_{\widehat{\mu}_n^{(cv)}} = \mathrm{E}\left[\widehat{\mu}_n^{(cv)}(\widehat{\alpha}_n) - \mu_f\right] = -\mathrm{E}\left[\widehat{\alpha}_n\left(\widehat{\mu}_n^{(c)} - \mu_f^{(c)}\right)\right], \tag{9.83}$$

which in general will be nonzero due to the nonlinear dependence of $\widehat{\alpha}_n$.

Example 9.11. ***Control Variate Adjusted Jump-Diffusion Payoff:*** *Zhu and Hanson [290] further reduced the variance of the thetic-antithetic adjusted jump-factor Black-Scholes mentioned in Example 9.10 using the error in the thetic-antithetic adjusted jump-factor,*

$$\Delta Y_i = 0.5\left(e^{S_i} + e^{S_i^{(a)}}\right) - e^{\lambda t_f \bar{J}},$$

where the partial sums S_i and $S_i^{(a)}$ are given in (9.75), $\bar{J} \equiv \mathrm{E}[J(Q)$ is the asset mean jump amplitude given in (9.36) of Example 9.3 and t_f is the option exercise time. The complex corrections to the bias $\beta_{\widehat{\mu}_n^{(cv)}}$ in (9.83) are given and proven in [290, 289] along with other results. The combination of antithetic and control variate variance reduction techniques were easy to implement and were efficient in spite of the theoretical complexity and the combination was better than either one separately.

For more formation in depth on the control variate technique, see Boyle, Broadie and Glasserman [39] and Glasserman [96].

Another topic that is important but beyond the scope of this book is the quasi-Monte Carlo method which uses quasi-random or low-discrepancy number sequences which are more genuine deterministic sequences than the pseudo-random number sequences commonly used. Their generation is more complex generally then the pseudo-random sequences, but their big benefit is that convergence is between genuine order $\text{ord}(1/\sqrt{n})$ and $\text{ord}(1/n)$, so can outperform the variance reduction techniques just discussed. See Niederreiter [218] for the basic theoretical background to the quasi-Monte Carlo method. For more general information, see Glasserman [96, Chapt. 5] and Jäckel [150, Chapt. 8]. The Sobol' [253] quasi-random numbers seem to be the best overall performers in various measures as demonstrated in [96, Figs. 514-5.16] and [150, Figs. 8.2-8.9]. Also, see *Numerical Recipes* [230, Sect. 7.7] of Press et al. for the Sobol' sequence code `sobseq`.

Suggested References for Further Reading

- Applebaum, 2004 [12].
- Beichl and Sullivan, 2000 [23].
- Boyle, 1977 [38].
- Boyle, Broadie and Glasserman, 1997 [39].
- Cyganowski and Kloeden, 2000 [65].
- Cyganowski, Grüne and Kloeden, 2002 [64].
- Cyganowski, Kloeden and Ombach, 2002 [66].
- Glasserman, 2003 [96].
- Glynn and Whitt, 1992 [97].
- Hammersley and Handscomb, 1964 [104].
- D. Higham, 2001 [139] and 2004 [140].
- D. Higham and Kloeden, 2002 [143] and 2005 [144].
- D. Higham, Mao and Stuart, 2002 [146].
- Jäckel, 2002 [150].
- Kalos and Whitlock, 1986 [157].
- Kloeden and Platen, 1992 [165].

- Kloeden, Platen and Schurz, 1994 [166].
- Lepage, 1978 [183].
- Maghsoodi , 1996 [191].
- Maghsoodi and Harris, 1987 [192].
- Metropolis, A. Rosenbluth, M Rosenbluth, A. Teller, and E. Teller, 1953 [207].
- Metropolis and Ulam, 1949 [208].
- Niederreiter, 1992 [218].
- Press and Farrar, 1990 [229].
- Press, Teukolsky, Vetterling and Flannery, 2002 [230].
- Zhu, 2005 [289].
- Zhu and Hanson, 2005 [290].

Chapter 10

Applications in Financial Engineering

From the point of view of the risk manager, inappropriate use of the normal distribution can lead to an understatement of risk, which must be balanced against the significant advantage of simplification.
—Alan Greenspan (1995), Joint Central Bank Research Conference [79].

Merton (1969, 1971, 1973) uses the formula from Itô's lemma and the continuous-time Bellman equation, but otherwise uses none of the concepts and methods of proof developed by Lebesgue and followers.
—Harry M. Markowitz in the forward to [245].

There is never enough time, unless you're serving it.
—Malcolm Forbes (1919 - 1990),
`http://www.quotationspage.com/quote/957.html`.

Stochastic effects play a major role in financial engineering applications, either using a combination of financial assets and other instrument to remove stochasticity all together through hedging or balancing securities, or just accommodating the financial portfolio analysis to stochastic effects. For general background, the formal derivation of the classical Black-Scholes [34] option pricing model is presented, but for students already familiar with the Black-Scholes formulation, they may prefer to skip to the next more rigorous section. Applying methods previously developed, this chapter presents the derivation of the Black-Scholes-Merton [34, 201, 203] formula for pricing European call and put options from the stock, bond and option portfolio diffusion model, including Merton fractions and self-similar solutions [203]. A related option pricing Merton study for underlying stock-bond jump-diffusion models is also discussed. In addition, optimal consumption and portfolio policies for con-

stant relative risk aversion (CRRA) utilities of terminal wealth and instantaneous consumption is discussed for marked jump-diffusions. The notion of scheduled event with distributed response (the so-called **Greenspan processes**) [235, 122, 129] is presented. The role of optimal stochastic control in finance is discussed. The stock jump-diffusion probability density is derived for the linear model treating the composite process as a triad of independent random variables [124, 123].

10.1 Classical Black-Scholes Option Pricing Model

The Black-Scholes option pricing model [34, 35] is perhaps the most used financial model in financial engineering, had been called the most seminal work in finance in the last 25 years during 1997 and is probably the most cited work in finance.

The Black-Scholes model is for a portfolio containing a stock option, hedged with the stock itself with price $S(t)$ and a risk-less bond with price $B(t)$ at time t providing a constant reference market rate of interest r. The option is assumed to be a **European option**, i.e., there is a contract with a fixed time-to-maturity T to either buy a number of shares of the stock at a given **exercise price** K per share at contract expiration time T (called a **European call option**) or to sell a number of shares of a stock at a given price K per share at a number of shares of a stock at a given exercise price K per share at contract expiration time T (called a **European put option**). The call and put options can be considered together, since they share the same financial market model except for different final boundary conditions at expiration $t = T$. The options contract is between the investor (buyer) and maker (writer) of the contract.

At the end of the term, $t = T$, of the contract, the investor's call option payoff or exercise profit is

$$\max[S(T) - K, 0]. \tag{10.1}$$

So the profit from exercising the option is positive only if the final stock price $S(T)$ per share exceeds the contract exercise price K, in which case the investor can buy the stocks at price K, i.e., exercise the option, and then sell the stocks in the market for price $S(T)$. Otherwise, the rational investor does nothing, i.e., does not exercise the option contract. At the start $t = 0$, the investor must bet that $S(T)$ will rise above K and the fixed cost of the bet is the option price, Y_0. Hence, the investor net profit for is the payoff (10.1) less the call option price Y_0 for the contract. The net profit position of the contract writer is just the opposite of that of the investor or contract buyer. See Hull [147, pp. 5-10] for a simple, concrete example. A simple version of Black-Scholes model will be given here, following Hull [147], but with our notation and added explanations.

The situation is reversed for the put option. At the end of the term of the contract, the investor's put option payoff or exercise profit is

$$\max[K - S(T), 0]. \tag{10.2}$$

So the profit is positive only if the final stock price $S(T)$ drops below the contract exercise price K, in which case the investor can sell the stocks at price K to the

contract maker and then buy stocks more cheaply in the market for $S(T)$, else the rational investor does nothing, i.e., does not exercise the option. At the start $t = 0$, the investor bets that $S(T)$ will fall below K and again the fixed cost of the bet is the option price. Again, the net profit is the payoff (10.2) less the put option price Y_0 for the contract.

Let the stock or other asset price $S(t)$ dynamics satisfy the linear SDE (often called **geometric Brownian motion**),

$$dS(t) = S(t)\left(\mu dt + \sigma dW(t)\right) \ , \quad S(0) = S_0 \ , \tag{10.3}$$

where μ is the constant rate of appreciation of the stock price and σ is the constant volatility (standard deviation) in the stock price. The bond price equation is not really needed, only that a risk-less investment grows at a constant rate r so that at time t the principal has grown by an exponential factor $\exp(rt)$ from time zero.

Let the option price be given by the function $Y = F(S(t), t)$ with exercise price K at exercise time $t = T$ when the starting stock price is S_0 at $t = 0$. By the stochastic chain rule, the option price changes according to the SDE,

$$\begin{aligned} dY(t) = dF(S(t), t) = &\left(\frac{\partial F}{\partial t} + \mu S(t)\frac{\partial F}{\partial S} + \frac{1}{2}\sigma^2 S^2(t)\frac{\partial^2 F}{\partial S^2}\right) dt \\ &+\sigma S(t)\frac{\partial F}{\partial S} dW(t) \ , \end{aligned} \tag{10.4}$$

where all partial derivatives are evaluated at $(S(t), t)$, e.g.,

$$\frac{\partial F}{\partial t} = \frac{\partial F}{\partial t}(S(t), t)$$

denotes the partial derivative of $F(s, t)$ with respect to the second argument t with the first argument s held fixed and evaluated at $(S(t), t)$ after differentiation. The major problem evaluating the initial option price $Y_0 = F(S_0, 0)$ is the volatility or uncertainty term

$$\sigma\frac{\partial F}{\partial S} dW(t)$$

in (10.4) makes any pricing decision difficult unless this term can be controlled or eliminated (i.e., hedged in the language of options pricing).

So to control or **hedge** the volitility term, the value of a **portfolio** of the option and stock is defined as

$$V(t) = N_F F(S(t), t) + N_S S(t) \ , \tag{10.5}$$

where N_F is the number of options and N_S is the number of stocks or other assets. Finding the change in the portfolio value is one of the not so clear assumptions in Black-Scholes option pricing derivations that is addressed in more detail in the next section, but there are also many other explanations, such as in D. Higham's nice introductory options book with emphasis on simulations [140]. For the simple

description here, it is assumed that the change in the numbers alone , $FdN_F + SN_S$ is negligible compared to other changes, i.e.,

$$dV(t) = N_F dF(S(t), t) + N_S dS(t) \,. \tag{10.6}$$

This formula is also called a **self-financing strategy**. Other arguments given are that the N_F and N_S are fixed during changes in F and S or that N_F and N_S change slowly compared to F and S. In fact, the seminal paper of Black and Scholes [34] did take a year or more to get published due to this and other questions [204, 51].

Next, we are interested in eliminating the deviation of the portfolio change for fixed F and S,

$$dV(t) - \mathrm{E}[dV(t)|F, S] = \sigma S \left(N_F \frac{\partial F}{\partial S} + N_S \right) dW(t).$$

So the optimal volatility eliminating hedge is to select the stock number to be

$$N_S^* = -N_F^* \frac{\partial F}{\partial S} \,,$$

where

$$\Delta_F = \partial F / \partial S \tag{10.7}$$

is called the portfolio **delta** in finance and the hedge is called a **delta hedge** [147, pp. 310ff]. In terms of fractions with $N_S^* + N_F^* = N$ for fixed N,

$$\frac{N_S^*}{N} = \frac{\frac{-\partial F}{\partial S}}{1 - \frac{\partial F}{\partial S}}, \quad \text{and} \quad \frac{N_F^*}{N} = \frac{1}{1 - \frac{\partial F}{\partial S}} \,,$$

providing, $\partial F / \partial S \neq 1$. At this point, we will ignore any contradiction with the self-financing assumption, relying in the end that Black-Scholes gives a reasonable and successful formula for option pricing.

Thus,

$$dV^*(t) = N_F^* \left(dF - \frac{\partial F}{\partial S} dS \right) = N_F^* \left(\frac{\partial F}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2 F}{\partial S^2} \right) dt \,, \tag{10.8}$$

using (10.3,10.4), while the optimal portfolio value becomes

$$V^*(t) = N_F^* \left(F - \frac{\partial F}{\partial S} S \right) \,. \tag{10.9}$$

In addition, it is necessary to avoid arbitrage, taking advantage of price differentials to make a profit without the trader making his or her own investment. So it is required that the portfolio earn a return at the risk-less market rate r or

$$dV^*(t) = rV^*(t)dt \,, \tag{10.10}$$

the no-arbitrage condition. Finally, the Black-Scholes PDE is formed by combining (10.10) with (10.8) and (10.9), then replacing the stock path function $S(t)$ by the independent stock variable S,

$$\frac{\partial F}{\partial t}(S,t) + \frac{1}{2}\sigma^2 \frac{\partial^2 F}{\partial S^2}(S,t) = r\left(F(S,t) - S\frac{\partial F}{\partial S}(S,t)\right) . \tag{10.11}$$

Note that the random volatility term and the mean appreciation (μ) term no longer appear, but the volatility coefficient appears due to the Itô diffusion coefficient correction. This PDE is a backward PDE for t on $[0,T)$ and S on $[0,\infty)$ with final condition at $t = T$ for any non-negative S,

$$F(S,T) = \mathcal{C}(S,T) = \max[S-K,0]$$

for a call option from (10.1) and

$$F(S,T) = \mathcal{P}(S,T) = \max[S-K,0]$$

for a put option from (10.2). The well-known formula [34, 140] for solution to this PDE can be found in terms of the normal distribution function, but only the results are given here since the details are presented for the more general case in the next section. In the case of the European call option, the Black-Scholes formula is

$$\mathcal{C}_0(S_0) \equiv \mathcal{C}(S_0,0) = S_0\Phi_n(d_1(S_0);0,1) - Ke^{-rT}\Phi_n(d_2(S_0);0,1) , \tag{10.12}$$

where the variable arguments of the normal distribution function $\Phi_n(w;\mu,\sigma^2)$ are

$$d_1(s) \equiv \frac{\ln(s/K) + (r+\sigma^2/2)T}{\sigma\sqrt{T}}$$

and

$$d_2(s) \equiv d_1(s) - \sigma\sqrt{T}.$$

In the case of the European put option, the Black-Scholes formula can be found by the well-known and very general **Put-Call Parity** that depends basically on the properties of the maximum function [203, 140],

$$\mathcal{P}_0(S_0) \equiv \mathcal{P}(S_0,0) = \mathcal{C}_0(S_0) + Ke^{-rT} - S_0 . \tag{10.13}$$

In 1900, Bachelier [16, 61], a student of Poincaré, published a theory of option pricing that was derived from his thesis, but his work was little noticed at the time. Unlike the Black-Scholes diffusion option pricing model based upon the geometric Brownian motion stochastic model, Bachelier's option pricing model was based upon additive Brownian motion, i.e., instead of being linear in the stock price as in the multiplicative noise (MultNoise1) case, the noise was independent of stock price and thus additive noise (4.31). Bachelier's paper was a very early, very complete and straight-forward application of stochastic processes in finance. The main drawback is the additive noise, since stock price fluctuations are now assumed to act in a compound or multiplicative fashion.

10.2 Merton's Three Asset Option Pricing Model Version of Black-Scholes

Sometimes the Black-Scholes model is called the Black-Scholes-Merton model, since Merton [201], in his **Theory of Rational Option Pricing** paper, gave substantial mathematical justification of the seminal Black-Scholes model using stochastic diffusion processes. Merton's paper includes generalizations of the Black-Scholes model that provide greater foundations and limitations for their model. Both the Black-Scholes and the Merton papers were published in Spring 1973, Merton having held up the publication of his paper out deference to Black and Scholes' original model. Robert C. Merton and Myron Scholes shared the 1997 Nobel prize in economics for the accomplishments, but unfortunately Fischer Black [204, 51] had passed away in 1995.

The version of the model presented here is based mainly on Merton's more general framework [201] (reprinted in Chapter 8 of [203]). The model is for a portfolio containing a European stock option, hedged with the stock itself with price $S(t)$ and a risk-less bond with price $B(t)$ at time t, but with more explicit assumptions than for the classical Black-Scholes.

The market model is comprised of a number of assumptions which will be enumerated and marked with **BSM** here, but formulated in the notation and spirit of this book. Multiple assumptions of Merton have been decomposed into single assumptions to make them easier to modify new problems. One of the objectives of this book to offer sufficient detail to enable the reader to become a practitioner making those new modifications. The more general model of Merton is treated here since many readers will be familiar with the simpler, classic versions of the Black-Scholes option pricing model which can be found in many of the references listed at the end of this chapter, e.g., Hull [147] or Wilmott et al. [282].

- **Assumption BSM1. Frictionless Markets:**
  There are no transaction fees for transactions involving the buying or selling of the three assets in the portfolio, excluding the original price of the option contract.

- **Assumption BSM2. No Dividends:**
  There are no dividends paid on the stock asset.

- **Assumption BSM3. Continuous Trading, without Jumps:**
  Trading among the assets is continuous, so discrete aspects of trading such as jumps are neglected. This assumption is consistent with the no transaction fees and no dividends of the prior two assumptions, since those are discrete events.

- **Assumption BSM4. Borrowing and Short Selling Allowed:**
  Short selling of stock or options is allowed within the term of the contract, with funds placed into the bond asset. Borrowing from the bond asset is allowed to increase the number of shares of the other two assets. Also, it is assumed that the borrowing rate is the same as the lending rate.

- **Assumption BSM5. Linear Stock-Price Stochastic Dynamics:**
  Let $S(t)$ be the **price of stock** per share at time t. Then the $S(t)$ satisfies a linear stochastic diffusion differential equation, written in terms of the rate of return or relative change in time dt:

$$dS(t)/S(t) = \mu_S(t)dt + \sigma_S(t)dW_S(t) \ , \tag{10.14}$$

  where

  - $dW_S(t)$ = stochastic diffusion differential process for the stock price process $S(t)$, such that $\mathrm{E}[dW_S(t)] = 0$ and

$$(dW_S)^2(t) \stackrel{ims}{=} dt,$$

    else

$$dW_S(t)dW_S(s) \stackrel{ims}{=} 0 \ , \quad \text{if} \ \ s \neq t \ ,$$

    by independent increments.
  - $\mu_S(t) = \mathrm{E}[dS(t)/S(t)]/dt$ = instantaneous **expected rate of return** on the stock in time dt.
  - $\sigma_S^2(t) = \mathrm{Var}[dS(t)/S(t)]/dt$ = instantaneous variance of the rate of return on the stock in time dt, while σ_S is the **volatility** of the stock return. Here volatility denotes a measure of uncertainty [147], but is derived from the French word meaning *to fly*.

  Here, a stock is considered a **risky asset**, compared to the bond asset. Since the option profit at exercise depends only on the stock price $S(T)$ at the expiration of the option, for some analysis it is more convenient to view the process in backward time $\tau = T - t$, also called the time-to-maturity, and to consider the stock price in that variable, i.e.,

$$\widehat{S}(\tau) \equiv S(T - \tau) \ .$$

- **Assumption BSM6. Linear Bond-Price Stochastic Dynamics:**
  Let $B(t)$ be the **price of bond** asset at time t, in particular a default-free zero-coupon bond or discounted loan with time-to-maturity T. Then the $B(t)$ satisfies a linear stochastic diffusion differential equation, written in terms of the rate of return or relative change in time dt:

$$dB(t)/B(t) = \mu_B(t)dt + \sigma_B(t)dW_B(t) \ , \tag{10.15}$$

  where

  - $dW_B(t)$ = Stochastic diffusion differential process for the bond return process $B(t)$, such that $\mathrm{E}[dW_B(t)] = 0$ and

$$(dW_B)^2(t) \stackrel{ims}{=} dt,$$

    otherwise

$$dW_B(t)dW_B(s) \stackrel{ims}{=} 0 \ , \quad \text{if} \ \ s \neq t \ ,$$

    by independent increments.

- $\mu_B(t) = \mathrm{E}[dB(t)/B(t)]/dt$ = instantaneous **expected rate of return** on the bond asset.
- $\sigma_B^2(t) = \mathrm{Var}[dB(t)/B(t)]/dt$ = instantaneous variance of the rate of return on the stock, while σ_B is the **volatility** of the stock return.

Here, a bond is usually considered a **lower risk asset**, compared to the higher risk or risky stock asset. Here, the variance or volatility will be taken as a **measure of riskiness**, so we say that the stock is **riskier** or **more risky** than the bond if $\sigma_S > \sigma_B$. We say that the bond is **risk-free** if $\sigma_B = 0$. However, Merton [203] has more precise measures of riskiness, though not as easy to apply (see Exercise 2).

In the more classical Black-Scholes model, the bond-price is assumed to be deterministic, so $\sigma_B(t) = 0$, and the mean rate is assumed to be constant, so $\mu_B(t) = r$. In this ideal case the bond is called **risk-free** or **risk-less**.

In the case where the bond is treated as a discounted loan, then the payback is at the final price $B(T)$, the initial discounted loan amount received is $B(0)$, which should be less than $B(T)$, so $(B(T) - B(0)) > 0$ is the amount discounted. Discounting is a backward time version of interest on principal. In the backward time problem, the time-to-maturity or time-to-go $\tau = T - t$ is the natural time variable.

In the non-stochastic interest rate problem, as in the traditional Black-Scholes formulation, $\sigma_B = 0$, $\mu_B = r$, the mean interest rate for borrowing and selling, and the bond price in backward time is

$$\widehat{B}(\tau) \equiv B(T - \tau).$$

So the bond price decays away from expiration,

$$d\widehat{B}(\tau) = -r\widehat{B}(\tau)d\tau$$

with the bond price decaying in τ due to discounting,

$$\widehat{B}(\tau) = \widehat{B}(0)e^{-r\tau} = B(T)e^{-r(T-t)} \ .$$

This backward time view is consistent with the options contract where the profit depends on the final stock price $S(T)$ and the objective is to find the number of shares as the initial price $S(0)$ in the final value problem for a stochastic differential equation.

- **Assumption BSM7. Bond and Stock Price Fluctuations are Correlated, but Not Serially:**
Thus, the correlation properties between the stock price noise and the bond price noise are

$$dW_B(t)dW_S(t) \stackrel{ims}{=} \rho dt \ , \tag{10.16}$$

$$dW_B(t)dW_S(s) \stackrel{ims}{=} 0 \ , \quad \text{if} \ \ s \neq t. \tag{10.17}$$

The former equation (10.16) expressing correlation on the same time increments at t (see Exercise 1 for the proof), while the latter equation (10.17) expresses the lack of serial correlation on disjoint time intervals when $s \neq t$, also preserving the independent increment property, where

$$\rho \equiv \frac{\text{Cov}[dS(t), dB(t)]}{\sqrt{\text{Var}[dS(t)]\text{Var}[dB(t)]}} = \frac{\text{Cov}[dW_S(t), dW_B(t)]}{dt} \tag{10.18}$$

= the instantaneous correlation coefficient between stock and bond returns, provided $\sigma_S(t)$ and $\sigma_B(t)$ are positive.

The joint density for $(dW_S(t), dW_B(t))$ is obviously the bivariate normal density in (B.146) of preliminaries Chapter B,

$$\phi_{(dS(t),dB(t))}(s,b) = \phi_n\left(\begin{bmatrix} s \\ b \end{bmatrix}; \begin{bmatrix} \mu_S(t) \\ \mu_B(t) \end{bmatrix} dt, \begin{bmatrix} \sigma_S^2(t) & \rho(\sigma_S\sigma_B)(t) \\ \rho(\sigma_S\sigma_B)(t) & \sigma_B^2(t) \end{bmatrix} dt\right) . \tag{10.19}$$

Merton [201, 203] claims that the lack of serial correlations is consistent with the **Efficient Markets Hypothesis**. In the simpler expositions of the Black-Scholes model, there are no correlations, so $\rho \equiv 0$, with $\sigma_B = 0$ and $\mu_B = r$, the common interest rate. The mean square limit for non-serial correlation (10.16) is left as an exercise for the reader.

- **Assumption BSM8. No Investor Preferences or Expectations, except for Agreement on Parameters:**
  The investors agree on and have reasonable knowledge the parameters, such as the means μ_S and μ_B, as well as the volatilities σ_S and σ_B.

- **Assumption BSM9. Option Price is a Function of Stock and Bond Prices:**
  The option price per share at time t,

$$Y(t) = F(S(t), B(t), t; T, K) \tag{10.20}$$

depends on the stock S and bond B price stochastic variables, as well as on time t and parameters such as the time-to-maturity T and the contracted expiration stock price K per share.

Alternatively, the relationship can be cast in terms of the time-to-maturity, $\tau = T - t$,

$$\widehat{Y}(\tau) = F(S(T-\tau), B(T-\tau), T-\tau; T, K) .$$

Although we are interested in the initial option price $Y(0) = \widehat{Y}(T)$, considering the time dependent option price $Y(t) = \widehat{Y}(\tau)$ allows analysis of the problem and yields more general results that permit conversion of the option contract

to another investor at the current option price $Y(t) = \widehat{Y}(\tau)$. In the case of constant coefficients, then the results will depend on the general time-to-exercise $tau = T - t$ without restriction to a fixed exercise time T.

Using a two-state-dimensional version of the stochastic diffusion chain rule, the return on the option asset, initially keeping all quadratic terms in the Taylor expansion, is

$$\begin{aligned} dY(t) &= dF(S(t), B(t), t; T, K) \\ &\stackrel{ims}{=} F_t dt + F_S dS(t) + F_B dB(t) \\ &\quad + \frac{1}{2}\left(F_{SS}\,(dS)^2(t) + 2F_{SB}\,dB(t)dS(t) + F_{BB}\,(dB)^2(t)\right), \end{aligned} \tag{10.21}$$

omitting higher order terms that obviously have zero mean limits. Here, $\{F_S, F_B, F_{SS}, F_{SB}, F_{BB}\}$ are the set of first and second partial derivatives of $F(S, B, t; T, K)$ with respect to the underlying portfolio assets S and B. Next, substitution for the return processes $S(t)$ and $B(t)$ is used, along with the quadratic differential forms in the mean square limit,

$$\begin{aligned} (dS)^2(t) &\stackrel{ims}{=} \sigma_S^2(t)S^2(t)dt \\ (dB)^2(t) &\stackrel{ims}{=} \sigma_B^2(t)B^2(t)dt \\ (dBdS)(t) &\stackrel{ims}{=} \rho \cdot \sigma_B(t)\sigma_S(t)B(t)S(t)dt\ , \end{aligned} \tag{10.22}$$

which simply follow from the corresponding mean square limit differential forms for $(dW_S)^2(t)$, $(dW_B)^2(t)$ and $(dW_B dW_S)(t)$, respectively, given under previous assumptions. This forces the geometric Brownian motion form on the option price,

$$dY(t) \stackrel{ims}{=} Y(t)\left(\mu_Y(t)dt + \sigma_{YS}(t)dW_S(t) + \sigma_{YB}(t)dW_B(t)\right)\ , \tag{10.23}$$

where the new option instantaneous return moment coefficients are defined as

$$\begin{aligned} Y(t)\mu_Y(t) &\equiv F_t + \mu_S S F_S + \mu_B B F_B \\ &\quad + \frac{1}{2}\left(\sigma_S^2 S^2 F_{SS} + 2\rho\sigma_S\sigma_B SBF_{SB} + \sigma_B^2 B^2 F_{BB}\right)\ , \end{aligned} \tag{10.24}$$

$$Y(t)\sigma_{YS}(t) \equiv \sigma_S S F_S\ , \tag{10.25}$$

$$Y(t)\sigma_{YB}(t) \equiv \sigma_B B F_B\ . \tag{10.26}$$

- **Assumption BSM10. Self-Financing Portfolio Investments:**
Let $N_S(t)$, $N_Y(t)$ and $N_B(t)$ be the instantaneous number of shares invested in the stock, option, and bond at time t, respectively, such that the instantaneous values of the assets in dollars are

$$V_S(t) = N_S(t)S(t)\ ,\ V_Y(t) = N_Y(t)Y(t)\ ,\ V_B(t) = N_B(t)B(t)\ , \tag{10.27}$$

respectively. However, it is assumed there is a **zero instantaneous aggregate portfolio value**,

$$V_P(t) = V_S(t) + V_Y(t) + V_B(t) = 0\ , \tag{10.28}$$

so that the bond value variable can be eliminated

$$V_B(t) = -(V_S(t) + V_Y(t)) \ . \tag{10.29}$$

Merton [203] defines a **self-financing portfolio** as a trading strategy in which no capital is put in or taken out until maturity. Such a strategy avoids an imbalance between the stock and its option, which would soon disappear as other investors took advantage of the imbalance. This strategy is also related to the avoidance of arbitrage profits and in Black-Sholes is $\mu = r$. Further, this strategy also includes a **no consumption** of assets assumption.

It is further assumed that the absolute instantaneous return from the value of the portfolio $V_P(t)$ is a linear combination of the instantaneous returns in each of the three assets, (S, Y, B), giving the **portfolio budget equation**

$$\begin{aligned} dV_P(t) &= N_S(t)dS(t) + N_Y(t)dY(t) + N_B(t)dB(t) \\ &= V_S(t)\frac{dS(t)}{S(t)} + V_Y(t)\frac{dY(t)}{Y(t)} + V_B(t)\frac{dB(t)}{B(t)} \ , \end{aligned} \tag{10.30}$$

using (10.27) to convert from number of shares to asset value assuming that none of the divisors are zero for the latter more convenient form in terms of rates of return. Note that the budget equation can not be expressed as the instantaneous rate of return since $V_P(t) = 0$.

Substituting for the three asset stochastic dynamics from (10.14, 10.15, 10.23) and eliminating the bond value $V_B(t)$ through (10.29),

$$\begin{aligned} dV_P(t) &= V_S\left(\frac{dS}{S} - \frac{dB}{B}\right) + V_Y\left(\frac{dY}{Y} - \frac{dB}{B}\right) \\ &= \left((\mu_S - \mu_B)\, V_S + (\mu_Y - \mu_B)\, V_Y\right) dt \\ &\quad + \left(\sigma_S V_S + \sigma_{YS} V_Y\right) dW_S(t) \\ &\quad + \left(-\sigma_B V_S + (\sigma_{YB} - \sigma_B)\, V_Y\right) dW_B(t) \ . \end{aligned} \tag{10.31}$$

See Merton [203, Chapter 5] text for more justification.

Note that (10.30) does not really follow the Itô stochastic calculus, but states that the absolute return on the portfolio is the number of shares weighted sum of the absolute returns on the portfolio assets. However, in [203, Chapter 5], Merton argues that the missing differential product terms, such as $dN_S(t)S(t)$ and $dN_S(t)dS(t)$, represent consumption or external gains to the portfolio, which would violate the self-financing assumption making the portfolio open rather than closed to just the three assets.

- **Assumption BSM11. Investor Hedging the Portfolio to Eliminate Volatility:**
  Since many investors as individuals or as a group act to avoid stochastic effects, they tune or hedge their trading strategy, as a protection against losses, by removing volatility through removing the coefficients of the stock and bond

fluctuations. A main purpose of the stock and bond underlying the option in the portfolio is to give sufficient flexibility to leverage or hedge the stock and bond assets to remove volatilities that would not be possible with the option alone. Hence, setting the coefficients of $dW_S(t)$ and $dW_B(t)$, respectively, to zero in (10.31),

$$(\sigma_S V_S^* + \sigma_{YS} V_Y^*) = 0 \tag{10.32}$$

$$-\sigma_B V_S^* + (\sigma_{YB} - \sigma_B)\, V_Y^* = 0\,. \tag{10.33}$$

The optimal system (10.32,10.33) has a non-trivial solution for the optimal values (V_S^*, V_B^*) provided the system is singular, i.e., the determinant of the system is zero,

$$0 = \mathrm{Det}\begin{bmatrix} \sigma_S & \sigma_{YS} \\ -\sigma_B & \sigma_{YB} - \sigma_B \end{bmatrix} = \sigma_S(\sigma_{YB} - \sigma_B) + \sigma_{YS}\sigma_B\,, \tag{10.34}$$

which leads to the **Merton volatility fraction**

$$\frac{\sigma_{YS}}{\sigma_S} = -\frac{\sigma_{YB} - \sigma_B}{\sigma_B}\,, \tag{10.35}$$

provided $\sigma_S \neq 0$ and $\sigma_B \neq 0$. The single optimal option-stock value relation that makes it work,

$$V_S^* = -\frac{\sigma_{YS} V_Y^*}{\sigma_S}\,, \tag{10.36}$$

recalling budget constraint on V_B^*, giving

$$V_B^* = -\,(V_S^* + V_Y^*) = -\left(1 - \frac{\sigma_{YS}}{\sigma_S}\right) V_Y^*\,. \tag{10.37}$$

In the case of the non-stochastic, constant rate bond process, as in the more traditional Black-Scholes model, $\mu_b = r$ and $\sigma_B = 0$, so $\sigma_{YB} = 0$ and the option price is assumed to be independent of the bond price B, i.e., $F = F(S(t), t; T, K)$ and $F_B \equiv 0$. Then only the optimal values (10.36) are obtained, i.e., there is no Merton volatility fraction in the traditional Black-Scholes model.

However, taking the Merton volatility fraction as valid and substituting in for the definitions of the option-stock volatility σ_{YS} and the option-bond volatility σ_{YB} from (10.25-10.26), respectively, then the option price turns out to be homogeneous [203] in S and B,

$$Y^* = Y_S^* S + Y_B^* B\,. \tag{10.38}$$

Since this result is based upon the Merton volatility fraction, it does not appear in the classical Black-Scholes model, the stock and bond dynamics no longer having common stochastic diffusion forms.

- **Assumption BSM12. Zero Expected Portfolio Return:**
  Further, to avoid arbitrage profits, the expected return must be zero as well. Thus, the coefficient of dt in (10.31) must be zero, aside from the assumption that $V_P(t) = 0$ would imply that $dV_P(t) = 0$, i.e.,

$$0 = (\mu_S - \mu_B)\, V_S^* + (\mu_Y - \mu_B)\, V_Y^* \tag{10.39}$$

$$= \left(-(\mu_S - \mu_B)\, \frac{\sigma_{YS}}{\sigma_S} + (\mu_Y - \mu_B) \right) V_Y^*, \tag{10.40}$$

  assuming $V_Y^* \neq 0$. Otherwise, there would be no option and no optimal values (10.36) that would follow from the Merton volatility fraction (10.35). This means that the portfolio returns are hedged to complete equilibrium, deterministically and stochastically. Thus, provided the option value $V_Y^* \neq 0$, then, by setting the coefficient of V_Y^* in (10.39) to zero, **Merton's Black-Scholes fraction** becomes simply **Merton's fraction** for the expected returns, i.e.,

$$\frac{\mu_Y - \mu_B}{\mu_S - \mu_B} = \frac{\sigma_{YS}}{\sigma_S}\ . \tag{10.41}$$

  Since it does not involve either of the bond related volatilities, σ_B or σ_{YB}, this primary Merton fraction holds for the Black-Scholes model as well. The Black-Scholes fraction (10.41) states that the net drift ratio equals the option-stock volatility ratio, where the net drift is relative to the market interest/discount rate μ_B.

10.2.1 PDE of Option Pricing

In order to derive the partial differential equation (PDE) of Black-Scholes-Merton option pricing, the definition of the option expected return μ_Y in (10.24) is viewed as a PDE for the option price function with the option trajectory $Y(t)$ replaced by the composite function equivalent $F(S, B, t; T, K)$ as a function of three independent variables (S, B, t), (S, B) having replaced the underlying state trajectories $(S(t), B(t))$,

$$\mu_Y F \equiv F_t + \mu_S S F_S + \mu_B B F_B \tag{10.42}$$

$$+\frac{1}{2}\left(\sigma_S^2 S^2 F_{SS} + 2\rho\sigma_S\sigma_B SB F_{SB} + \sigma_B^2 B^2 F_{BB}\right)\ .$$

It is conceptually important to separate the view of S and B as deterministic, independent PDE variables and the view of $S(t)$ and $B(t)$ as the random SDE state trajectories in time, but use each view in the appropriate place.

Next, μ_Y is eliminated using the Black-Scholes fraction (10.41) with $\mu_Y = \mu_B + (\mu_S - \mu_B)\sigma_{YS}/\sigma_S$ and the option-stock induced volatility σ_{YS} is eliminated using its definition in (10.25), i.e., $\sigma_{YS} = \sigma_S S F_S / F$, while the option price F can be eliminated by Merton's homogeneous condition (10.38) with Y replaced by F,

$$F = S F_S + B F_B\ ,$$

incidentally eliminating both first partials F_S and F_B, so

$$0 = F_t + \frac{1}{2}\left(\sigma_S^2 S^2 F_{SS} + 2\rho\sigma_S\sigma_B SBF_{SB} + \sigma_B^2 B^2 F_{BB}\right) . \tag{10.43}$$

This Merton's **PDE of option pricing** needs side conditions, final condition at the expiration time and boundary conditions in the asset variables, the PDE and conditions forming a final value problem (FVP). For the FVP, the natural time variable is the time-to-maturity or time-to-go $\tau = T - t$ and $F_t = -F_\tau$, so the backward formulated PDE (10.43) in forward time t can be written as a forward diffusion or parabolic PDE in backward time τ,

$$F_\tau = \frac{1}{2}\left(\sigma_S^2 S^2 F_{SS} + 2\rho\sigma_S\sigma_B SBF_{SB} + \sigma_B^2 B^2 F_{BB}\right) . \tag{10.44}$$

It is conceptionally important to remember that the **PDE problem**, (10.44) plus any final and boundary conditions, is a deterministic problem in realized independent variables $(S, B, t = T - \tau)$ all stochasticity being eliminated, in contrast to the **SDE problem** in the stochastic path variables $(S(t), B(t), Y(t))$ which depends on the independent variable t and underlying stochastic diffusion processes.

In the classical Black-Scholes model, the bond price has no volatility $\sigma_B(t) = 0$, so the Merton homogeneous result (10.38) does not hold since it is based upon the Merton volatility fraction which is invalid if $\sigma_B(t) = 0$. Thus, starting back at the view of the definition of μ_Y as a PDE (10.42) setting all B partial derivatives to zero, but eliminating μ_Y using the Black-Scholes fraction (10.41) and σ_{YS} using (10.25), letting the option price function in backward time be defined as

$$\widehat{F}(S, \tau; T, K) \equiv F(S, T - \tau; T, K) ,$$

leads to Merton's **Black-Scholes option pricing PDE**, including a bond term,

$$\widehat{F}_\tau = \frac{1}{2}\sigma_S^2 S^2 \widehat{F}_{SS} + \mu_B(S\widehat{F}_S + B\widehat{F}_B - \widehat{F}) . \tag{10.45}$$

If the assumption that the mean interest/discount rate is the constant market rate, $\mu_B = r$ along with constant stock volatility σ_S, then the standard Black-Scholes option pricing PDE is obtained.

However, many texts do not use Merton's elaborate assumptions, that we have decomposed into a larger number of individual assumptions here, so these texts use a different hedging argument to produce the Black-Scholes PDE and the constant rate coefficient r. Dropping the zero aggregate assumption, the portfolio value is then

$$V_P(t) = N_S(t)S(t) + N_Y(t)Y(t) \tag{10.46}$$

in terms of the number of shares times the price per share for the option and the underlying stock. Similarly, the change in the portfolio value is given by the budget equation,

$$dV_P(t) = N_S(t)dS(t) + N_Y(t)dY(t), \tag{10.47}$$

ignoring the missing differential forms as in Merton's more general version. Upon eliminating the resultant stochastic terms to form a riskless portfolio, the coefficients of $dW_S(t)$, again yields the stock-option relationship, relating the number of stock shares to that of the options

$$N_S = -N_Y \widehat{F}_S \ , \tag{10.48}$$

called **delta hedging** since $\Delta \equiv \partial \widehat{F}/\partial S$ is called the **Delta** of the option [283], where the definition of σ_{YS} in 10.25) has been used.

Thus,

$$V_P = N_Y \cdot (F - S\widehat{F}_S),$$

where the process $Y(t)$ has been replaced by the composite function definition $Y = F$ in (10.20), and

$$dV_P = N_Y \cdot \left(-\widehat{F}_\tau + \frac{1}{2} S^2 \widehat{F}_{SS} \right) dt \ .$$

Finally, it is assumed that the portfolio will earn at the riskless rate, avoiding arbitrage profits without risk,

$$dV_P(t) = rV_p(t)dt \tag{10.49}$$

which upon eliminating V_P and dV_P leads to the **Black-Scholes option pricing PDE**,

$$\widehat{F}_\tau = \frac{1}{2}\sigma_S^2 S^2 \widehat{F}_{SS} + r(S\widehat{F}_S - \widehat{F}) \ , \tag{10.50}$$

independent of N_Y as long as $N_Y \neq 0$ and, as typically written, no longer including the bond term as in Merton's version (10.45).

The Black-Scholes option pricing equation (10.45) is a parabolic or diffusion PDE in two asset values, S and B, but degenerate in B since there is no diffusion term in B and only a drift or mean rate term $rB\widehat{F}_B$.

Two elementary solutions of (10.45) can easily be verified:

- Only a stock asset: $\widehat{F}(S, B, \tau; T, K) = S$.
- Only a deterministic bond asset: $\widehat{F}(S, B, \tau; T, K) = B(T)\exp(-r\tau)$.

10.2.2 Final and Boundary Conditions for Option Pricing PDE

In the case of the European call option, the final option price, for any value S of $S(T)$, satisfies the final option profit conditions given in (10.1) for calls or (10.2) for puts, translated directly as

$$\begin{aligned} F(S(T), B(T), T; T, K) &= \left\{ \begin{array}{l} \max[S(T) - K, 0], \text{ call} \\ \max[K - S(T), 0], \text{ put} \end{array} \right\} \\ &= \max[\theta(S(T) - K), 0] \ , \end{aligned} \tag{10.51}$$

where $\theta = 1$ for calls and $\theta = -1$ for puts. Since $S(T)$ and $B(T)$ are arbitrary but non-negative, we can replace them by the independent variables S and B respectively to form the final condition for the PDE,

$$F(S, B, T; T, K) = \max[\theta(S - K), 0] \,, \tag{10.52}$$

but we will return to the original form (10.51) when transforming to new variables.

For the other boundary conditions, the discussion will be simplified to the risk-free bond case, i.e., $\sigma_B(t) = 0$, as assumed in the classical Black-Scholes case (10.50), except that the time-dependent interest/discount rate, $\mu_B(t) = r(t)$, will be retained. In the case of risky bonds, the boundary conditions are given by diffusion PDEs instead of explicit functions or values, so solving the PDE (10.44) by computational methods, as in Chapter 8 or in [108, 230, 264, 283], is more practical.

The number of boundary conditions depends on the highest order partial derivative for each independent state variable in the PDE, one condition if it is first order and two conditions if it is second order. Thus, for (10.44) it is two boundary conditions in the stock and one in the bond. Time is not a state variable, but there is one final condition (technically an initial condition for the backward time variable τ) since the time derivative is first order.

At the zero stock price, $S = 0$, the Merton's Black-Scholes PDE (10.45) reduces to

$$\widehat{F}_\tau(0, B, \tau; T, K) = r(B\widehat{F}_B - \widehat{F}) \,, \tag{10.53}$$

upon setting S to zero in the coefficients, assuming the derivatives are bounded, which is a risky assumption prior to finding the solution. This is a first order PDE, all of which are classified as hyperbolic PDEs, and the usual method of constructing a solution is called the **method of characteristics** [251]. Noting that the PDE problem is a deterministic problem, the PDE (10.53) is compared to the deterministic (non-Itô!) chain rule for $\widetilde{F}(B, \tau) \equiv \widehat{F}(0, B, \tau; T, K)$,

$$d\widetilde{F} = \widetilde{F}_\tau d\tau + \widetilde{F}_B dB \,, \tag{10.54}$$

assuming that the differentials $d\tau$ and dB can be varied independently, the ordinary differential equations (ODEs) for the characteristic path are written maintaining relative proportions between the differentials of (10.54) and the corresponding coefficients of (10.53),

$$\frac{d\tau}{1} = -\frac{dB}{rB} = -\frac{d\widetilde{F}}{r\widetilde{F}} \,.$$

Solving these ODEs successively in pairs,

$$B = \widetilde{B}(\tau) = \kappa e^{-R(\tau)} \,, \tag{10.55}$$

where κ is a characteristic path constant of integration and the cumulative rate for time-dependent $r(t)$ is

$$R(\tau) \equiv \int_0^\tau r(T - s)ds \equiv \int_0^\tau \widehat{r}(s)ds \,, \tag{10.56}$$

and

$$\widetilde{F} = f(\kappa)e^{-R(\tau)} ,$$

where $f = f(\kappa)$ is an arbitrary function of integration depending on the constant κ from a prior integration. Using the first integral (10.55) to eliminate κ in favor of $\widetilde{B}$ and τ yields

$$\widetilde{F}(\widetilde{B}(\tau), \tau) = f\left(\widetilde{B}(\tau)e^{R(\tau)}\right)e^{-R(\tau)} . \tag{10.57}$$

It is not necessary to know much about the method of characteristics, since the reader can verify the solution by the usual substitution procedure. The arbitrary function f can be eliminated by applying the final condition (10.52) at $\tau = 0$ with $R(0) = 0$,

$$\widetilde{F}(\widetilde{B}(0), 0) = f(\widetilde{B}(0)) = F(0, \widetilde{B}(0), T; T, K) = \max[\theta(-K), 0] = 0.5(1-\theta)K .$$

Since $\widetilde{B}(0) = B(T)$ is considered arbitrary at this point, $f(\widetilde{B}) = 0.5(1-\theta)K$, a constant (*beware*: Merton [201] assumes $B(T) = 1$), leading to the complete particular solution

$$\widetilde{F}(B, \tau) = \widehat{F}(0, B, \tau; T, K) = 0.5(1-\theta)Ke^{-R(\tau)} , \tag{10.58}$$

independent of $B = \widetilde{B}(\tau)$. Note that $\widetilde{B}(\tau)$ is a deterministic path function of a deterministic **ODE problem** since it is derived from the deterministic PDE problem, (10.53) plus conditions, so is different from the stochastic path function $\widehat{B}(\tau)$ for the SDE problem, or more precisely the stochastic ODE problem. The boundary condition (10.58) corresponds to a boundary condition used by Wilmott [283] for finite differences applied to Black-Scholes type models.

However, since we cannot assume the partial derivatives are bounded for the full Merton model (10.44), we will only assume that the option price will be bounded in the limit of zero stock price:

$$\widehat{F}(S, B, \tau; T, K) \textbf{ is bounded as } S \to 0^+ . \tag{10.59}$$

For large S, it is more difficult to find the proper boundary condition. However, one heuristic choice is to assume that for large S the diffusion term will be exponentially small so the drift terms will dominate:

$$\widehat{F}_\tau \simeq r(S\widehat{F}_S + B\widehat{F}_B - \widehat{F}) . \tag{10.60}$$

As with the small stock price limit, the conjecture (10.60) needs to be verified for a solution. Again applying the method of characteristics to $\widetilde{F}(S, B, \tau) \equiv \widehat{F}(S, B, \tau; T, K)$, or checking by substitution, but with four variables, $\{\tau, B, S, \widehat{F}\}$, instead of three,

$$\frac{d\tau}{1} = -\frac{dB}{rB} = -\frac{dS}{rS} = -\frac{d\widehat{F}}{r\widehat{F}} .$$

Integration leads to three constants or functions of integration, two of which can be eliminated in favor of the independent variables S and B,

$$\widehat{F}(S, B, \tau; T, K) = g\left(Se^{R(\tau)}, Be^{R(\tau)}\right) e^{-R(\tau)} , \tag{10.61}$$

where $g = g(S\exp(R(\tau)), B\exp(R(\tau)))$ is an arbitrary function of integration obtained by integrating both the stock and bond characteristic ODEs effectively generating two constants of integration, and $R(\tau)$ is given in (10.56). Applying the final condition (10.52) when $S > K$ yields

$$\widehat{F}(S, B, 0; T, K) = \max[\theta(S - K), 0] = 0.5(1 + \theta)(S - K) ,$$

so that g is a constant function and the complete particular solution

$$\widehat{F}(S, B, \tau; T, K) \simeq 0.5(1 + \theta)(S - Ke^{-R(\tau)}) . \tag{10.62}$$

A similar boundary condition is also specified in Wilmott's [283] finite difference applications. However, it turns out we will not need this condition here, but the condition suggests that the option price will not be bounded as $S \to +\infty$.

The bond boundary condition or conditions are not as straightforward, since the final bond price per share does not appear explicitly in the final option profit formula. At the zero bond price, $B = 0$, the Black-Scholes PDE (10.50) reduces to

$$\widehat{F}_\tau(S, 0, \tau; T, E) = r(S\widehat{F}_S - \widehat{F}) + \frac{1}{2}\sigma_S^2 S^2 \widehat{F}_{SS} , \tag{10.63}$$

upon setting B to zero in the coefficients, assuming the derivatives are bounded. However, 10.63 is a diffusion equation rather than a boundary value, so there has been very little simplification of the original Black-Scholes PDE except the dimension has been reduced to one from two state variables. This may still be useful for computational methods. The reduction in dimension is similar for the Merton version (10.44) of the Black-Scholes option pricing PDE, the only difference is that the drift term is absent. For either PDE, setting $B = B(T)$ in the PDE leads to no simplification since $B(T)$ would be arbitrary. There is still hope, since Merton has a way of transforming away $B(T)$ analytically, but this transformation is modified here.

10.2.3 Transforming PDE to Standard Diffusion PDE

Since the underlying stock and bound price models are linear stochastic diffusion equations, the expectation is that the distribution of the option price should be somehow related to the log-normal distribution studied in Chapter 4. However, here we have two state variables instead of one, so it will be useful to get rid of the bond $\widehat{B}$ dependence since the dependence is so weak that the bond does not appear in the final condition. For this purpose, it is noted that the dimensions of $\widehat{B}$, $\widehat{S}$, $\widehat{F}$ and K are all in the price of dollars per share. Thus, according to **Buckingham's pi theorem** [42] of **dimensional analysis**, the solution can be put into the form

of intrinsic **dimension-less groups** collecting all powers (the pi's) of variables and parameters in the problem to eliminate any extraneous scalings. Two such groups for independent and dependent variables that lead to a self-similar solution without B are

$$x = \frac{\widehat{B}(0)S}{K \cdot B} \,, \tag{10.64}$$

$$G(x,\tau) = \frac{\widehat{B}(0)\widehat{F}(S,B,\tau)}{K \cdot B} \,, \tag{10.65}$$

where the scale factor $K \cdot \widehat{B}/\widehat{B}(0)$ is equivalent to Merton's [201] if we set the final bond price $\widehat{B}(0) = 1$ dollar per share. Note that if $y(\tau) = K \cdot \widehat{B}(\tau)/\widehat{B}(0)$ has the final payoff $y(0) = K$ which is the exercise price and $dy(\tau) = Kd\widehat{B}(\tau)/\widehat{B}(0)$. See Wilmott [283] for more on the use of similarity transformations in the financial context.

The partial derivatives of the proposed self-similar transformation to eliminate the bond explicitly are

$$\begin{aligned}
\frac{\partial x}{\partial S} = \frac{x}{S} \,,\ \frac{\partial x}{\partial B} &= -\frac{x}{B} \,, \\
S\widehat{F}_S = SG_x \,,\ B\widehat{F}_B &= -\frac{KB}{\widehat{B}(0)}(xG_x - G) \,,\ \widehat{F}_\tau = \frac{KB}{\widehat{B}(0)}G_\tau \,, \\
S^2\widehat{F}_{SS} &= \frac{KB}{\widehat{B}(0)}x^2G_{xx} \,, \\
SB\widehat{F}_{SB} &= -\frac{KB}{\widehat{B}(0)}x^2G_{xx} \,, \\
B^2\widehat{F}_{BB} &= \frac{KB}{\widehat{B}(0)}x^2G_{xx} \,.
\end{aligned}$$

Upon substitution of the PDE of option pricing (10.44), a singular diffusion equation is obtained with variable coefficients,

$$G_\tau(x,\tau) = \frac{1}{2}\widehat{\sigma}^2(\tau)x^2G_{xx} \,, \tag{10.66}$$

where

$$\widehat{\sigma}^2(\tau) = (\sigma_S^2 - 2\rho\sigma_S\sigma_B + \sigma_B^2)(T-\tau) \tag{10.67}$$

is a combined volatility term where all the volatilities on the right hand side are evaluated at the common argument of $(T-\tau)$, confirming the validity of the conjectures self-similar solution transformation to transform away the bond variable B, subject to consistent boundary and initial conditions. The boundedness boundary condition (10.59) as $S \to 0^+$ is

$$G(0^+,\tau) \textbf{ is bounded} \,. \tag{10.68}$$

As $S \to +\infty$, the option boundary condition should also be bounded for a put, but $O(S)$ for a call is expected. For the final condition it is helpful to consult the original forward form

$$F(S(T), B(T), T; T, K) = \max[\theta(S(T) - K), 0] \, ,$$

leading to

$$G(x, 0^+) = \max[\theta(x - 1), 0] \, , \tag{10.69}$$

where the factor $B(T) = \widehat{B}(0)$ washes out as by our proper scaling or by Merton's unscaled dollar bond. This completely justifies the assumption of a self-similar transformation heuristically, since it works.

Note that the diffusion PDE (10.66) has a variable diffusion coefficient that is quadratic in x and vanishes as $x \to 0^+$, so that the PDE is called a singular diffusion. However, we still have not transformed the backward time variable τ and we have not used a logarithm transformation like the one we used in Chapter 4. In order to obtain a standard diffusions PDE, with coefficient $\frac{1}{2}$, let

$$u = u(\tau) = \int_0^\tau \widehat{\sigma}^2(s) ds \, , \tag{10.70}$$

$$w = w(x, \tau) = \ln(x) + \frac{1}{2} u(\tau) \, , \tag{10.71}$$

$$G(x, \tau) = x \Phi(w(x, \tau), u(\tau)) \, , \tag{10.72}$$

combining several of Merton's [201] transformations. The new time variable u is a diffusion time that helps eliminate the correlation coefficient and other terms. The inverse of the independent variable logarithmic transformation is given by $x = exp(w - u/2)$, with the diffusion time correction. The new dependent variable

$$\Phi(w, u) = \frac{G(x, \tau)}{x} = \frac{\widehat{F}(S, B, \tau)}{S} \, ,$$

provided $S > 0$, is thus the dimensionless ratio of the option price $\widehat{F}$ to the stock price S, comprising another self-similar transformation and that transformation is common to both $\widehat{F}$ and G. Applying this transformation, being easier than the first, the standard diffusion equation is obtained,

$$\Phi_u(w, u) = \frac{1}{2} \Phi_{ww}(w, u) \, , \tag{10.73}$$

where $-\infty < w < +\infty$ and $0 = u(0) < u \leq u(T)$. The partial derivatives have the following meaning,

$$\Phi_u = \left(\frac{\partial \Phi}{\partial u} \right)_w \quad \text{and} \quad \Phi_{ww} = \left(\frac{\partial^2 \Phi}{\partial w^2} \right)_u \, .$$

On the other hand, the side conditions are not so standard with the final condition (10.69) at $\tau = 0$ for G being transformed to

$$\Phi(w, 0^+) = e^{-w} \max[\theta(e^w - 1), 0] = \max[\theta(1 - e^{-w}), 0] \, , \tag{10.74}$$

where the reader should confirm that this is correct in all cases, since it is generally not correct to bring a variable into a maximum argument. However, for the boundary condition a singular limit is avoided by keeping the x factor multiplying $\Phi(w,u)$, so

$$x(w,u)\Phi(w,u) \tag{10.75}$$

should bounded as $w \to -\infty$ when $x \to 0^+$.

The solution of (10.73) can be written in terms of the complementary error function erfc or the normal distribution Φ_n, but they are related through several identities, two of which are in (B.20,B.21). Merton [201], [203, Chapter 5] uses erfc, while Black and Scholes [34] use the standard normal distribution which in our notation is $\Phi_n(x;0,1)$. The simplest fundamental solution

$$\Phi_1(w,u) \equiv \Phi_n(w;0,u) \ , \tag{10.76}$$

of (10.73) can be derived using Fourier transform methods [102, Chapter 9] or can be derived using the self-similar solution technique used here earlier to remove the bond dependence. See also the introduction to the diffusion equation (B.26) in Chapter B. However, it may be much easier to verify

$$\Phi_{1,u}(w,u) = \frac{1}{2}\Phi_{1,ww}(w,u) \ , \tag{10.77}$$

using a symbolic computation system such as Maple™ or *Mathematica*™. The simple diffusion solution Φ_1 in (10.76) is just too simple and does not satisfy the final condition (10.74) at $u = 0^+$ which can be written in terms of either the standard unit step function $H(x)$ in (B.158) or the averaged unit step function $H_a(x)$ in (B.159)

$$\Phi(w,0^+) = \theta(1-e^{-w})H(\theta w) = \theta(1-e^{-w})H_a(\theta w). \tag{10.78}$$

Since either step function will do, the coefficient vanishes at $w = 0$, but instead the simple solution Φ_1 satisfies the final condition,

$$\Phi_1(w,0^+) = H_a(w), \tag{10.79}$$

as the reader can verify by examining the cases $w > 0$, $w = 0$ and $w < 0$ as $u \to 0^+$.

Thus, another solution is needed to provide the extra variable factor e^{-w}. Specializing to the call option when $\theta = +1$, the second solution is

$$\Phi_2(w,u) \equiv e^{-w+u/2}\Phi_n(w;u,u) \ , \tag{10.80}$$

which can be shown to satisfy the standard diffusion equation (10.73) and a different final condition

$$\Phi_2(w,0^+) = e^{-w}H_a(w) \ , \tag{10.81}$$

than that of Φ_1 in (10.79). The boundedness condition (10.75) is trivial as $w \to -\infty$ since both Φ_1 and $e^{w-u/2}\Phi_2$ vanish by the definition of Φ_n with $u > 0$. Thus, the transformed solution for the call option price is

$$\begin{aligned}\Phi^{(call)}(w,u) &= \Phi_1(w,u) - \Phi_2(w,u) \\ &\equiv \Phi_n(w;0,u) - e^{-w+u/2}\Phi_n(w;u,u) \qquad (10.82)\\ &= \Phi_n\left(\frac{w}{\sqrt{u}};0,1\right) - e^{-w+u/2}\Phi_n\left(\frac{w-u}{\sqrt{u}};0,1\right)\ , \qquad (10.83)\end{aligned}$$

upon transforming to standard normal distributions. Thus, $\Phi^{(call)}$ satisfies the final condition,

$$\Phi^{(call)}(w,0^+) = (1-e^{-w})H_a(w)\ . \qquad (10.84)$$

The solution form resembles solutions of the diffusion equation on a semi-infinite domain found by the classical method of reflection. Transforming back to the original variables, one can compare to the original Black-Scholes form with $\mu_B = r$ and $\sigma_B^2 = 0$, so $\widehat{B}(\tau) = \widehat{B}(0)\exp(-r\tau)$ and $u = \sigma_S^2\tau$. The correlation term with ρ (10.18) vanishes with σ_B.

The European put option price solution ($\theta = -1$) is somewhat different, relying on normal distributions complementary to those of the the European call option with two component solutions,

$$\begin{aligned}\Phi^{(\mathrm{put})}(w,u) &= \Phi_3(w,u) - \Phi_4(w,u) \\ &\equiv e^{-w+u/2}\left(1-\Phi_n(w;u,u)\right) - \left(1-\Phi_n(w;0,u)\right) \qquad (10.85)\\ &= e^{-w+u/2}\left(1-\Phi_n\left(\frac{w-u}{\sqrt{u}};0,1\right)\right) \\ &\quad -\left(1-\Phi_n\left(\frac{w}{\sqrt{u}};0,1\right)\right)\ , \qquad (10.86)\end{aligned}$$

where again the final form is in terms of standard normal distributions. The reader can verify that $\Phi^{(\mathrm{put})}(w,u)$ satisfies the standard diffusion equation (10.73) and the put option price final condition,

$$\Phi^{(\mathrm{put})}(w,0^+) = (e^{-w}-1)H_a(-w) \qquad (10.87)$$

and the zero stock limit boundedness condition that $e^{w-u/2}\Phi^{(\mathrm{put})}(w,u)$ is bounded as $w \to -\infty$, not zero as in the call case, Maple or *Mathematica* being the preferred tools. The put and call option prices are related in a general way according to the principle of **put-call parity**, i.e., in transformed variables,

$$\Phi^{(\mathrm{put})}(w,u) - \Phi^{(call)}(w,u) = \exp(-w+u/2) - 1.$$

See also Exerecise 6 on p. 346.

The boundary condition limits of the solutions essentially follow from the corresponding extreme limits of the normal distribution function,

$$\Phi_n(w;\mu,\sigma^2) \to \left\{ \begin{array}{ll} 0, & w \to -\infty \\ 1, & w \to +\infty \end{array} \right\}\ , \qquad (10.88)$$

except in one case. Thus, for the intermediate transformed call option price multiplied by the transformed stock option $x\Phi^{(call)}$ using (10.82) for $\Phi^{(call)}(w,u)$ satisfies the limiting conditions,

$$x\Phi^{(call)}(w,u) \to \left\{ \begin{array}{cc} 0, & w \to -\infty \ \& \ x \to 0^+ \\ x-1, & w \to +\infty \ \& \ x \to +\infty \end{array} \right\} , \tag{10.89}$$

consistent with the derived limits (10.58) for Black-Scholes call and put option pricing in the boundary conditions subsection. The put option price is formulated in terms of the complementary normal probability distribution, $1-\Phi_n(w;\mu,\sigma^2)$ which vanishes exponentially as $w \to +\infty$ and $x \to +\infty$, so results in an indeterminate form, $\infty \cdot 0$, for $x\Phi^{(\mathrm{put})}$. However, this form can be resolved using **l'Hospital's rule** and the fact that $x = \exp(w - u/2)$,

$$x\left(1-\Phi_n(w;\mu,\sigma^2)\right) = \frac{1-\Phi_n(w;\mu,\sigma^2)}{e^{-w+u/2}} \to \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(w-m)^2/(2\sigma^2)+w} \to 0 \ ,$$

since the larger degree monomial in the exponent dominates the smaller one. Finally, the put option price extreme conditions are

$$x\Phi^{(\mathrm{put})}(w,u) \to \left\{ \begin{array}{cc} 1-x, & w \to -\infty \ \& \ x \to 0^+ \\ 0, & w \to +\infty \ \& \ x \to +\infty \end{array} \right\} , \tag{10.90}$$

again consistent with prior derived limits (10.62) for Black-Scholes call and put option pricing. Note that the extreme boundary conditions strongly reflect the final condition.

Reversing the transformations used to convert the answers $\Phi^{(call)}(w,u)$ (10.82-10.83) for the call option price and $\Phi^{(\mathrm{put})}(w,u)$ (10.85-10.86) for the put option price back to the actual option price $Y(t) = F(S(t),B(t),t;T;K)$ is left as Exercise 5 at the end of this chapter.

While the put option pricing results are not in Merton's continuous returns paper [201] ([203, Chapter 8]), there are many other results and more exploration with the removal of assumptions, such as the no dividends exclusion. In his companion discontinuous returns paper [202] ([203, Chapter 9]), Merton presents one of the first treatments of jump-diffusions in finance.

10.3 Jump-Diffusion Option Pricing

Since the 1973 Black-Scholes-Merton option pricing model is based upon the pure diffusion stochastic model, there is one obvious missing feature that large market fluctuations or jumps such as crashes or rallies which characterize extreme market psychology are not represented. There are several papers on the statistical importance for including jumps in financial market models, e.g., see Ball and Torous [18] on stocks and options, Jarrow and Rosenfeld [153]on the capital asset pricing model (CPAM) or Jorion [155] on foreign exchange and stocks.

There are other qualitative features that characterize real market log-return distributions that can not be reproduced by the pure diffusion model of Black-Scholes-Merton, but can be modeled, in part, by adding jumps to the diffusion

process. One feature is that real markets have negatively skewed log-return distributions, provided a sufficient number of years of daily return data is used [127], so that the log-return skewness coefficient (B.11),

$$\eta_3[X] \equiv \frac{\mathrm{E}[(X - \mathrm{E}[X])^3]}{(\mathrm{Var}[X])^{3/2}} < 0,$$

where

$$X = \Delta \ln(S(t_i)) = \ln(S(t_{i+1})) - \ln(S(t_i))$$

is the log-return for trading day t_{i+1} for $i = 1 : n_s - 1$ trading days, while $\eta_3[X] = 0$ for the intrinsically skewless normally distributed log-return model on the pure diffusion process. Hence, real markets in the long run are found to be pessimistic due to more negative log-returns, including crashes, than positive log-returns.

Another feature is that real market distributions are found to be leptokurtic so that the log-return kurtosis coefficient (B.12),

$$\eta_4[X] \equiv \frac{\mathrm{E}[(X - \mathrm{E}[X])^4]}{(\mathrm{Var}[X])^2} > 3,$$

for $X = \Delta \ln(S(t_i)) = \ln(S(t_{i+1})) - \ln(S(t_i))$, whereas the the normally distributed pure diffusion process is **mesokurtic** (also said to have zero excess kurtosis. $\eta_4[X] - 3$) since $\eta_4[X] = 3$. Leptokurtic means that the distribution is more peaked (kurtic is derived from the word for crown) at the maximum and consequently has fatter tails than the normal distribution.

Still another characteristic if the **volatility smile** which refers to the curvature of the implied volatility, volatility implied by the log-normal Black-Scholes formula, versus the strike price. For more information on volatility smiles and their relation to non-log-normal distributions which they signify, see such references as Hull [147].

Merton [202] in 1976 pioneered the analysis of option pricing for stock returns governed by a jump-diffusion model. Merton chose the normal distribution for the jump-amplitude distribution for the log-return. Here, the option pricing with jump-diffusions is described in terms of the jump-diffusion formulations in this book. The stock price is assumed to be subject to extreme changes over a very short period of time due to significant changes in the firm or in the market. Further details can be found in Zhu [289], and Zhu and Hanson [290]. Thus, consider the jump-diffusion model for the stock price $S(t)$ at time t,

$$dS(t) = S(t)(\mu_d dt + \sigma_d dW(t) + J(Q)dP(t)) \ , \quad S(0) = S_0 > 0 \ , \qquad (10.91)$$

where μ_d and σ_d are designated as the diffusion parameters for the standard diffusion $dW(t)$, while $J(Q) = \exp(Q) - 1 > -1$ is the jump-amplitude for the jumps of Poisson process $dP(t)$, such that the symbolic jump from means

$$J(Q)dP(t) = \sum_{k=1}^{dP(t)} J(Q_k),$$

for integers $k \geq 1$, otherwise the sum is zero if $k = 0$ and where the marks Q_k are IID normally distributed. Note, unlike Merton in [202], in (10.91) there are not the same notational shifts in the diffusion drift and jump amplitude, so that $\mathrm{E}[dS(t)/S(t)] = \mu_d dt + \lambda dt \mathrm{E}[J(Q)]$ and Merton's $Y - 1$ is the same as $J(Q)$ while $\alpha = \mu_d$.

By the stochastic chain rule the log-return satisfies

$$d\ln(S(t)) = \mu_{ld} dt + \sigma_d dW(t) + \sum_{k=1}^{dP(t)} Q_k \ , \tag{10.92}$$

when $\mu_{ld} = \mu_d - 0.5\sigma_d^2$ is the diffusion-corrected mean appreciation coefficient. Under the assumption of constant coefficients, the solution of (10.92) is immediate,

$$S(t) = S_0 \exp(\mu_{ld} t + \sigma_d W(t) + \sum_{k=1}^{P(t)} Q_k) \ . \tag{10.93}$$

The solution is positive as long as $S_0 > 0$ and Q_k is assumed real, a consequence of the **geometric jump-diffusion** assumptions.

10.3.1 Jump-Diffusions with Normal Jump-Amplitudes

Since the marks Q_k are independent and identically distributed normally, the mark density is defined in our notation as

$$\phi_Q(q) = \phi_n(q; \mu_j, \sigma_j^2) \tag{10.94}$$

where ϕ_n denotes a normal density with mean $\mu_j = \mathrm{E}[Q]$ and variance $\sigma_j^2 = \mathrm{Var}[Q]$.

If the discrete version

$$\begin{aligned} \Delta\ln(S(t)) &= \mu_{ld}\Delta t + \sigma_d \Delta W(t) + \sum_{k=1}^{\Delta P(t)} Q_k \\ &= \mu_{ld}\Delta t + \sigma_d \Delta W(t) + \mu_j(\Delta P(t) - \lambda \Delta t) + \sum_{k=1}^{\Delta P(t)} (Q_k - \mu_j) \ , \end{aligned} \tag{10.95}$$

of the log-return SDE (10.92) is used to approximate the log-return difference,

$$\Delta\ln(S(t)) \equiv \ln(S(t + \Delta t) - \ln((S(t)),$$

where the last line of (10.95) has the stochastic terms collected into independent and zero mean forms to facilitate moments calculations. The standard moments (mean plus central moments for higher moments) can be calculated (See Theorem 5.17 on p. 151, [131] and [289]) using (10.95 rather than the solution (10.93). Thus,

$$\begin{aligned} M_1 &\equiv \mathrm{E}[\Delta\ln(S(t))] = (\mu_{ld} + \lambda\mu_j)\Delta t, \\ M_2 &\equiv \mathrm{Var}[\Delta\ln(S(t))] = (\sigma_d^2 + \lambda(\mu_j^2 + \sigma_j^2))\Delta t, \\ M_3 &\equiv \mathrm{E}[(\Delta\ln(S(t)) - M_1)^3] = \lambda\mu_j(\mu_j^2 + 3\sigma_j^2)\Delta t, \\ M_4 &\equiv \mathrm{E}[(\Delta\ln(S(t)) - M_1)^4] = \lambda(\mu_j^4 + 6\mu_j^2\sigma^2 + 3\sigma_j^4)\Delta t + 3(\sigma_d^2 + \lambda(\mu_j^2 + \sigma_j^2)^2\Delta t^2. \end{aligned}$$

The variance normalized third moment is the skewness coefficient,

$$\eta_3[\Delta \ln(S(t))] = \frac{\lambda \mu_j (\mu_j^2 + 3\sigma_j^2)}{(\sigma_d^2 + \lambda(\mu_j^2 + \sigma_j^2))^{3/2} (\Delta t)^{1/2}}, \tag{10.96}$$

so $\eta_3[\Delta \ln(S(t))] < 0$ if the log-normal jump-amplitude mean $\mu_j < 0$, since the jump rate λ must be positive for there to be jumps. The variance normalized fourth moment is the kurtosis coefficient,

$$\eta_4[\Delta \ln(S(t))] = \frac{\lambda(\mu_j^4 + 6\mu_j^2\sigma_j^2 + 3\sigma_j^4)}{(\sigma_d^2 + \lambda(\mu_j^2 + \sigma_j^2))^2 \Delta t} + 3, \tag{10.97}$$

so the $\eta_4[\Delta \ln(S(t))] > 3$ provided $\mu_j \neq 0$ and $\sigma_j \neq 0$. Thus, the jump-diffusion with log-normally distributed jump amplitudes provides more realistic log-return distribution properties with skewness whose direction depends on the sign of the mark mean μ_j and leptokurtosis for nontrivial mark distributions.

Another advantage, particularly in analysis, follows from the convolution result that the sum of normals is normally distributed. This is expressed in Corollary 5.21 on p. 157, so for the jump-diffusion with log-normally distributed jump-amplitude, the density with a small modification for the difference and constant coefficients is given as an infinite sum of translated normal densities over all Poisson jumps by

$$\phi_{\Delta \ln(S(t))}(x) = \sum_{k=1}^{\infty} p_k(\lambda \Delta t) \phi_n(x; \mu_{ld} \Delta t + k\mu_j, \sigma_d^2 \Delta t + k\sigma_j^2)\,, \tag{10.98}$$

where $p_k(\lambda \Delta t)$ is the Poisson distribution (B.50) with parameter $\lambda \Delta t$ and $\phi_n(x; \mu, \sigma^2)$ denotes the normal density with general parameters μ and σ^2.

10.3.2 Risk-Neutral Option Pricing for Jump-Diffusions

Rather than follow Merton's 1976 [202] paper to directly explain his approach using the PDE formulation of the previous section, we will approach the option pricing in the presence of both diffusion and jumps by directly applying a **risk-neutral** assumption that the discounted earnings on a European call option is at the existing market rate r, i.e., the risk-neutral call option price has the form

$$\mathcal{C}^{(\mathrm{rn})}(S_0, T) \equiv e^{-rT} \mathrm{E}^{(\mathrm{rn})}[\max[S(T) - K, 0]]\,, \tag{10.99}$$

where T is the option exercise time, K is the strike price, $\exp(-rT)$ is the discount factor and $\mathrm{E}^{(\mathrm{rn})}$ denotes the risk-neutral expected value [147, pp. 248-250], depending on the initial asset price S_0 as well.

As Merton points out, the classical Black-Scholes hedge or the **delta hedge** (10.7) is no longer sufficient to eliminate all risk when there are jumps in the underlying asset price that result in non-marginal changes. This could come from non-systematic information about the firm to cause extreme changes in value. There are special cases that are of little interest and there is always the possibility of using

the Black-Scholes hedge to eliminate the diffusive **volatility-risk** during the quiet period between jumps, but when a jump event arrives there is the possibility of a large loss or other unexpected change in value of the option, i.e., the so-called **jump-risk** will not be covered. In short, there are too many random variables in a jump-diffusion to **delta hedge** away with a single stock. For instance, in the compound Poisson process there is the pure counting part of the process and then there is the uncountable IID log-jump-amplitudes or marks Q_k that would need to be hedged.

Letting

$$\mu_J \equiv \mathrm{E}[J(Q)] = \int_{-\infty}^{+\infty} \phi_n(q)\,(e^q - 1)\,dq = e^{\mu_j + 0.5\sigma_j^2} - 1 \tag{10.100}$$

be the mean jump-amplitude of the asset price, then the mean asset price at the strike time T using iterated expectations on the closed form solution (10.93),

$$\begin{aligned}
\mathrm{E}[S(T)] &= S_0 e^{(\mu_d - 0.5\sigma_d^2)T} \mathrm{E}\left[e^{\sigma_d W(T)} e^{\sum_{k=1}^{N(T)} Q_k}\right] \\
&= S_0 e^{(\mu_d - 0.5\sigma_d^2)T} \mathrm{E}_{W(T)}\left[e^{\sigma_d W(T)}\right] \mathrm{E}_{N(T)}\left[\prod_{k=0}^{N(T)} \mathrm{E}_{Q_k|N(T)}\left[e^{Q_k}\,\middle|\,N(T)\right]\right] \\
&= S_0 e^{(\mu_d - 0.5\sigma_d^2)T} e^{0.5\sigma_d^2 T} \mathrm{E}_{N(T)}\left[\prod_{k=1}^{N(T)} (\mu_J + 1)\right] \\
&= S_0 e^{\mu_d T} \mathrm{E}_{N(T)}\left[(\mu_J + 1)^{N(t)}\right] = S_0 e^{\mu_d T} \sum_{k=1}^{\infty} p_k(\lambda T)(\mu_J + 1)^k \\
&= S_0 e^{\mu_d T} e^{-\lambda T} \sum_{k=1}^{\infty} (\lambda T(\mu_J + 1))^k / k! = S_0 e^{\mu_d T + \lambda T \mu_J} \\
&= S_0 e^{(\mu_d + \lambda\mu_J)T}\,,
\end{aligned} \tag{10.101}$$

where the IID property of the Q_k and the Poisson distribution $p_k(\lambda T)$ (B.50) have also been used.

In the risk-neutral world (see Hull [147, pp. 248-250]) then

$$\mathrm{E}[S(T)] = S_0 e^{(\mu_d + \lambda\mu_J)T} = S_0 e^{rT},$$

so the jump diffusion rate in a risk-neutral world must be

$$\mu_d + \lambda\mu_J = r,$$

the sum of the diffusive and jump mean rates. For consistency with the benchmark Black-Scholes model, this relation will be used to eliminate the diffusive mean rate in a risk-neutral world

$$\mu_d = \mu_d^{(\mathrm{rn})} \equiv r - \lambda\mu_J, \tag{10.102}$$

allowing the following formulation of the risk-neutral option.

Definition 10.1. ***Jump-Diffusion Risk-Neutral European Call Option:*** *Applying the general jump-diffusion solution (10.93), with (10.102) to the risk-neutral European call option payoff (10.99) and in terms of the general jump partial sum random variable*

$$\widehat{\mathcal{S}}_k = \sum_{i=1}^{k} Q_i$$

with density $\phi_{\widehat{\mathcal{S}}_k}(s_k)$), yields the form,

$$\begin{aligned}
\mathcal{C}^{(\mathrm{rn})}(S_0,T) \equiv\ & e^{-rT}\mathrm{E}^{(\mathrm{rn})}[\max[S(T)-K,0]] \\
\equiv\ & e^{-rT}\mathrm{E}\Big[\max\Big[S_0 e^{(r-\lambda\mu_J-\sigma_d^2/2)T+\sigma_d W(T)+\sum_{k=1}^{P(T)}} - K,0\Big]\Big] \\
=\ & e^{-rT}\textstyle\sum_{k=0}^{\infty} p_k(\lambda T)\int_{-\infty}^{+\infty} dw\ \phi_n(w;0,T)\int_{-\infty}^{+\infty} ds_k \phi_{\widehat{\mathcal{S}}_k}(s_k) \\
& \cdot\max\Big[S_0 e^{(r-\lambda\mu_J-\sigma_d^2/2)T+\sigma_d w+s_k} - K,0\Big] \\
=\ & e^{-rT}\textstyle\sum_{k=0}^{\infty} p_k(\lambda T)\int_{-\infty}^{+\infty} dw\ \phi_n(w;0,T) \\
& \cdot\mathrm{E}_{\widehat{\mathcal{S}}_k}\Big[\max\Big[S_0 e^{(r-\lambda\mu_J-\sigma_d^2/2)T+\sigma_d w+\widehat{\mathcal{S}}_k} - K,0\Big]\Big],
\end{aligned} \tag{10.103}$$

Remark 10.2. *The random sum $\widehat{\mathcal{S}}_k$ is used here, rather that the mark Q_i as in (10.101), since the maximum function in (10.103) needs a different splitting of the expectations.*

Theorem 10.3. ***Risk-Neutral Call Prices as an Infinite Poisson Sum of Shifted Black-Scholes Call Prices – General Jump-Diffusion Case:*** *For the general jump-diffusion,*

$$\mathcal{C}^{(\mathrm{rn})}(S_0,T) = \sum_{k=0}^{\infty} p_k(\lambda T)\mathrm{E}_{\widehat{\mathcal{S}}_k}\Big[\mathcal{C}^{(\mathrm{bs})}\Big(S_0 e^{\widehat{\mathcal{S}}_k-\lambda\mu_J T},T;K,\sigma_d^2,r\Big)\Big], \tag{10.104}$$

where the Black-Sholes call price function

$$\mathcal{C}^{(\mathrm{bs})}(s,T;K,\sigma_d^2,r) = \ s\Phi\left(d_1(s)\right) - Ke^{-rT}\Phi\left(d_2(s)\right), \tag{10.105}$$

$$\Phi(x) \equiv \Phi_n(x;0,1) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy \tag{10.106}$$

is the standardized normal distribution,

$$d_1(s) = \left(\ln(s/K) + (r+\sigma_d^2/2)T\right)/(\sigma_d\sqrt{T}) \ \text{ and } \ d_2(s) = d_1(s) - \sigma_d\sqrt{T} \tag{10.107}$$

are Black-Scholes normal argument functions.

Proof. Note that the argument of the last maximum in (10.103) has a root at $w = w_0(s_k)$ when

$$S_0 e^{(r-\lambda\mu_J-\sigma_d^2/2)T+\sigma_d w+s_k} = K$$

or when

$$w_0(s_k) = (\ln(K/S_0) - (r - \lambda\mu_J - \sigma_d^2/2)T + s_k)/\sigma_d, \tag{10.108}$$

allowing the removal of the maximum function. Some further manipulations with the normal integrals permits the transformation to an infinite Poisson sum over Black-Scholes call functions with shifted arguments,

$$\begin{aligned}
\mathcal{C}^{(\mathrm{rn})}(S_0,T) &= e^{-rT}\sum_{k=0}^{\infty} p_k(\lambda T)\mathrm{E}_{\widehat{\mathcal{S}}_k}\left[\int_{w_0(\widehat{\mathcal{S}}_k)}^{+\infty} dw\phi_n(w;0,T)\right. \\
&\qquad \left.\cdot\left(S_0 e^{(r-\lambda\mu_J-\sigma_d^2/2)T+\sigma_d w+\widehat{\mathcal{S}}_k} - K\right)\right], \\
&= \sum_{k=0}^{\infty} p_k(\lambda T)\mathrm{E}_{\widehat{\mathcal{S}}_k}\left[S_0 e^{\widehat{\mathcal{S}}_k-\lambda\mu_J T}A(S_0 e^{\widehat{\mathcal{S}}_k-\lambda\mu_J T}) - Ke^{-rT}B(S_0 e^{\widehat{\mathcal{S}}_k-\lambda\mu_J T})\right]
\end{aligned}$$

where the intermediate functions $A(s)$ and $B(s)$ are derived below.

$$\begin{aligned}
A(s) &= e^{-\sigma_d^2 T/2}\textstyle\int_{w_0(s)}^{\infty} dw\ \phi_n(w;(0,T)e^{\sigma_d w} \\
&= e^{-\sigma_d^2 T/2}\tfrac{1}{\sqrt{2\pi T}}\textstyle\int_{w_0(s)}^{\infty} dw\ \epsilon^{-w^2/(2T)+\sigma_d w} \\
&= e^{-\sigma_d^2 T/2}\tfrac{1}{\sqrt{2\pi T}}\textstyle\int_{w_0(s)}^{\infty} dw\ \epsilon^{-(w-\sigma_d T)^2/(2T)} \\
&= \tfrac{1}{\sqrt{2\pi}}\textstyle\int_{(w_0(s)-\sigma_d T)/\sqrt{T}}^{\infty} dy\ \epsilon^{-y^2/2} \\
&= \left(1-\Phi\left((w_0(s)-\sigma_d T)/\sqrt{T}\right)\right) \\
&= \Phi\left((\sigma_d T - w_0(s))/\sqrt{T}\right) \\
&= \Phi\left(d_1\left(S_0 e^{s-\lambda\mu_J T}\right)\right),
\end{aligned}$$

since by (10.108) and (10.107), $(\sigma_d T - w_0(s))/\sqrt{T} = d_1(S_0 e^{s-\lambda\mu_J T})$. The simpler second argument quickly follows from

$$\begin{aligned}
B(s) &= \textstyle\int_{w_0(s)}^{\infty} dw\ \phi_n(w;(0,T) = \Phi\left(-w_0(s)/\sqrt{T}\right) \\
&= \Phi\left(d_1\left(S_0 e^{s-\lambda\mu_J T}\right) + \sigma_d\sqrt{T}\right) = \Phi\left(d_2\left(S_0 e^{s-\lambda\mu_J T}\right)\right),
\end{aligned}$$

using (10.108) and (10.107) again. Reassembling $A(s)$ and $B(s)$ from the current equation for $\mathcal{C}^{(\mathrm{rn})}$ yields (10.105) from the relation

$$C^{(\mathrm{bs})}(s,T;K,\sigma_d^2,r) = sA(s) - Ke^{-rT}B(s),$$

and thus (10.104) follows. □

Remark 10.4. *The primary argument s of $\mathcal{C}^{(\mathrm{bs})}$ is shifted for each jump number k by a factor $\exp(\widehat{\mathcal{S}}_k - \lambda\mu_J T)$ that depends only on the jump process (the result in this form is valid for general jump-diffusions as treated in this book).*

If the mark density is normal, $\phi_Q(q) = \phi_n(q; \mu_j, \sigma_j^2)$, then the European call option formula can be simplified.

Theorem 10.5. ***Risk-Neutral Call Prices as an Infinite Poisson Sum of Shifted Black-Scholes Call Prices – Log-Normal-Jump-Amplitude Jump-Diffusion Case:***
For the log-normal-jump-amplitude jump-diffusion,

$$\mathcal{C}_n^{(\mathrm{rn})}(S_0, T) = \sum_{k=0}^{\infty} p_k(\lambda T)\mathcal{C}_n^{(\mathrm{bs})}\Big(S_0 e^{k(\mu_j + \sigma_j^2/2) - \lambda\mu_J T}, T; K, \sigma_k^2(T)/T, r\Big), \quad (10.109)$$

where the Black-Sholes call price function

$$\mathcal{C}_n^{(\mathrm{bs})}(s, T; K, \sigma_k^2(T)/T, r) = s\Phi\Big(\widehat{d}_1\left(s; \sigma_k^2(T)\right)\Big) - Ke^{-rT}\Phi\Big(\widehat{d}_2\left(s, \sigma_k^2(T)\right)\Big), \quad (10.110)$$

$$\begin{aligned} \widehat{d}_1(s; \sigma_k^2(T)) &= \left(\ln(s/K) + rT + \sigma_k^2(T)/2\right)/\sigma_k(T), \\ \widehat{d}_2(s; \sigma_k^2(T)) &= d_1(s) - \sigma_k(T) \end{aligned} \quad (10.111)$$

are Black-Scholes normal argument functions, and

$$\sigma_k^2(T) = \sigma_d^2 T + k\sigma_j^2 \quad (10.112)$$

is the log-return variance.

Proof. In order to simplify the expectation calculations, let

$$X = \sigma_d W(T) + \left(\widehat{\mathcal{S}}_k - k\mu_j\right)$$

be the zero mean part of the risk-neutral log-return process obtained by subtracting the mean

$$\mu_k(T) = (r - \lambda_J - \sigma_d^2/2)T + k\mu_j \quad (10.113)$$

and leaving the variance (10.112),

$$\sigma_k^2(T) = \sigma_d^2 T + k\sigma_j^2,$$

so by the normal convolution corollary (10.98) the density is reduced to

$$\phi_X(x) = \phi_n(x; 0, \sigma_k^2(T)). \quad (10.114)$$

The payoff cutoff to remove the maximum function in the normal case then is

$$x_k(T) = \ln(K/S_0) - \mu_k(T).$$

Thus, the normal risk-neutral call price is derived using normal integral identities as follows,

$$\begin{aligned}
\mathcal{C}_n^{(\mathrm{rn})}(S_0,T) &= e^{-rT}\textstyle\sum_{k=0}^{\infty} p_k(\lambda T)\int_{-\infty}^{+\infty} dx\ \phi_n(x;0,\sigma_k^2(T))\max\left[S_0e^{\mu_k(T)+x}-K,0\right] \\
&= e^{-rT}\textstyle\sum_{k=0}^{\infty} p_k(\lambda T)\int_{x_k(T)}^{+\infty} dx\ \phi_n(x;0,\sigma_k^2(T))\left(S_0e^{\mu_k(T)+x}-K\right) \\
&= e^{-rT}\textstyle\sum_{k=0}^{\infty} p_k(\lambda T)\frac{1}{\sqrt{2\pi\sigma_k^2(T)}} \\
&\quad \cdot\Big(S_0e^{\mu_k(T)+\sigma_k^2(T)/2}\textstyle\int_{x_k(T)}^{+\infty} dx\ e^{-(x-\sigma_k^2(T))^2/(2\sigma_k^2(T))} \\
&\quad -K\textstyle\int_{x_k(T)}^{+\infty} dx\ e^{-x^2/(2\sigma_k^2(T))}\Big) \\
&= \textstyle\sum_{k=0}^{\infty} p_k(\lambda T)\left(S_0e^{\mu_k(T)-rT+\sigma_k^2(T)/2}\Phi\left(\frac{\sigma_k^2(T)-x_k(T)}{\sigma_k(T)}\right)\right. \\
&\quad \left.-Ke^{-rT}\Phi\left(\frac{-x_k(T)}{\sigma_k(T)}\right)\right) \\
&= \textstyle\sum_{k=0}^{\infty} p_k(\lambda T) \\
&\quad \cdot\Big(S_0e^{k(\mu_j+\sigma_j^2/2)-\lambda\mu_JT}\Phi\left(\widehat{d}_1\left(S_0e^{k(\mu_j+\sigma_j^2/2)-\lambda\mu_JT};\sigma_k^2(T)\right)\right) \\
&\quad -Ke^{-rT}\Phi\left(\widehat{d}_2\left(S_0e^{k(\mu_j+\sigma_j^2/2)-\lambda\mu_JT};\sigma_k^2(T)\right)\right)\Big)\ ,
\end{aligned}$$

finally by using (10.111) with (10.113) and (10.112).

Note that by several IID and normal identities,

$$\mathrm{E}[e^{\widehat{S}_k}] = \mathrm{E}\left[e^{\sum_{i=1}^k Q_i}\right] = \prod_{i=1}^k \mathrm{E}\left[e^{Q_i}\right] = \prod_{i=1}^k e^{\mu_j+\sigma_j^2/2} = e^{k(\mu_j+\sigma_j^2/2)},$$

giving the meaning of this exponential term in (10.109) for the final normal jump-diffusion call option result $\mathcal{C}_n^{(\mathrm{rn})}(S_0,T)$. □

Option pricing for other jump-diffusions can not be written in as simple a form and the Poisson terms increase in complexity exponentially. The use of the double-exponential (Laplace) log-jump-amplitude jump-diffusion has been developed by Kou and co-worker [169, 170]. Zhu and Hanson [290] have developed a Monte-Carlo estimation of risk-neutral option pricing for uniform log-jump-amplitude jump-diffusions. Zhu [289] has made a comprehensive study and comparison of various exponential and uniform log-jump-amplitude jump-diffusions using refined Monte-Carlo estimations of option prices with several variance reduction techniques. Recently, Yan and Hanson [287] have treated option pricing for the uniform log-jump-amplitude jump-diffusion combined with stochastic volatility (SVJD) using characteristic functions and fast Fourier transforms following the general methodology of Carr and Madan [47]. Yan and Hanson [130] computationally solve the SVJD problem using a systematic finite difference formulation of the free-boundary American put partial integro-differential complementary problem (PIDCP), implemented using a successive over-relaxtion (SOR) method projected on the maximum payoff function.

Some other hedging methods for jump-diffusions, like mean-variance hedging, are treated in a more abstract way by Runggaldier [239], Bingham and Kiesel [33],

and Cont and Tankov [59] using a generalization of jump-diffusions allowing infinite jump-rates called Lévy processes (see Chapter 12 in this book).

10.4 Optimal Portfolio and Consumption Models

Prior to Merton's 1973 mathematical justication and generalization of the Black-Scholes model [34] in [201], he did pioneering work on the portfolio and consumption problem in continuous-time. Beginning in 1969 Merton's paper [198] ([203, Chapter 4]) on lifetime portfolio selection with constant relative risk-aversion (CRRA) utilities laid out the background for the widely cited 1971 paper [199, 200] (reprinted in [203, Chapter 5]) on the optimal portfolio and consumption theory with the more general hyperbolic absolute risk-aversion (HARA) utilities that exhibit explicit solutions. While the paper was primarily on geometric Brownian motion (pure diffusion), generalization to jump-diffusions consisting of Brownian motion and compound Poisson processes with general random finite amplitude is discussed very briefly in [199].

While Merton was often on the leading edge of continuous-time finance and pushing generality of financial models by incorporating the latest financial and stochastic theories, one can get cut on the leading edge. There are a number of errors in the 1971 Merton paper [199, 200] due to the lack of proper boundary conditions and problems with the general HARA utilities. In particular, there are difficulties due to enforcing non-negative wealth, handing zero wealth (bankruptcy) and maintaining the non-negativity of consumption. These errors are very thoroughly discussed in Sethi's [245] massive assembly of papers by Sethi and his coauthors that give corrections and generations to the consumption and investment portfolios with an emphasis on bankruptcy and pure diffusion. The basic problems are clearly discussed in Sethi's introduction [245, Chapter 1], while important basic papers are the paper of Karatzas, Lehoczhy, Sethi, Shreve [159] (reprint [245, Chapter 2]) on exact solutions the infinite horizon case and the paper of Sethi and Taksar [246] (reprint [245, Chapter 3]) pinpointing the errors in Merton's 1971 paper [199, 200]. The errors were mainly in certain ranges of the HARA utilities and these difficulties led to a more thorough exploration of the consumption and portfolio problem.

In this section, the jump-diffusion version for the consumption and portfolio problem is treated with a version of the CRRA utilities that avoids the problematic parameter range of the general HARA utilities. In particular, the text-oriented presentation here is partly based on a portfolio optimization paper with time-dependence and uniformly distributed log-jump-amplitudes of Hanson and Westman [126] with some corrections.

10.4.1 Log-Uniform Amplitude Jump-Diffusion for Log-Return

Let $S(t)$ be the price of a single financial asset at time t, such as a stock or mutual fund, governed by a geometric jump-diffusion stochastic differential equation (SDE)

with time-dependent coefficients,

$$dS(t) = S(t)\left(\mu_d(t)dt + \sigma_d(t)dG(t) + \sum_{k=1}^{dP(t)} J\left(T_k^-, Q_k\right)\right), \qquad (10.115)$$

with $S(0) = S_0,\ \ S(t) > 0$, where $\mu_d(t)$ is the mean appreciation return rate, $\sigma_d(t)$ is the volatility, $G(t)$ is a continuous Gaussian process with zero-mean and t-variance (G is used for the diffusion component of the noise since W in this section will denote the wealth), $P(t)$ is a discontinuous, Poisson process with jump rate $\lambda(t)$, and associated jump-amplitude $J(t,Q)$, $-1 < J(t,Q) < \infty$ to avoid bankruptcy at a single jump, with log-return mark Q mean $\mu_j(t)$ and variance $\sigma_j^2(t)$. At the kth Poisson jump, T_k^- is the pre-jump time and Q_k is the corresponding IID random pick for the mark. The stochastic processes $G(t)$ and $P(t)$ are assumed to be Markov and pairwise independent. The jump-amplitude mark Q, given that a Poisson jump in time occurs, is also independently distributed. The stock price SDE (10.115) is similar in our prior work [124, 123], except that time-dependent coefficients introduce more realism.

Since the stock price process is a geometric jump-diffusion, the common multiplicative factor of $S(t)$ on the right can be removed by a logarithmic transformation yielding the SDE of the stock price log-return,

$$d\ln(S(t)) = \mu_{ld}(t)dt + \sigma_d(t)dG(t) + \sum_{k=1}^{dP(t)} \ln\left(1 + J\left(T_k^-, Q_k\right)\right), \quad (10.116)$$

where $\mu_{ld}(t) \equiv \mu_d(t) - \sigma_d^2(t)/2$ is the log-diffusion drift and $\ln(1 + J(t,q))$ the stock log-return jump-amplitude is the logarithm of the relative post-jump-amplitude.

Since $J(t,q) > -1$, it is convenient to select the mark process to be the log-jump-amplitude random variable,

$$Q = \ln\left(1 + J(t,Q)\right), \qquad (10.117)$$

on the mark space $\mathcal{Q} = (-\infty, +\infty)$. Though this is a convenient mark selection, it implies the independence of the jump-amplitude in time, but not of the log-jump-amplitude distribution $\Phi_Q(q;t)$ for Q. For comparison to the Standard and Poor's (S&P500) log-return data, the discrete log-return difference form

$$\Delta\ln(S_i) \equiv \ln(S_{i+1}) - \ln(S_i) = \ln(S_{i+1}/S_i)$$

will be used at time $t_{i+1} = t_i + \Delta t_i$. The corresponding the log-return differential $d\ln(S(t))$ in SDE (10.116) is written in the approximate, mean-zero, independent process, discrete form,

$$\begin{aligned}\Delta\ln(S(t_i)) \simeq\ & \left(\mu_{ld}(t_i) + \lambda(t_i)\mu_j(t_i)\right)\Delta t_i + \sigma_d(t_i)\Delta G(t_i) \\ & + \mu_j(t_i)\left(\Delta P(t_i) - \lambda(t_i)\Delta t_i\right) + \textstyle\sum_{k=1}^{\Delta P(t_i)}\left(Q_k - \mu_j(t_i)\right),\end{aligned} \qquad (10.118)$$

where $\Delta G(t_i) \equiv G(t_{i+1}) - G(t_i)$ and $\Delta P(t_i) \equiv P(t_{i+1}) - P(t_i)$.

10.4.2 Log-Uniform Jump-Amplitude Model

Extreme jumps in the market are rare events making it difficult or impossible to separate out the jumps from a background of continuous diffusive changes (see Aït-Sahalia [5]) to determine their distribution. Extreme jumps are limited by **circuit breakers** [11] introduced by the New York Stock Exchange in 1988 as a response to the crash of 1987, so a finite jump-amplitude distribution like the uniform distribution is appropriate. Thus, consider the uniform density on $[a(t), b(t)]$ for the marks Q,

$$\phi_Q(q;t) \equiv \left\{ \begin{array}{ll} \frac{1}{b(t)-a(t)}, & a(t) \le q \le b(t) \\ 0, & \text{otherwise} \end{array} \right\} , \tag{10.119}$$

where $a(t) < 0 < b(t)$ to allow for both crashes ($q < 0$) and rallies ($q > 0$) .

The basic moments of the uniformly Q (uq) density $\phi_Q(q;t_i)$ yields the mean

$$\mathrm{E}_Q[Q] = \mu_j(t_i) = (a(t_i) + b(t_i))/2 \, , \tag{10.120}$$

variance

$$\mathrm{Var}_Q[Q] = \sigma_j^2(t_i) = (b(t_i) - a(t_i))^2/12, \tag{10.121}$$

third central moment

$$M_3^{(\mathrm{uq})}(t_i) \equiv \mathrm{E}_Q\left[(Q - \mu_j(t_i))^3\right] = 0$$

and fourth central moment

$$M_4^{(\mathrm{uq})}(t_i) \equiv \mathrm{E}_Q\left[(Q - \mu_j(t_i))^3\right] = 9\sigma_j^4(t_i)/5.$$

In terms of the original jump-amplitude $J(t, Q)$, the mean is

$$\bar{J}(t_i) \equiv \mathrm{E}_Q\left[J(Q, t_i)\right] = \mathrm{E}_Q\left[e^Q - 1\right] = \frac{e^{b(t_i)} - e^{a(t_i)}}{b(t_i) - a(t_i)} - 1.$$

The first four moments of the uniform jump-diffusion (UJD) log-return difference using (10.118) are

$$M_1^{(\mathrm{ujd})}(t_i) \equiv \mathrm{E}[\Delta \ln(S(t))] = (\mu_{ld}(t_i) + \lambda(t_i)\mu_j(t_i))\Delta t_i, \tag{10.122}$$

$$M_2^{(\mathrm{ujd})}(t_i) \equiv \mathrm{Var}[\Delta \ln(S(t_i))] = \left(\sigma_d^2(t_i) + \lambda(t_i)\left(\mu_j^2(t_i) + \sigma_j^2(t_i)\right)\right)\Delta t_i, \tag{10.123}$$

$$\begin{aligned} M_3^{(\mathrm{ujd})}(t_i) \equiv\ & \mathrm{E}\left[\left(\Delta \ln(S(t_i)) - M_1^{(\mathrm{ujd})}(t_i)\right)^3\right] \\ =\ & \lambda(t_i)\mu_j(t_i)\left(\mu_j^2(t_i) + 3\sigma_j^2(t_i)\right)\Delta t_i, \end{aligned} \tag{10.124}$$

$$\begin{aligned} M_4^{(\mathrm{ujd})}(t_i) \equiv & \ \mathrm{E}\left[\left(\Delta \ln(S(t_i)) - M_1^{(\mathrm{ujd})}(t_i)\right)^4\right] \\ = & \ \lambda(t_i)\left(\mu_j^4(t_i) + 6\mu_j^2(t_i)\sigma_j^2(t_i) + 9\sigma_j^4(t_i)/5\right)\Delta t_i \\ & +3\left(\sigma_d^2(t_i) + \lambda(t_i)\left(\mu_j^2(t_i) + \sigma_j^2(t_i)\right)\right)^2 (\Delta t_i)^2. \end{aligned} \tag{10.125}$$

The $M_m^{(\mathrm{ujd})}(t_i)$ moment calculations, in particular, need Lemma 5.15 from Chapter 5 for the four powers of partial sums of zero-mean IID random variables $\widehat{Q}_k = Q_k - \mu_j$, so

$$\mathrm{E}\left[\left(\sum_{k=1}^{n} \widehat{Q}_k\right)^m\right] = \left\{ \begin{array}{ll} 0, & m=1 \\ nM_2^{(\mathrm{uq})}(t_i) = n\sigma_j^2(t_i), & m=2 \\ nM_3^{(\mathrm{uq})}(t_i) = 0, & m=3 \\ nM_4^{(\mathrm{uq})}(t_i) + 3n(n-1)\left(M_2^{(\mathrm{uq})}(t_i)\right)^2, & m=4 \end{array} \right\},$$

where $n = \Delta P(t_i)$.

Let the uniform jump-diffusion be denoted by

$$X_i = \mathcal{G}_i + \sum_{k=1}^{\Delta P(t_i)} Q_k$$

where $\mathcal{G}_i = \mu_{ld}(t_i)\Delta t_i + \sigma_d(t_i)\Delta G(t_i)$ is the nonstandard Gaussian process, then the density for the uniform jump-diffusion X_i, is derived from the law of total probability (B.92) summing over all Poisson jumps and the nested convolution property (B.100),

$$\phi_{\mathrm{ujd}}(x) = \sum_{k=0}^{\infty} p_k(\lambda(t_i)\Delta t_i)\phi_{\mathrm{ujd}}^{(k)}(x),$$

where $p_k(\Lambda)$ is the usual Poisson counting distribution with corresponding kth density coefficient $\phi_{\mathrm{ujd}}^{(k)}(x)$ given by

$$\phi_{\mathrm{ujd}}^{(k)}(x) = \left(\phi_{\mathcal{G}_i}\,(*\phi_Q)^k\right)(x),$$

through the nested convolution property. The complexity of these coefficients grows exponentially with k. However, the first few are, using (5.77) for $k=0$,

$$\phi_{\mathrm{ujd}}^{(0)}(x) = \phi_{\mathcal{G}_i}(x) = \phi_n(x;\mu,\sigma^2),$$

where for brevity $\mu = \mu_{ld}(t_i)\Delta t_i$ and $\sigma^2 = \sigma_d^2(t_i)\Delta t_i$, now dropping the (t_i) argument for brevity and using (5.78) for $k=1$,

$$\phi_{\mathrm{ujd}}^{(1)}(x) = \left(\phi_{\mathcal{G}_i} * \phi_Q\right)(x) = \phi_{sn}(x-b, x-a; \mu, \sigma^2) \equiv \tfrac{\Phi_n(x-b,x-a;\mu,\sigma^2)}{b-a},$$

where $\phi_{sn}(x-b, x-a; \mu, \sigma^2)$ is called the **secant-normal density** (5.79), and finally from (5.80) with the **triangular density** (5.81) for $k=2$,

$$\begin{aligned} \phi_{\mathrm{ujd}}^{(2)}(x) = & \left(\phi_{\mathcal{G}_i}(*\phi_Q)^2\right)(x) = \tfrac{2b-x+\mu}{b-a}\phi_{sn}(x-2b, x-a-b; \mu, \sigma^2) \\ & + \tfrac{x-2a-\mu}{b-a}\phi_{sn}(x-a-b, x-2a; \mu, \sigma^2) \\ & + \tfrac{\sigma^2}{(b-a)^2}\left(\phi_n(x-2b;\mu,\sigma^2) - 2\phi_n(x-a-b;\mu,\sigma^2) + \phi_n(x-2a;\mu,\sigma^2)\right). \end{aligned}$$

There are five stochastic jump-diffusion model parameter processes to be estimated,

$$\{\mu_d(t), \sigma_d^2(t), \mu_j(t), \sigma_j^2(t), \lambda(t)\},$$

assuming that the interest rate process $r(t)$ is given deterministically and the time steps Δt_i over the given market period. Using the definitions of the jump mean-variance parameters $\{\mu_j(t), \sigma_j^2(t)\}$ in (10.120-10.121), the uniform jump range $\{a(t), b(t)\}$ can be estimated instead. The parameter estimations using variants of maximum likelihood methods is beyond the scope of this chapter, but the reader can consult our work in [126, 291] in the time-dependent parameter case and [124, 125, 127, 128, 131] for other background in the time-independent parameter case.

10.4.3 Optimal Portfolio and Consumption Policies Application

Let a portfolio contain a riskless asset, the bond, with price $B(t)$ dollars at time t in years, and a risky asset, the stock, with price $S(t)$ at time t. Let the instantaneous portfolio change fractions be $U_0(t)$ for the bond and $U_1(t)$ for the stock, such that the total satisfies $U_0(t) + U_1(t) = 1$. The bounds for $U_0(t)$ and $U_1(t)$ will be developed later from the jump-amplitude distribution and the non-negativity of wealth condition.

Let the bond price process be deterministic and exponential,

$$dB(t) = r(t)B(t)dt \ , \quad B(0) = B_0, \tag{10.126}$$

where $r(t)$ is the bond rate of interest at time t. The stock price $S(t)$ has been given in the jump-diffusion SDE (10.115). The portfolio wealth process changes due to changes in the portfolio fractions less the instantaneous consumption of wealth $C(t)dt$,

$$\begin{aligned} dW(t) = W(t) \big(r(t)dt + U_1(t) \big((\mu_d(t) - r(t))dt \\ + \sigma_d(t)dG(t) + \textstyle\sum_{k=1}^{dP(t)} J(Q_k) \big) \big) - C(t)dt \ , \end{aligned} \tag{10.127}$$

such that $W(t) \geq 0$ and that the consumption rate is constrained relative to wealth $0 \leq C(t) \leq C_0^{(\max)} W(t)$, consistent with non-negative constraints that Sethi and Taksar [246] show are needed. In addition, the stock fraction is bounded by fixed constants. $U_0^{(\min)} \leq U_1(t) \leq U_0^{(\max)}$, so borrowing and short-selling is permited. In (10.127), $U_0(t) = 1 - U_1(t)$ has been eliminated [123, 126, 291].

The investor's portfolio objective is to maximize the conditional, expected current value of the discounted utility $\mathcal{U}_f(w$ of final wealth at the end of the investment final time t_f and the discounted utility of instantaneous consumption preferences $\mathcal{U}(c)$, i.e., the optimal value of the portfolio satisfies

$$\begin{aligned} v^*(t,w) = \ & \max_{\{u,c\}[t,t_f)} \Big[\mathrm{E} \Big[e^{-\overline{\beta}(t,t_f)} \mathcal{U}_f(W(t_f)) \\ & + \textstyle\int_t^{t_f} e^{-\overline{\beta}(t,s)} \mathcal{U}(C(s)) \, ds \Big| \mathcal{C} \Big] \Big] \ , \end{aligned} \tag{10.128}$$

conditioned on the state-control set $\mathcal{C} = \{W(t) = w, U_1(t) = u, C(t) = c\}$, where the time horizon is assumed to be finite, $0 \le t < t_f$, and $\overline{\beta}(t,s)$ is the cumulative time-discount over time in (t,s) with $\overline{\beta}(t,t) = 0$ and discount rate $\beta(t) = (\partial\overline{\beta}/\partial s)(t,t)$ at time t. In order to avoid Merton's [199] difficulties with HARA utility functions too general for the portfolio and consumption problem, $\mathcal{U}'(C) \to +\infty$ as $C \to 0^+$ will be assumed for the utility of consumption, while a similar form will be used for the final utility $\mathcal{U}_f(W)$. Thus, the instantaneous consumption $c = C(t)$ and stock portfolio fraction $u = U_1(t)$ serve as two control variables, while the wealth $w = W(t)$ is the single state variable.

Absorbing Boundary Condition at Zero Wealth:

Eq. (10.128) is subject to zero wealth absorbing natural boundary condition. This avoids arbitrage as pointed out by Karatzas, Lehoczky, Sethi and Shreve [159] ([245, Chapter 2]). It is necessary to enforce non-negativity feasibility conditions on both wealth and consumption. They derive formally explicit solutions from a consumption-investment dynamic programming model with a an infinite horizon, that qualitatively correct the results of Merton [199, 200] ([203, Chapter 6]). See also Sethi and Taksar [246] for specific errors in [199, 200] and Sethi's excellent summary [245, Chapter 1].

Here the Merton boundary condition correction in his 1990 text [203, Chap. 6] is used,

$$v^*(t,0^+) = \mathcal{U}_f(0)e^{-\overline{\beta}(t,t_f)} + \mathcal{U}(0)\int_t^{t_f} e^{-\overline{\beta}(t,s)}ds \ , \tag{10.129}$$

since the consumption must be zero when the wealth is zero. The terminal wealth condition, $v^*(t_f,w) = \mathcal{U}_f(w)$, must also be satisfied.

Portfolio Stochastic Dynamic Programming:

Assuming the optimal value $v^*(t,w)$ is continuously differentiable in t and twice continuously differentiable in w, then the stochastic dynamic programming equation (see our papers [123, 126, 291]) follows from an application of the (Itô) stochastic chain rule to the principle of optimality,

$$\begin{aligned} 0 = \ & v_t^*(t,w) - \beta(t)v^*(t,w) + \mathcal{U}(c^*) \\ & + [(r(t) + (\mu_d(t) - r(t))u^*)w - c^*]\, v_w^*(t,w) \\ & + \tfrac{1}{2}\sigma_d^2(t)(u^*)^2 w^2 v_{ww}^*(t,w) \\ & + \tfrac{\lambda(t)}{b(t)-a(t)}\int_{a(t)}^{b(t)} \cdot (v^*(t,\alpha(u^*,q)w) - v^*(t,w))\, dq, \end{aligned} \tag{10.130}$$

where $u^* = u^*(t,w) \in [U_0^{(\min)}, U_0^{(\max)}]$ and $c^* = c^*(t,w) \in [0, C_0^{(\max)}w]$ are the optimal controls if they exist, while $v_w^*(t,w)$ and $v_{ww}^*(t,w)$ are the partial derivatives with respect to wealth w when $0 \le t < t_f$. Upon a jump, the wealth changes by a factor

$$\alpha(u,q) \equiv 1 + (e^q - 1)u,$$

in the post-jump wealth argument of (10.130).

Non-Negativity of Wealth and Jump Distribution:

Non-negativity of wealth implies an additional consistency condition for the control since the jump in wealth argument $\alpha(u^*, q)w = (1 + (e^q - 1)u^*)w$ in the stochastic dynamic programming equation (10.130) requires $\alpha(u, q) \geq 0$ on the support interval of the jump-amplitude mark density $\phi_Q(q)$. Hence, it will make a difference in the optimal portfolio stock fraction u^* bounds if the support interval $[a(t), b(t)]$ is finite or if the support interval is $(-\infty, +\infty)$, i.e., had infinite range. Our results will be restricted to the usual case, the $a(t) < 0 < b(t)$, i.e., both crashes and rallies are modeled.

Lemma 10.6 (Bounds on Optimal Stock Fraction due to Non-Negativity of Wealth Jump Argument[291]). *If the support of $\phi_Q(q)$ is the finite interval $q \in [a(t), b(t)]$ with $a(t) < 0 < b(t)$, then $u^*(t, w)$ is restricted by (10.130) to*

$$\frac{-1}{\left(e^{b(t)} - 1\right)} \leq u^*(t, w) \leq \frac{1}{\left(1 - e^{a(t)}\right)}, \tag{10.131}$$

but if the support of $\phi_Q(q; t)$ is fully infinite, i.e., $(-\infty, +\infty)$, then $u^*(t, w)$ is restricted by (10.130) to

$$0 \leq u^*(t, w) \leq 1. \tag{10.132}$$

Proof. It is necessary that $\alpha(u, q) \geq 0$ so that $\alpha(u, q)w \geq 0$ when the wealth and its jump-in-wealth in the HJBE (10.130) argument need to be non-negative, $w \geq 0$. The borderline case is when instantaneous stock fraction case is zero, i.e., $u = 0$, so $\alpha(0, q) = 1 > 0$.

Next consider the case when the support, $a(t) \leq q \leq b(t)$, is finite. When $u > 0$, then

$$0 \leq 1 - \left(1 - e^{a(t)}\right) u \leq \alpha(u, q) \leq 1 + \left(e^{b(t)} - 1\right) u.$$

Since $e^{a(t)} < 1 < e^{b(t)}$, the worse case for enforcing $\alpha(u, q) \geq 0$ is on the left, so

$$u \leq \frac{+1}{\left(1 - e^{a(t)}\right)} = \frac{-1}{J(t, a(t))}.$$

When $u < 0$, then

$$0 \leq 1 - \left(e^{b(t)} - 1\right)(-u) \leq \alpha(u, q) \leq 1 + \left(1 - e^{a(t)}\right)(-u).$$

The worse case for enforcing $\alpha(u, q) \geq 0$ is again on the left, so upon reversing signs,

$$u \geq \frac{-1}{\left(e^{b(t)-1}\right)} = \frac{-1}{J(t, b(t))},$$

completing both sides of the finite case (10.131), which can be written in terms of the original jump-amplitude coefficient $-1/J(t, b(t)) \leq u^*(t, w) \leq -1/J(t, a(t))$.

In the infinite range jump model case when $-\infty < q < +\infty$, $0 < e^q < \infty$. Thus, when $u > 0$,

$$0 \leq 1 - u < \alpha(u, q) < \infty,$$

so $u \leq 1$. However, when $u < 0$, then

$$-\infty < \alpha(u, q) < 1 - u,$$

so $u < 0$ leads to a contradiction in that $\alpha(u, q)$ is unbounded below and $u \geq 0$, proving (10.132), which is the limiting case of (10.131) when $a(t) \to -\infty$ and $b(t) \to -\infty$. □

Remark 10.7. *This lemma gives the constraints on the instantaneous stock fraction $u^*(t, w)$ that limit the jumps to the jumps that at most just wipe out the investor's wealth. Unlike the case of pure diffusion where the functional term has local dependence on the wealth mainly through partial derivatives, the case of jump-diffusion has global dependence through jump integrals over finite differences with jump modified wealth arguments, leading to additional constraints under non-negative wealth conditions that do not appear for pure diffusions. The additional constraint comes not from the current wealth or nearby wealth but from the discontinuous wealth created by a jump.*

In the case of the fitted log-uniform jump-amplitude model, the range of the jump-amplitude marks $[a(t), b(t)]$ is covered by the estimated largest range,

$$[a^{(\min)}, b^{(\max)}] = \left[\min_t(a(t)), \max_t(b(t))\right] \simeq [-7.113e\text{-}2, +4.990e\text{-}2],$$

over the period from 1992-2001 corresponding to $t = 1{:}10$ using [126] results. The corresponding overall estimated range of the optimal instantaneous stock fraction $u^*(t, w)$ is then

$$[u^{(\min)}, u^{(\max)}] = \left[\frac{-1}{\left(e^{b^{(\max)}} - 1\right)}, \frac{+1}{\left(1 - e^{a^{(\min)}}\right)}\right] \simeq [-19.54, +14.56.],$$

in large contrast to the highly restricted infinite range models where

$$[\min(u^*(t, w)), \max(u^*(t, w))] = [0, 1]$$

is fixed for any t.

Regular Optimal Control Policies:

In the absence of constraints on the controls, then the maximum controls are the regular optimal controls $u^{(\text{reg})}(t, w)$ and $c^{(\text{reg})}(t, w)$, which are given implicitly, provided they are attainable and there is sufficient differentiability in c and u, by the dual critical conditions,

$$\mathcal{U}'(c^{(\text{reg})}(t, w)) = v^*_w(t, w)\ , \tag{10.133}$$

$$\begin{array}{l} \sigma_d^2(t)w^2 v^*_{ww}(t,w)u^{(\mathrm{reg})}(t,w) = -(\mu_d(t)-r(t))wv^*_w(t,w) \\ -\lambda(t)w\frac{1}{b(t)-a(t)}\int_{a(t)}^{b(t)}(e^q-1)v^*_w(t,\alpha(u^{(\mathrm{reg})}(t,w),q)w)\ dq\ , \end{array} \tag{10.134}$$

for the optimal consumption and portfolio policies with respect to the terminal wealth and instantaneous consumption utilities (6.2). Note that (10.133-10.134) define the set of regular controls implicitly.

10.4.4 CRRA Utility and Canonical Solution Reduction:

For the risk-averse investor, the utilities are assumed to be the constant relative risk-aversion (CRRA) power utilities [203, 122], with the same power for both wealth and consumption,

$$\mathcal{U}(x) = \mathcal{U}_f(x) = x^\gamma/\gamma\ ,\quad x \geq 0\ ,\quad 0 < \gamma < 1\ . \tag{10.135}$$

The CRRA utility designation arises since the relative risk aversion is the negative of the derivative ($\mathcal{U}''(x)$) in the marginal utility ($\mathcal{U}'(x)$) relative to the average change in the marginal utility ($\mathcal{U}'(x)/x$), or here

$$RRA(x) \equiv -\mathcal{U}''(x)/(\mathcal{U}'(x)/x) = (1-\gamma) > 0, \tag{10.136}$$

i.e., a positive constant, and is a special case of the more general HARA utilities.

The CRRA power utilities for the optimal consumption and portfolio problem lead to a **canonical reduction** of the stochastic dynamic programming PDE problem to a simpler ODE problem in time, by the separation of wealth and time dependence,

$$v^*(t,w) = \mathcal{U}(w)v_0(t), \tag{10.137}$$

where only the time function $v_0(t)$ is to be determined. The regular consumption control is a linear function of the wealth,

$$c^{(\mathrm{reg})}(t,w) \equiv w \cdot c_0^{(\mathrm{reg})}(t) = w/v_0^{1/(1-\gamma)}(t), \tag{10.138}$$

using (10.133) and $\mathcal{U}'(x) = x^{\gamma-1}$ from (10.135). The regular stock fraction u from (10.134) is a wealth independent control, but is given in uniform case implicit form:

$$u^{(\mathrm{reg})}(t,w) = \ u_0^{(\mathrm{reg})}(t) \equiv \frac{1}{(1-\gamma)\sigma_d^2(t)}\left[\mu_d(t) - r(t) + \lambda(t)I_1\left(u_0^{(\mathrm{reg})}(t)\right)\right], \tag{10.139}$$

$$I_1(u) = \frac{1}{b(t)-a(t)}\int_{a(t)}^{b(t)}(e^q-1)\alpha^{\gamma-1}(t,w)dq. \tag{10.140}$$

The wealth independent property of the regular stock fraction is essential for the separability of the optimal value function (10.137). Since (10.139) only defines $u_0^{(\mathrm{reg})}(t)$ implicitly in fixed point form, $u_0^{(\mathrm{reg})}(t)$ must be found by an iteration such as Newton's method, while the general Gauss-Statistics quadrature [277] can be used for jump integrals (see [123]).

The optimal controls, when there are constraints, are given in piecewise form as

$$c^*(t,w)/w = c_0^*(t) \equiv \max\left[\min\left[c_0^{(\mathrm{reg})}(t), C_0^{(\max)}\right], 0\right],$$

provided $w > 0$, and

$$u^*(t,w) = u_0^*(t) \equiv \max\left[\min\left[u_0^{(\mathrm{reg})}(t), U_0^{(\max)}\right], U_0^{(\min)}\right],$$

is independent of w along with $u_0^{(\mathrm{reg})}(t)$.

Substitution of the separable power solution (10.137) and the regular controls (10.138-10.139) into the stochastic dynamic programming equation (10.130), leads to an apparent Bernoulli type ODE,

$$0 = v_0'(t) + (1-\gamma)\left(g_1(t,u_0^*(t))v_0(t) + g_2(t)v_0^{\frac{\gamma}{\gamma-1}}(t)\right), \tag{10.141}$$

$$\begin{aligned} g_1(t,u) \equiv\ & \tfrac{1}{1-\gamma}\left[-\beta(t) + \gamma\left(r(t) + u(\mu_d(t) - r(t))\right)\right. \\ & \left. - \tfrac{\gamma(1-\gamma)}{2}\sigma_d^2(t)u^2 + \lambda(t)(I_2(t,u) - 1)\right], \end{aligned} \tag{10.142}$$

$$g_2(t) \equiv \frac{1}{1-\gamma}\left[\left(\frac{c_0^*(t)}{c_0^{(\mathrm{reg})}(t)}\right)^{\gamma} - \gamma\left(\frac{c_0^*(t)}{c_0^{(\mathrm{reg})}(t)}\right)\right], \tag{10.143}$$

$$I_2(t,u) \equiv \frac{1}{b(t)-a(t)}\int_{a(t)}^{b(t)} \alpha^{\gamma}(u,q)\, dq\,, \tag{10.144}$$

for $0 \le t < t_f$. The coupling of $v_0(t)$ to the time dependent part of the consumption term $c_0^{(\mathrm{reg})}(t)$ in $g_2(t)$ and the relationship of $c_0^{(\mathrm{reg})}(t)$ to $v_0(t)$ in (10.138) means that the differential equation (10.141) is implicit and highly nonlinear and thus (10.141) is only of Bernoulli type formally. The apparent Bernoulli equation (10.141) can be transformed to an apparent linear differential equation by using the Bernoulli linearizing transformation $\theta(t) = v_0^{1/(1-\gamma)}(t)$, to obtain,

$$0 = \theta'(t) + g_1(t,u_0^*)\theta(t) + g_2(t),$$

whose general solution can be inverse transformed to the general solution for the separated time, but implicit, function,

$$v_0(t) = \theta^{1-\gamma}(t) = \left[e^{-g_1(t,u_0^*(t))(t_f-t)}\left(1 + \int_t^{t_f} g_2(\tau)e^{g_1(t,u_0^*(t))(t_f-\tau)}d\tau\right)\right]^{1-\gamma}. \tag{10.145}$$

In order to illustrate this stochastic application, a computational approximation of the solution is presented. The main computational changes from the procedure used in [123] are that the jump-amplitude distribution is now uniform and the portfolio parameters, as well as the jump-amplitude distribution are time-dependent. Parameter time-dependence is approximated by quadratic interpolation

over the years from 1992-2001. The terminal time is taken to be $t_f = 11$, one year beyond this range. For this numerical study, the economic rates are taken to be time-independent, so the bond interest rate is $r(t) = 5.75\%$ and the time-discount rate is $\beta(t) = 5.25\%$. The portfolio stock fraction constraints are

$$[U_0^{(\min)}, U_0^{(\max)}] = [-10, 10] \text{ and } C_0^{(\max)} = 0.75$$

for consumption relative to wealth.

In Figure 10.1, the optimal portfolio stock fraction $u^*(t)$ is displayed. The portfolio policy is not monotonic in time and the minimum control constraint at $U_0^{(\min)}$ is active during the first half year in $t \in [0, t_f]$, while the maximum constraint is not activated since $u^*(t)$ remains significantly below that constraint. The $u^*(t)$ non-monotonic behavior is very interesting compared to the constant behavior in the constant parameter model in [123]. Likely the stock fraction grew initially due to the early relatively quiet period, then peaked at the beginning of the fourth year (1996 in the S&P500 data) as the market became noisier and continued to decline due to the final relatively noisier period. In Figure 10.2 the optimal, expected,

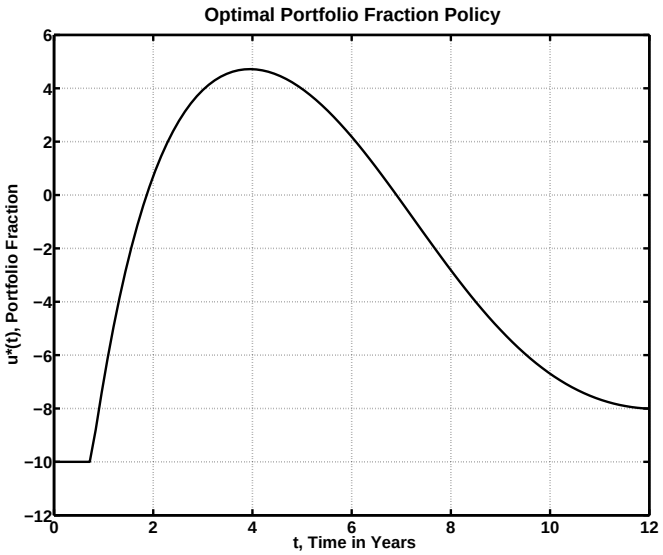


Figure 10.1. *Optimal portfolio stock fraction policy* $u^*(t)$ *on* $t \in [0, 12]$ *subject to the control constraint set* $[U_0^{(\min)}, U_0^{(\max)}] = [-10, 10]$.

cumulative consumption, $c^*(t, w)$, is displayed in three dimensions. The optimal consumption policy $c^*(t, w)$ results in this computational example are qualitatively similar to that of the time-independent parameter case in [123].

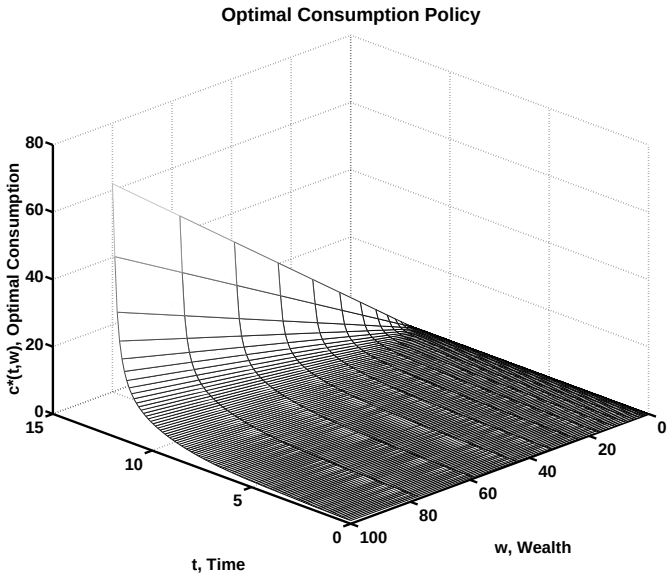


Figure 10.2. *Optimal consumption policy* $c^*(t, w)$ *for* $(t, w) \in [0, 12] \times [0, 100]$.

10.5 Important Financial Events Model: The Greenspan Process

Many financially critical announcements can have significant effects in the market, such as those on interest rates, unemployment statistics, budget deficits, trade deficits, prices of supplies such as oil, weather extremes and many others. Some of these announcements are scheduled like those of the Federal Reserve Board, labor reports or business earnings. The response to these scheduled announcements are sometimes difficult to predict, because market investors may have already factored in unfavorable or favorable news. Whereas, unscheduled announcements present both uncertainties in time and response making the compound Poisson processes a reasonable model. The Poisson model would be unsuitable for scheduled announcements. In [129], Hanson and Westman proposed a quasi-deterministic stochastic jump process that resembled the compound Poisson process only in the random jump-amplitude components, but otherwise jump at scheduled or deterministic times. This theoretical basis for our paper was the optimal portfolio problem for important external events paper [235] of Rishel . Our contribution was primarily constructing the intricate computational procedure for the problem and formulating the problem as a full stochastic differential equation model. The formulation appears to be of interest in other financial problems where there are uncertain, scheduled payments such as dividends.

Scheduled jumps affect the market. The response magnitude of the jumps

can be random, as described by Rishel [235]. On February 17, 2000 there were large market fluctuations caused by the semi-annual economic report of the now former Federal Reserve Board Chairman Alan Greenspan to Congress concerning the raising of interest rates among other things. The next day was followed by a *double witching day* with the simultaneous expiration of contracts on stock options and indices. Although these events and the market responses to them are quite complicated, these quasi-deterministic processes are strongly motivated by the influential announcement events by Chairman Greenspan and thus they might be called "*Greenspan processes*."

The optimal portfolio and consumption work [122, 129] will be summarized and reformulated with the constructs of this book. The reformulation uses a more concrete formulation of the quasi-deterministic processes than the more general, abstract Poisson random measure-like formulation in [129]. Also the problem was reduced to a single risky asset model from the multi-asset model in [129]

10.5.1 Stochastic Scheduled and Unscheduled Events Model, with Stochastic Parameter Processes

Let the usual Poisson process $P(t)$ denote an unscheduled events process which occur at the random times T_k for $k = 1, 2, \ldots$, with random jump-amplitudes $J(p_k, \mathbf{A}(T_k^-))$ where p is the corresponding random mark and $\mathbf{A}(t)$ is an auxiliary parameter, vector process. Let the quasi-deterministic process or *Greenspan process* $Q(t)$ denote a scheduled events process at fixed times τ_ℓ with random jump-amplitude $K(\widehat{q}; \mathbf{A}(\tau_\ell^-))$ where $\widehat{q}$ is the corresponding random mark. Both processes are right-continuous.

Let the portfolio consist of one almost risk-less asset $B(t)$ at time t and one risky asset $S(t)$. The risk-less asset $B(t)$ satisfies the familiar form,

$$dB(t) = r(\mathbf{A}(t))B(t)dt, \quad B(0) = B_0, \tag{10.146}$$

where the almost risk-less asset interest $r(\mathbf{a})$ depends on a mildly random parameter vector $\mathbf{A}(t) = [A_1(t); A_2(t)]$ associated with unscheduled and scheduled event processes. Here, $A_1(t))$ is a parameter for unscheduled events driven by Poisson process $dP(t)$, with jump-amplitude $J_1(q)$ and random mark p, and satisfied by

$$dA_1(t) = A_1(t)J_1(q)dP(t) = \sum_{k=1}^{dP(t)} A_1(T_k^-)J_1(q_k), \tag{10.147}$$

where $(T_k; q_k)$ are the kth Poisson time-mark parameters. The process $A_1(t)$ can be called a *geometric Poisson process* since the noise is linear in $A_1(t)$, making the noise multiplicative. For reasons cited in the previous section, the range of the unscheduled process mark is finite, so $a \le q \le b$. Also, $A_2(t)$ is a parameter process for scheduled events driven by the quasi-deterministic process $dQ(t)$, with jump-amplitude $K_2(\widehat{q})$ with random mark $\widehat{q}$, and satisfied by

$$dA_2(t) = A_2(t)K_2(\widehat{q})dQ(t) = \sum_{\ell=1}^{dQ(t)} A_2(\tau_\ell^-)K_2(\tau_\ell^-; \widehat{q}_\ell), \tag{10.148}$$

where τ_ℓ is a scheduled event time such that $\tau_{\ell+1} > \tau_\ell$ and $\widehat{q}_\ell$ is the ℓth realized jump-amplitude mark for $\ell = 1{:}M$, where $\tau_M = \max(\tau_\ell) < t_f$ with t_f being the portfolio final time and $\widehat{a} \leq \widehat{q} \leq \widehat{b}$. The process $A_2(t)$ is also a multiplicative or geometric noise process.

The risky portfolio asset with price $S(t)$ satisfies the SDE

$$\begin{aligned} dS(t)=&S(t)(\mu(\mathbf{A}(t)) + \sigma(\mathbf{A}(t))dG(t) + J(q;\mathbf{A}(t))dP(t) \\ &+K(\widehat{q};\mathbf{A}(t))dQ(t), \end{aligned} \tag{10.149}$$

where $S(0) = S_0$, $0 \leq S(t) < t_f$, $\mu(\mathbf{a})$ is the mean stock appreciation rate, $G(t)$ is a standard Wiener or Gaussian process , $\sigma(\mathbf{a})$ is the standard deviation coefficient corresponding to $dG(t)$, $\mathbf{A}(t)$ is an auxiliary parameter process, while the compound unscheduled and scheduled jump processes with jump-amplitudes short-hand notation can properly be defined as

$$S(t)J(q;\mathbf{A}(t))dP(t) = \sum_{k=1}^{dP(t)} S(T_k^-)J(q_k;\mathbf{A}(T_k^-)), \tag{10.150}$$

and

$$S(t)K(\widehat{q};\mathbf{A}(t))dQ(t) = \sum_{\ell=1}^{dQ(t)} S(\tau_\ell^-)K(\widehat{q}_\ell;\mathbf{A}(\tau_\ell^-)). \tag{10.151}$$

The primary difference between forms (10.150) and (10.151) is that in the former $dP(t)$ and T_k^- are stochastic with $\mathrm{E}[dP(t)] = \lambda dt$ and $T_{k+1} - T_k$ exponentially distributed (B.56), while in the latter $dQ(t)$ and τ_ℓ are deterministic so $\mathrm{E}[dQ(t)] = dQ(t)$ and $\mathrm{E}[\tau_\ell] = \tau_\ell$.

10.5.2 Further Properties of Quasi-Deterministic or Scheduled Event Processes: $K(\widehat{q};\mathrm{A}(t))dQ(t)$

The scheduled jump of the $dQ(t)$ of (10.151) is scheduled at prescribed times τ_ℓ and jump-counts $\ell = 1{:}M$, such that $\tau_{\ell+1} > \tau_\ell$ and $\tau_M = \max(\tau_\ell) < t_f$. At these times, random jump-amplitiudes $K(\widehat{q}_\ell;\mathbf{A}(\tau_\ell^-))$ where $\widehat{q}_\ell$ is the random mark or background random variable for which the probability distribution can be more conveniently specified. The $\mathbf{A}(t)$ is an auxiliary parameter process that is optional for the jump-amplitude function K associated with $dQ(t)$. The $dQ(t)$ is a pure deterministic counting process the triggers the random jump-amplitude.

The expectation of the event response jump-amplitude $K(\widehat{q};\mathbf{A}(t))$ conditioned on the parameter process is

$$\mathrm{E}[K(\widehat{q};\mathbf{A}(t))|\mathbf{A}(t) = \mathbf{a}] = \mathrm{E}[K(\widehat{q};\mathbf{a})] \equiv \overline{K}(\mathbf{a}).$$

The jump in the ith stock at a jump of the ith scheduled event processes is given by

$$[S](\tau_\ell) = S(\tau_\ell^+) - S(\tau_\ell^-) = K(\widehat{q}_\ell;\mathbf{A}(\tau_\ell^-))S(\tau_\ell^-),$$

for $\tau_\ell < t_f$ where t_f is the terminal time and stocks due to right-continuity property of the scheduled jump processes.

Similarly, for the scheduled parameter process $A_2(t)$, the jump at τ_ℓ is given by

$$[A_2](\tau_\ell) = A_2(\tau_\ell^+) - A_2(\tau_\ell^-) = K_2(\widehat{q}_\ell)A_2(\tau_\ell^-),$$

which in turn is similar to the jump of the unscheduled parameter process $A_1(t)$,

$$[A_1](T_k) = A_1(T_k^+) - A_2(T_k^-) = J_1(q_k)A_2(T_k^-).$$

10.5.3 Optimal Portfolio Utility, Stock Fraction and Consumption

The set-up of this optimal portfolio problem is similar to that of the prior section, so the focus will be mainly on differences arising from including the quasi-deterministic scheduled event processes and skipping similar intermediate steps. Let $W(t)$ be the portfolio wealth at time t, $U_1(t)$ is the vector of the ***instantaneous*** fraction of wealth invested in the risky assets at vector price $S(t)$, such that the risk-less asset fraction at price $B(t)$ satisfies

$$U_0(t) = 1 - U_1(t),$$

and $C(t)$ is the consumption of wealth. As in the prior sections, the jump-amplitude distributions will be assumed to be of finite range, so that the risky asset fractions will not be restricted to $[0, 1]$, but will be restricted to some larger and reasonable range $[U^{(\min)}, U^{(\max)}]$.

Following Eq. (10.127) of the previous section, the portfolio wealth process, relative changes due to relative changes in the portfolio fractions less the instantaneous consumption of wealth $C(t)dt$, is governed by the SDE,

$$\begin{aligned} dW(t) = &(W(t)\,(r(\mathbf{A}(t)) + U_1(t)(\mu(\mathbf{A}(t)) - r(\mathbf{A}(t))) - C(t))\,dt \\ &+W(t)U_1(t)\sigma(\mathbf{A}(t))dG(t) + W(t)U_1(t)J(q;\mathbf{A}(t))dP(t) \\ &+W(t)U_1(t)K(\widehat{q};\mathbf{A}(t))dQ(t)\ , \end{aligned} \tag{10.152}$$

with the necessary conditions that $W(t) \geq 0$ and that the consumption rate is constrained relative to wealth $0 \leq C(t) \leq C_0^{(\max)}W(t)$. For the stochastic dynamic programming formulation, it is necessary to know the jumps in the wealth for both unscheduled and scheduled jump-times, which are

$$[W](T_k) = W(T_k^+) - W(T_k^-) = W(T_k^-)U_1(T_k^-)J_1(q_k) \tag{10.153}$$

and

$$[W](\tau_\ell) = W(\tau_\ell^+) - W(\tau_\ell^-) = W(\tau_\ell^-)U_1(\tau_\ell^-)K_2(\widehat{q}_\ell)\ . \tag{10.154}$$

The investor's objective is to maximize the conditional, expected current value of the discounted utility $\mathcal{U}_f(w;\mathbf{a})$ of final wealth at the end of the investment final

time t_f and the discounted utility of instantaneous consumption preferences $\mathcal{U}(c)$, so that the optimal value of the portfolio satisfies

$$v^*(t,w;\mathbf{a}) = \max_{\{u,c\}[t,t_f)} \left[\mathrm{E}\left[e^{-\overline{\beta}(t,t_f)} \mathcal{U}_f(W(t_f);\mathbf{a}) + \int_t^{t_f} e^{-\overline{\beta}(t,s)} \mathcal{U}(C(s))\, ds \Big| \mathcal{C} \right] \right], \tag{10.155}$$

conditioned on the state-control set $\mathcal{C} = \{W(t) = w, U_1(t) = u, C(t) = c, \mathbf{A}(t) = \mathbf{a}\}$, where the time horizon is assumed to be finite, $0 \le t < t_f$, and $\overline{\beta}(t,s) \equiv \int_t^s \beta(\mathbf{A}(z))dz$ is the integral over the instant nominal discount rate $\beta(\mathbf{A}(t))$ on $[t,s]$. The instantaneous consumption $c = C(t)$ and stock portfolio fraction vector $u = U_1(t)$ serve as two control variables, while the wealth $w = W(t)$ is the single state variable.

Again, Merton's zero-wealth boundary condition correction given in his 1990 text [203, Chap. 6] is used, but here with the extra parameter argument,

$$v^*(t,0^+;\mathbf{a}) = \mathcal{U}_f(0;\mathbf{a}) e^{-\overline{\beta}(t,t_f)} + \mathcal{U}(0) \int_t^{t_f} e^{-\overline{\beta}(t,s)} ds, \tag{10.156}$$

since the consumption must be zero when the wealth is zero. The terminal wealth condition

$$v^*(t_f,w;\mathbf{a}) = \mathcal{U}_f(w;\mathbf{a}), \tag{10.157}$$

must also be satisfied and provides the start of the stochastic dynamic programming problem, a backward time problem.

The *constant relative risk-aversion* (CRRA) power utilities (10.135-10.136) are also used here, as in the last section, for the risk-averse investor, with the same power for consumption and wealth, but now with parameter values,

$$\begin{aligned} \mathcal{U}(c) &= c^\gamma/\gamma\,, \quad c \ge 0\,, \quad 0 < \gamma < 1\,. \\ \mathcal{U}_i(a_i) &= |a_i|^{\gamma_i}/\gamma_i\,, \quad a_i \ne 0\,, \quad \gamma_i \ne 0\,, \quad i = 1,2\,, \\ \mathcal{U}_f(w;\mathbf{a}) &= \mathcal{U}(w)\mathcal{U}_1(a_1)\mathcal{U}(a_2)\,, \quad w \ge 0\,, \quad \mathbf{a} = [a_1;a_2]\,. \end{aligned} \tag{10.158}$$

The utilities satisfy general properties, which in the case of consumption for example: (1) it is non-negative, $\mathcal{U}(c) \ge 0$, (2) the marginal utility is favorable toward consumption, $\mathcal{U}'(c) > 0$, (3) but at a decreasing rate, $\mathcal{U}''(c) < 0$.

The application of stochastic dynamic programming to the standard jump-diffusion with only Gaussian and Poisson noise leads to a single PDE in time t and wealth, as in the previous section, because the Gaussian and Poisson noise, in particular the Poisson jump times, average out with the expectation used in the objective. However, in the present problem with scheduled quasi-deterministic jumps, the scheduled jump-times are not averaged out by the expectation operator. Thus, between scheduled jump-times, τ_ℓ for $i = 1{:}N{+}1$ and jump-counters ℓ, the optimal value function $v^*(t,w;\mathbf{a})$ using the Principle of Optimality and expanding

using the SDEs and the stochastic chain rule to dt-precision,

$$\begin{aligned}
0 = \; & v^*_t(t,w;\mathbf{a}) - \beta(\mathbf{a}) v^*(t,w;\mathbf{a}) \\
& + \max_{u,c} \Big[\mathcal{U}(c) + ((r(\mathbf{a}) + u(\mu(\mathbf{a}) - r(\mathbf{a})))w - c)\, v^*_w(t,w;\mathbf{a}) \\
& + \tfrac{1}{2}(u\sigma(\mathbf{a}))^2 w^2 v^*_{ww}(t,w;\mathbf{a}) \\
& + \lambda \textstyle\int_a^b (v^*(t, w(1+uJ(q;\mathbf{a})); a_1(1+J_1(q)), a_2) - v^*(t,w;\mathbf{a}))\, \phi_q(q) dq \Big] \\
= \; & v^*_t(t,w;\mathbf{a}) - \beta(\mathbf{a}) v^*(t,w;\mathbf{a}) + \mathcal{U}(c^*) \\
& + ((r(\mathbf{a}) + u^*(\mu(\mathbf{a}) - r(\mathbf{a})))w - c^*)\, v^*_w(t,w;\mathbf{a}) \\
& + \tfrac{1}{2}(u^*\sigma(\mathbf{a}))^2 w^2 v^*_{ww}(t,w;\mathbf{a}) \\
& + \lambda \textstyle\int_a^b (v^*(t, w(1+u^*J(q;\mathbf{a})); a_1(1+J_1(q)), a_2) - v^*(t,w;\mathbf{a}))\, \phi_q(q) dq \;,
\end{aligned} \tag{10.159}$$

valid starting from the terminal wealth condition (10.157) and otherwise holding on open time intervals in backward order determined by the scheduled jump-times from (τ_M, t_f) to $(\tau_{\ell-1}, \tau_\ell)$ for $\ell = M{:}{-}1{:}2$ (the triple construct has the form *start : step : stop* as in MATLAB™) and $(0, \tau_1)$. Here, $u^* = u^*(t,w;\mathbf{a})$ and $c^* = c^*(t,w;\mathbf{a})$ are the optimal arguments of the maximum in the first part of (10.159) and are subject to previously stated constraints.

While the unscheduled, Poisson jumps are instantaneous and random, the expectation from the objective averages the jumps with $\mathrm{E}[dP(t)] = \lambda dt$ which is the same order as the contributions of the continuous terms in (10.159), the scheduled jumps are instantaneous and deterministic so they do not average, the continuous terms contribute zero in that instant and only the scheduled jump integral survives. Hence, at the scheduled jump-time τ_ℓ for $\ell = M{:}{-}1{:}\,1$ there is a new *stochastic dynamic programming jump condition*,

$$v^*(\tau_\ell^-, w;\mathbf{a}) = \textstyle\int_{\widehat{a}}^{\widehat{b}} v^* \left(\tau_\ell^+, w(1+u_\ell^- K(\widehat{q};\mathbf{a})); a_1, a_2(1+K_2(\widehat{q})\right) \phi_q(q) dq \;, \tag{10.160}$$

where $u_\ell^- = u^*(\tau_\ell^-, w;\mathbf{a})$. This condition does not arise in the usual jump-diffusion problem with only unscheduled jumps. Note that the value of $v^*(\tau_\ell^+, w;\mathbf{a})$ is to be found from integrating (10.159) from $\tau_{\ell+1}$ to τ_ℓ, so that the jump-condition (10.160) provides the new backward value $v^*(\tau_\ell^-, w;\mathbf{a})$ which is the start for the integration of (10.159) on $(\tau_{\ell-1}, \tau_\ell)$.

Since there is a non-negativity condition on wealth, that condition also applies to the wealth arguments in (10.159) and (10.160), so

$$(1+u^*J(q;\mathbf{a})) \geq 0$$

and

$$(1+u^*K(\widehat{q};\mathbf{a})) \geq 0$$

are additionally required, respectively.

If the consumption and stock fraction are unconstrained, then the regular controls, $c^{(\mathrm{reg})}(t,w;\mathbf{a})$ and $u^{(\mathrm{reg})}(t,w;\mathbf{a})$ are implicitly obtained, assuming sufficient differentiability,

$$\mathcal{U}'\left(c^{(\mathrm{reg})}(t,w;\mathbf{a})\right) = v^*_w(t,w;\mathbf{a}) \tag{10.161}$$

and

$$\begin{aligned}(\sigma(\mathbf{a})w)^2 v^*_{ww}(t,w;\mathbf{a})u^*(t,w;\mathbf{a}) = -(\mu(\mathbf{a})-r(\mathbf{a}))wv^*_w(t,w;\mathbf{a})\\ -\lambda w\textstyle\int_a^b J(q)v^*_w(t,w(1+u^*J(q));a_1(1+J_1(q)),a_2)\phi_q(q)dq\ .\end{aligned} \tag{10.162}$$

Since these regular control policies introduce both implicitness and nonlinearities into the PDE of stochastic dynamic programming (10.159), the solution will require computational iterations. There is also a jump in the regular stock fraction from (10.160) and is given implicitly by

$$0 = w\textstyle\int_{\widehat{a}}^{\widehat{b}} K(\widehat{q};\mathbf{a})v^*(\tau_\ell^+, w(1+u^-_{(\mathrm{reg}),\ell}K(\widehat{q};\mathbf{a}));a_1,a_2(1+K_2(\widehat{q})) \tag{10.163}$$

where $u_\ell^{(\mathrm{reg})-} = u^{(\mathrm{reg})}(\tau_\ell^-, w;\mathbf{a})$. The optimal policies (c^*,u^*) are found by applying the constraints to the regular control policies $(c^{(\mathrm{reg})},u^{(\mathrm{reg})})$.

10.5.4 Canonical CRRA Model Solution

The great advantage of the CRRA power utilities (10.158) for the portfolio and consumption optimization problem is that the solution is separable in the form

$$v^*(t,w;\mathbf{a}) = \mathcal{U}_f(w;\mathbf{a})v_0(t;\mathbf{a})\ , \tag{10.164}$$

so the wealth state can be completely stripped away in terms of a given utility function $\mathcal{U}_f(w;\mathbf{a})$, avoiding the exponential computational complexity of the *curse of dimensionality*. Also, the terminal condition (10.158) is easily satisfied as long as the remaining time-dependent part of the solution satisfies

$$v_0(t_f;\mathbf{a}) = 1,$$

and since $\mathcal{U}(0^+;\mathbf{a}) = 0 = \mathcal{U}(0^+)$ the zero-wealth absorbing boundary condition (10.156) $v^*(t,0^+;\mathbf{a}) = 0$.

Substituting the canonical solution into the implicit equation (10.161) for $c^{(\mathrm{reg})}(t,w;\mathbf{a})$ yields a preliminary solution linear in w and in terms of $v_0(t;\mathbf{a})$,

$$c^{(\mathrm{reg})}(t,w;\mathbf{a}) = w\cdot c_0^{(\mathrm{reg})}(t;\mathbf{a}) \equiv \frac{w\psi_2(\mathbf{a})}{v_0^{1/(1-\gamma)}(t;\mathbf{a})}, \tag{10.165}$$

where $\psi_2(\mathbf{a}) \equiv 1/(\mathcal{U}_1(a_1)\mathcal{U}_2(a_2))^{1/(1-\gamma)}$, using some algebra. The corresponding optimal consumption is given by

$$c^*(t,w;\mathbf{a}) = wc_0^*(t;\mathbf{a}) = w\max\left(c_0^{(\mathrm{reg})}(t;\mathbf{a}), C_0^{(\max)}\right)\ . \tag{10.166}$$

However, the reduction of the $u^{(\mathrm{reg})}(t,w;\mathbf{a})$ does not eliminate the implicitness, but yields a solution independent of w, i.e., $u^{(\mathrm{reg})}(t,w;\mathbf{a}) = u_0^{(\mathrm{reg})}(t;\mathbf{a})$, a prime criterion for separability, where

$$u_0^{(\mathrm{reg})}(t;\mathbf{a}) = \frac{1}{(1-\gamma)\sigma^2(\mathbf{a})}\left(\mu(\mathbf{a})-r(\mathbf{a})+\frac{\lambda}{\gamma}I_1'\left(t,u_0^{(\mathrm{reg})}(t;\mathbf{a});\mathbf{a}\right)\right), \tag{10.167}$$

and where

$$I_1'(t,u;\mathbf{a}) \equiv \gamma^2 \int_a^b J(q;\mathbf{a}) \frac{\mathcal{U}(1+uJ(q;\mathbf{a}))}{(1+uJ(q;\mathbf{a}))} \mathcal{U}(1+J_1(q)) \psi_1(t,q;\mathbf{a}) \phi_q(q) dq \ ,$$

$$\psi_1(t,q;\mathbf{a}) \equiv \frac{v_0(t;(1+J_1(q))a_1,a_2)}{v_0(t;a_1,a_2)} \ ,$$

noting that $\psi_1(t,q;\mathbf{a})$ is the primary source of implicitness. The corresponding optimal portfolio fraction is given by

$$u^*(t,w;\mathbf{a}) = \max\left(U^{(\min)}, \min\left(U^{(\max)}, u_0^{(\mathrm{reg})}(t;\mathbf{a})\right)\right) \ . \tag{10.168}$$

Substituting the PDE (10.159) and CRRA separated solution (10.164) along with the optimal controls (10.166-10.168), leads to an implicit Bernoulli-type ordinary differential equation,

$$0 = v_0'(t;\mathbf{a}) + (1-\gamma)\left(\psi_3'(t,u^*(t;\mathbf{a});\mathbf{a}) v_0(t;\mathbf{a}) + \widehat{\psi}_2(t;\mathbf{a}) v_0^{\frac{\gamma}{\gamma-1}}(t;\mathbf{a})\right) \ , \tag{10.169}$$

where

$$\begin{aligned}
\psi_3'(t,u;\mathbf{a}) = \ & \partial\psi_3(t,u;\mathbf{a})/\partial t \equiv \tfrac{1}{1-\gamma}\left(-\beta(\mathbf{a}) + \gamma\left(r(\mathbf{a}) - (\mu(\mathbf{a}) - r(\mathbf{a}))u\right)\right. \\
& \left. - \tfrac{\gamma(1-\gamma)}{2\sigma^2(\mathbf{a})} u^2 + \lambda(I_1(t,u;\mathbf{a}) - 1)\right) \ , \\
\widehat{\psi}_2(t;\mathbf{a}) \equiv \ & \tfrac{1}{1-\gamma}\left(\left(\tfrac{c_0^*(t;\mathbf{a})}{c_0^{(\mathrm{reg})}(t;\mathbf{a})}\right)^\gamma - \gamma\left(\tfrac{c_0^*(t;\mathbf{a})}{c_0^{(\mathrm{reg})}(t;\mathbf{a})}\right)\right)\psi_2(\mathbf{a}) \ , \\
I_1(t,u;\mathbf{a}) - 1 \equiv \ & \gamma \int_a^b \mathcal{U}(1+uJ(q;\mathbf{a}))\psi_1(t,q;\mathbf{a}) \mathcal{U}_1(1+J_1(q))\psi_1(t,q;\mathbf{a})\phi_q(q)dq \ ,
\end{aligned}$$

when t is on $(\tau_{\ell-1},\tau_\ell)$ for $\ell = (M+1):{-1}:1$, conveniently defining $\tau_{M+1} \equiv t_f$ and $\tau_0 \equiv 0$.

The implicit, nonlinear Bernoulli equation can be linearized by the transformation

$$\theta(t) = v_0^{1/(1-\gamma)}(t;\mathbf{a}),$$

so (10.169) becomes

$$0 = \theta'(t) + \psi_3'(t,u^*(t;\mathbf{a});\mathbf{a})\theta(t) + \widehat{\psi}_2(t;\mathbf{a}) \ , \tag{10.170}$$

which can easily be solved for $\theta(t)$ but only formerly in terms of the implicit dependence on the controls which requires iteration to obtain a fully explicit solution.

Besides iterations, the computation of the solution has many complications in terms of integrating the jump integrals embedded in the coefficients, merging a regular time grid with the scheduled jumps and assembling solutions on the unscheduled subintervals with the jump conditions at the scheduled jumps in time. A summary for the computational algorithm is given in [129] along with the solutions for a test case of discrete jumps and various parameter values. The merger of the regular-time grid and the jump-time grid is illustrated in the simple jump-adapted code fragment in Fig. 9.7 of Subsubsect. 9.1.5 on p. 264. This is the complication that is the most asked question about this problem, mostly because it has many other applications in finance where there are jumps added onto a continuous process, such as discrete transaction costs, dividends and death benefits.

10.6 Exercises

Many of these exercises, depending on the instructor, whether numerical or theoretical, can be done by MATLAB, Maple or *Mathematica*, but if theoretical, the Symbolic Toolbox in MATLAB will be needed.

1. Show that the Itô mean square limit for correlated bond-stock price noise at time t (10.16)

$$dW_B(t)dW_S(t) \stackrel{ims}{=} \rho dt \ , \tag{10.171}$$

   is valid. Are there any special treatments required if $\rho = 0$ or $\rho = \pm 1$? You may use the bivariate normal density in (B.146) or Table B.1 of selected moments of preliminaries Chapter B.

2. Merton [201] ([203], p. 266) gives a stricter definition of more risky or riskier,

   Security $X_1(t)$ is **more risky** than security $X_2(t)$ if

$$X_1(t) = qX_2(t) + \epsilon \ ,$$

   where $(q, X_2(t), \epsilon)$ are mutually independent, $\mathrm{E}[q] = 1$, $\mathrm{E}[X_2(t)] = \mu_2(t)$, $\mathrm{E}[\epsilon] = 0$, $\mathrm{Var}[q] = \sigma_q^2$, $\mathrm{Var}[X_1(t)] = \sigma_1^2(t)$, $\mathrm{Var}[X_2(t)] = \sigma_2^2(t) > 0$ and $\mathrm{Var}[\epsilon] = \sigma_\epsilon^2$.

   (a) Show that

$$\sigma_1^2(t) = (1 + \sigma_q^2)\sigma_2^2(t) + \mu_2^2(t)\sigma_q^2 + \sigma_\epsilon^2 > \sigma_2^2(t) \ ,$$

   (b) Can you demonstrate this for a financial application or critically evaluate the applicability of the definition?

3. Verify that the call option pricing solution $\Phi^{(call)}(w, u)$ (10.82) or (10.83) satisfies the

   (a) standard diffusion PDE (10.73),

   (b) call final condition (10.84).

   Either Maple or *Mathematica* is recommended.

4. Verify that the put option pricing solution $\Phi^{(\mathrm{put})}(w, u)$ (10.85) or (10.86) satisfies the

   (a) standard diffusion PDE (10.73),

   (b) put final condition (10.87).

   Either Maple or *Mathematica* is recommended.

5. (a) Reverse the transformations to obtain option pricing solutions for

$$F^{(call)}(S, B, t; T, K) \text{ and } F^{(\text{put})}(S, B, t; T, K)$$

from the transformed solutions $\Phi^{(call)}(w, u)$ and $\Phi^{(\text{put})}(w, u)$, respectively, through restoring the original variables B, S, F and $\tau = T - t$.

(b) Reduce the final restored form to the Black-Scholes assumptions on volatilities and mean rates.

6. Show that the transformed call and put option solutions satisfy a more usual call-put parity principle, %

$$\left(\widehat{F}^{((\text{put}))} - \widehat{F}^{((call))}\right)(\widehat{S}, \widehat{B}, \tau; T, K) = K \exp(-R(\tau)) - \widehat{S}(\tau),$$

if certain conditions are satisfied and specify those conditions.

Suggested References for Further Reading

- Aït-Sahalia, 2004 [5].
- Aourir, et al., 2002 [11].
- Bachelier, 1900 [16].
- Ball and Torous, 1985 [18].
- Black, 1989 [35].
- Black and Scholes, 1973 [34].
- Bingham and Kiesel, 2004 [33].
- Bossaerts, 2002, [41].
- Bridgeman, 1963 [42].
- Carr and Madan, 1999 [47].
- Chichilnisky, 1996 [51].
- Cont and Tankov, 2004 [59].
- Courtault et al., 2000 [61].
- Cox and Rubinstein, 1985 [63].
- Duffie, 1992 [74].
- Haberman, 1983 [102].
- Hanson, 1996 [108].

- Hanson and Westman, 2001 [122], 2002 [124, 125, 123, 126], 2003 [127, 128] and 2004 [129, 131].
- Heath and Schweizer, 2000 [135].
- D. Higham, 2004 [140].
- Hull, 2000 [147].
- Jarrow and Rosenfeld, 1984 [153].
- Jorion, 1989 [155].
- Kamien and Schwartz, 1981 [158].
- Karatzas, et al., 1986 [159].
- Karatzas and Shreve, 1998 [160].
- Klebaner, 1998 [164].
- Kou, 2002 [169].
- Kou and Wang, 2004 [170].
- Lipton, 2001 [186].
- Merton, 1969 [198], 1971 [199] , 1973a [200], 1973b [201], 1976 [202] and 1990 [203].
- Merton and Scholes, 1996 [204].
- Mikosch, 1998 [209].
- Neftci, 2000 [217].
- Pliska, 1997 [225].
- Press et al., 2002 [230].
- Rishel, 1999 [235].
- Rogers and Williams, 2000 [236].
- Runggaldier, 2003 [239].
- Sethi, 1997 [245].
- Sethi and Taksar, 1988 [246].
- Shreve, 2004 [248].
- Sneddon, 1957 [251].
- Tavella and Randall, 2000 [264].

- Westman and Hanson, 2000, [277].
- Wilmott, 2000 [283].
- Wilmott, Howison and Dewynne, 1996 [282].
- Yan and Hanson, 2006 [287].
- Zhu, 2005 [289].
- Zhu and Hanson, 2005 [290] and 2006 [291].

Chapter 11

Applications in Mathematical Biology and Medicine

Despite assertions in both the lay and the professional literature, it is now known that normal physiology is anything but "regular." ...
Loss of event-to-event variability occurs during normal aging and also occurs pathologically in critical illiness.
—Dr. Timothy G. Buchman (2004) [45].

Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better
—Joel E. Cohen (2004) [58].

The application to optimal harvesting in uncertain environments is made in the presence of both background Gaussian noise and catastrophic jump events. Many problems in nature exhibit random effects and undergo catastrophic changes for which the stochastic calculus of continuous Wiener processes alone is inadequate.

11.1 Stochastic Bioeconomics: Optimal Harvesting Applications

For deterministic problems of optimal harvesting of renewable resources, the seminal reference by C. W. Clark is *Mathematical Bioeconomics: The Optimal Management of Renewable Resources* [56]. The book is nicely self-contained with introduction to the necessary economics, calculus of variations and optimal control theory. An excellent survey of stochastic bio-economics is given by Anderson and Sutinen in [9].

In this chapter, examples of optimal harvesting problems in random environments are illustrated. The first application is optimal harvesting with random population fluctuations [242]. A second application is optimal harvesting with random population fluctuations, but also with price fluctuations [115], so is a two-

dimensional state generalization of the first application.

11.1.1 Optimal Harvesting of Logistically Growing Population Undergoing Random Jumps

This problem of natural logistic growth of a renewable resource subject to random disasters and bonanzas was treated by Ryan and Hanson [242]. The parameter data was motivated by the **boom and bust** characteristics of Antarctic pelagic whaling at the time as studied by Clark and Lamberson [57]. The problem is summarized in the notation of this book, so for more information the reader should refer to [242].

Let $X(t)$ be the amount of biomass (mass of the biological species) of the harvested species at time t with stochastic dynamics given by

$$dX(t) = X(t)\left(r(1 - X(t)/K) - qU(t)\right)dt + X(t)\sum_{i=1}^{n_p}\nu_i dP_i(t), \qquad (11.1)$$

where $X(0) = x_0 > 0$ is the initial biomass, $r > 0$ is the constant intrinsic growth rate and $K > 0$ is the constant biomass carrying capacity that reflects the size of the population that the environment can support in absence of harvesting and other factors. Hence, the natural growth function $f(x) = rx(1 - x/K)$ is called the logistic function since as $x \to K$ a saturation effect due to crowding limits growth. Under the assumption of linear harvesting, the rate of harvesting is $H(t) = h(X(t), U(t)) = qU(t)X(t)$, where $U(t) \geq 0$ is the harvesting effort or rate and also the control variable, while $q > 0$ is called the **catchability coefficient** and is a measure of the efficiency of the harvest. The population suffers from rare random jumps from various sources for $i = 1 : n_p$ linear in the biomass $X(t)$ with proportions $-1 < \nu_i$. The negative values $-1 < \nu_i < 0$ denote disastrous effects but limited by a lower bound so that the population will not be wiped out by a single disaster, while the positive values $\nu_i > 0$ denote bonanzas or beneficial effects. The randomness of the jumps is modeled by a set of n_p Poisson processes $P_i(t)$ with common infinitesimal means and variances

$$\mathrm{E}[dP_i(t)] = \lambda_i dt = \mathrm{Var}[dP_i(t)],$$

for $i = 1 : n_p$, where $\lambda_i > 0$ is the ith jump rate. The actual jump at the jth jump time $t_{i,j}$ of the ith Poisson process is given in jump notation by

$$[X](t_{i,j}) \equiv X(t_{i,j}^+) - X(t_{i,j}^-) = \nu_i X(t_{i,j}^-)\,.$$

The motivation for the multitude of jump terms in (11.1) is that large random fluctuations can come from many causes, like climatic changes, over-fishing and epi-zootics (see [212, 138, 250, 242], for instance).

In [241], Ryan and Hanson treated the optimal harvesting case where the natural growth of the biomass is exponential with jumps, i.e., $1/K = 0$ and the natural growth function is linear, $f(x) = rx$, so the overall growth of $X(t)$ is exponential with harvesting and jumps. The model (11.1) is a pure jump model with logistic drift because the focus is on the effects of jumps on the harvesting

bioeconomics, although diffusion terms could have been easily added to the model. For $r > 0$ with no harvesting and jumps, the exponential model $dX(t) = rX(t)dt$ leads to unbounded exponential growth, while the logistic model $dX(t) = rX(t)(1 - X(t)/K)dt$ leads to saturated growth as $X(t) \to K^-$ if $x_0 < K$ or limiting decay as $X(t) \to K^+$ if $x_0 > K$. The density dependent (nonlinear) jump case is treated by Hanson and Ryan in [113].

The economic value of the harvest, starting at time t with biomass x and ending at the final time t_f, is given by the expected, discounted present value,

$$\overline{V}[X,U](x,t) = \mathrm{E}\left[\int_t^{t_f} e^{-\delta s} h(X(s),U(s))R(X(s),U(s))ds \,\middle|\, X(t)=x, U(t)=u\right], \tag{11.2}$$

where δ is the continuous, inflation-corrected discount rate with discounting starting at $t = 0$ and $\exp(-\delta t)$ is the discount factor which accounts for the opportunity costs of investing money elsewhere in a secure investment. The instantaneous net harvest revenue per unit harvest is

$$R(x,u) = \left(ph(x,u) - C(u)\right)/h(x,u).$$

It can be assumed that $x > 0$ and $u > 0$ to avoid dividing by zero, but the net revenue always appears in the product form $h(x,u)R(x,u)$, so the divide check is not needed. The price of a unit of a harvested biomass ($h = qux$) is p and

$$C(u) = c_1 u + c_2 u^2$$

is the total cost of the harvesting effort when the biomass or stock size is x given that $c_2 > 0$ so that $C(u)$ is a genuine quadratic. Note that $C(u)$ is assumed to be quadratic in the effort, which suggests that the effort is more costly the larger it becomes. In the case of fisheries, this might mean that more inefficient fishing boats or less experienced fisherman are used as the fishing effort $U(t)$ increases. The effort is assumed to be bounded, i.e., constrained, so that

$$0 \le U^{(\min)} \le U(t) \le U^{(\max)} < \infty \tag{11.3}$$

and the objective is to seek the maximum, expected current value

$$v^*(x,t) = \max_U \left[\, \overline{V}[X,U](x,t)\right].$$

Thus, the goal is to calculate optimal value $\overline{V}^*(x,t)$ and the optimal feedback control or effort

$$u^*(x,t) = \operatorname*{argmax}_U \left[\, \overline{V}[X,U](x,t)\right]$$

for $0 \le t < t_f$. However, the initial optimal expected, current value $\overline{V}^*(x,0)$ is the optimal expected, discounted present value of future revenues.

In order to facilitate the application of the Hamilton-Jacobi-Bellman (HJB) equation theorem 6.3 to the discounted current value form in (11.2) with the so-called **cost function** $C(\mathbf{x}, \mathbf{u}, t) = \exp(-\delta t)h(\mathbf{x}, u)R(\mathbf{x}, u)$ here, the discount factor $\exp(-\delta t)$ can be removed in the pseudo-Hamiltonian by converting from the present value $v^*(x,t)$ of Chapter 6 to the current value $\mathcal{V}^*(x,t)$ by the transformation

$$v^*(x,t) = \exp(-\delta t)\mathcal{V}^*(x,t).$$

Thus, $v_t^*(x,t) = \exp(-\delta t)(\mathcal{V}_t^*(x,t) - \delta t\, \mathcal{V}^*(x,t))$, where v_t^* and $\mathcal{V}_t^*$ are the partial derivatives of the value functions with respect to time. Note that initially both value functions coincide, $v^*(x,0) = \mathcal{V}^*(x,0)$.

Kamien and Schwartz [158] define the difference between the present and current value in terms of the present and current value Hamiltonians. The current value Hamiltonian $\mathcal{H}(x,u,t)$ is related to the present value Hamiltonian $\mathcal{H}(x,u,t)$ by

$$\begin{aligned}\mathcal{H}(x,u,t) &\equiv e^{+\delta t}\mathcal{H}(x,u,t)\\ &= (pqux - c_1u - c_2u^2) + (rx(1-x/K) - qux)\mathcal{V}_x^*(x,t)\\ &\quad + \sum_{i=1}^{n_p} \lambda_i(t)\left(\mathcal{V}^*(x+\nu_i x,t) - \mathcal{V}^*(x,t)\right),\end{aligned}$$

cancelling out the discount factor $\exp(-\delta t)$. Further, separating out the control terms, the HJBE is

$$\begin{aligned}0 &= \mathcal{V}_t^*(x,t) - \delta\mathcal{V}^*(x,t) + rx(1-x/K)\mathcal{V}_x^*(x,t) + \widehat{\mathcal{S}}^*(x,t)\\ &\quad + \sum_{i=1}^{n_p} \lambda_i(t)\left(\mathcal{V}^*(x+\nu_i x,t) - \mathcal{V}^*(x,t)\right),\end{aligned} \tag{11.4}$$

where the control switching term contains all control terms in the quadratic form:

$$\widehat{\mathcal{S}}(x,u,t) \equiv ((p - \mathcal{V}_x^*(x,t))qx - c_1 - c_2u)u\,,$$

including only the control dependent terms. The interior critical point of $\widehat{\mathcal{S}}(x,u,t)$ with respect to the control u is the regular optimal control,

$$u^{(\mathrm{reg})}(x,t) = \frac{0.5}{c_2}\left((p - \mathcal{V}_x^*(x,t))\,qx - c_1\right)\,, \tag{11.5}$$

since $c_2 > 0$, with the regular control being easily computed in terms of the gradient $\mathcal{V}_x^*(x,t)$ due to the quadratic cost assumption. As in the case of many applications, the control is constrained like in (11.3), so the constrained optimal control is the composite **bang-regular-bang** control function,

$$u^*(x,t) = \left\{ \begin{array}{ll} U^{(\min)}, & u^{(\mathrm{reg})}(x,t) \le U^{(\min)} \\ u^{(\mathrm{reg})}(x,t), & U^{(\min)} \le u^{(\mathrm{reg})}(x,t) \le U^{(\max)} \\ U^{(\max)}, & U^{(\max)} \le u^{(\mathrm{reg})}(x,t) \end{array} \right\}. \tag{11.6}$$

Consequently, the optimal control switch term is

$$\widehat{\mathcal{S}}^*(x,t) \equiv \widehat{\mathcal{S}}(x,u^*(x,t),t) = c_2 u^*(x,t)\left(2u^{(\mathrm{reg})}(x,t) - u^*(x,t)\right) ,$$

after some algebraic manipulations. When $u^{(\mathrm{reg})}(x,t)$ is within the constraints (11.3), the switch term will be quadratic in $u^{(\mathrm{reg})}(x,t)$, i.e., $\mathcal{S}^*(x,t) = c_2(u^{(\mathrm{reg})})^2(x,t)$, and consequently quadratic in the value gradient $\mathcal{V}_x^*(x,t)$, so the PDE of stochastic dynamic programming will be PDE with a quadratic nonlinearity. The gradient $\mathcal{V}_x^*(x,t)$ is the so-called **shadow price** [56] for the way it directly modifies the price p in (11.5) and represents the expected value of future harvests [56]. The PDE is also a partial differential-difference equation, since the discrete Poisson jumps lead to difference terms in (11.4) rather than the mark integral over difference terms as more generally presented in Chapter 6.

The final condition for the backward HJB equation is $\mathcal{V}^*(x,t_f) = 0$ for $x > 0$ in absence of salvage or terminal costs. Thus, the final regular control or effort at $t = t_f$ is given by

$$u^{(\mathrm{reg})}(x,t_f) = (pqx - c_1)/c_2 = c_1(x - x_f)/(2c_2x_f)$$

where $x_f \equiv c_1/(pq)$ is also the deterministic equilibrium stock value x_∞ [56]. However, in this stochastic case, if $c_1 \neq 0$, the final minimum control switch point is

$$x_{f,\min} = x_f\left(1 + 2c_2U^{(\min)}/c_1\right)$$

and the final maximum control switch point is

$$x_{f,\max} = x_f\left(1 + 2c_2U^{(\max)}/c_1\right).$$

As the biomass approaches extinction levels, $X(t) \to 0^+$, the rate of change $dX(t)$ (11.1) vanishes along with it, but the net revenue $R(x,u)$ should have become negative since costs dominate at low biomass. Hence, it will be assumed in addition that $R(x,u) \geq 0$, i.e., replacing $R(x,u)$ by $\max[R(x,u),0]$, so that the extinction boundary condition is

$$\mathcal{V}^*(0^+,t) = 0$$

for $0 < t < t_f$.

A very reasonable approximation to the solution can be obtained from the **quasi-deterministic approximation**,

$$\begin{aligned} dX^{(\mathrm{qdet})}(t) &\equiv \mathrm{E}\left[dX(t)|\, X(t) = X^{(\mathrm{qdet})}(t), U(t) = U^{(\mathrm{qdet})}(t)\right] \\ &= r^{(\mathrm{qdet})}X^{(\mathrm{qdet})}(t)\left(1 - X^{(\mathrm{qdet})}(t)/K^{(\mathrm{qdet})}\right)dt, \end{aligned}$$

where $r^{(\mathrm{qdet})} \equiv r + \sum_{i=1}^{n_p}\lambda_i\nu_i$ and $K^{(\mathrm{qdet})} \equiv Kr^{(\mathrm{qdet})}/r$, comprising an approximate logistic model. For this simplified model, the HJBE will no longer have difference terms since the jumps have been averaged out, but the optimal control will still be of the form (11.6).

Due to the complexity of the PDE, numerical methods are needed to approximate the solution. The HJBE can be solved with PDE-oriented finite difference methods [108] or the probability-oriented Markov chain approximation [179]. The finite difference method requires a sufficiently small mesh ratio for a comparison regular parabolic PDE [108] in the jump-diffusion case, while the Markov chain approximation, if the Markov chain probabilities are properly constructed, automatically comes with a weak convergence property [179]. For the current application in [242] and also in [241], Hanson and Ryan used the PDE-oriented finite difference method of [108] with predictor-corrector procedures to iterate on the nonlinear terms and precision-preserving interpolation to approximate the jump terms by values at neighboring finite difference nodes. Both methods are variations of the finite difference method and are summarized in Chapter 8 in Sections 8.1 for the PDE-oriented method and 8.2 for the Markov chain approximation, respectively.

The primary bio-economic parameters used in [242] come from [57], i.e., r, K, q, p and c_1, while other parameters like δ, t_f, λ_i and ν_i are reasonable estimates. Many of these estimated parameters were subjected to sensitivity tests in [242] in the many numerical results presented there. Some of the parameters are now obsolete, since whaling is no longer permitted in many countries or else highly restricted. Interest and discount rates are much lower now than then. However, significant sensitivity in u^* and V^* was found to the parameters δ, c_2 and $\lambda_i\nu_i/r$ for both a bonanza-dominated case with $\lambda_i\nu_i/r = 2\delta_{i,1}$ and a disaster-dominated case with $\lambda_i\nu_i/r = -0.5\delta_{i,2}$, where here $\delta_{i,j}$ is the Kronecker delta. In particular, in the **cheap control** limit as $c_2 \to 0^+$, the **bang-regular-bang** control law approaches a **bang-bang** control law in absence of a regular control component.

11.1.2 Optimal Harvesting with Both Price and Population Random Dynamics

The optimal harvesting problem, under joint population and price fluctuations in a random jump environment of Hanson and Ryan [115], is also an example of a two-dimensional state problem. Here, the problem is briefly summarized in the notation of this book. For general introduction to stochastic resource modeling, the reader can consult Anderson and Sutinen [9] or Mangel [193].

Again, let $X_1(t)$ be the amount of biomass (mass of the biological species) of the harvested species population at time t with stochastic dynamics consisting of logistic deterministic dynamics, discrete Poisson jumps and now with background, stochastic diffusion,

$$dX_1(t) = X_1(t)\Bigg((r_1(1 - X_1(t)/K_1) - q_1U_1(t))\,dt + \sigma_1 dW_1(t) + \sum_{i=1}^{n_1} \nu_{i,1} dP_{i,1}(t)\Bigg), \tag{11.7}$$

$X_1(0) = x_{1,0} > 0$, where the extra subscript 1 designates population parameters or processes, i.e., the essential biological component of the bio-economic process. The

diffusion process $\sigma_1 dW_1(t)$ satisfies zero mean and $\sigma_1^2 dt$ variance properties, with $\sigma_1 > 0$ assumed. For the Poisson processes, $\nu_{i,1} > -1$ and $\lambda_{i,1} > 0$ for $i = 1{:}n_1$.

The economic process or price process $p(t)$ depends on the time-dependent bio-mass harvest rate $H(t) = h(X_1(t), U_1(t)) = q_1 U_1(t) X_1(t)$ and other stochastic processes. Since on the average $p(t)$ decreases as $h(t)$ increases [115] following **supply-demand principles**, the price is assumed to satisfy

$$p(t) = (p_0/H(t) + p_1)\, X_2(t), \tag{11.8}$$

where p_0 is a constant supply-demand coefficient such that $p(t)H(t)$ is the gross harvest return, p_1 is the constant price per unit harvested bio-mass coefficient and $X_2(t)$ is a fluctuating inflationary factor [115] satisfying the SDE

$$dX_2(t) = X_2(t) \left(r_2 dt + \sigma_2 dW_2(t) + \sum_{i=1}^{n_2} \nu_{i,2} dP_{i,2}(t) \right), \tag{11.9}$$

$X_2(0) = x_{2,0} > 0$, where the extra subscript 2 designates parameters and processes in the price process SDE, $\sigma_2 > 0$, $\nu_{i,2} > -1$ and $\lambda_{i,2} > 0$ for $i = 1{:}n_2$. It is assumed that all primary stochastic processes, $P_{i,1}(t)$, $P_{i,2}(t)$, $W_1(t)$ and $W_2(t)$, are pair-wise independent.

The economic value of the harvest, starting at time t with biomass x_1 and ending at the final time t_f, is given by the expected, discounted current value,

$$\overline{V}[\mathbf{X}, U_1](\mathbf{x}, t) = \mathrm{E}\left[\int_t^{t_f} e^{-\delta s} H(s) R(\mathbf{X}(s), U_1(s)) ds \,\middle|\, \mathbf{X}(t) = \mathbf{x}, U_1(t) = u_1 \right], \tag{11.10}$$

where δ is the continuous, nominal discount rate, uncorrected by inflation since inflation is included in $X_2(t)$, with discounting starting at t. The random vector state is $\mathbf{X}(t) = [X_1(t)\ X_2(t)]^\top$ and $\mathbf{x} = [x_1\ x_2]^\top$ is a sampled vector state, such that

$$R(\mathbf{x}, u_1) = \left((p_0 + p_1 h(x_1, u_1))\, x_2 - C(u_1) \right) / h(x_1, u_1)$$

is the instantaneous net harvest revenue rate per unit biomass. It can be assumed that $x_1 > 0$ and $u_1 > 0$ to avoid dividing by zero, but the net revenue always appears in the product form $h(x_1, u_1) R(\mathbf{x}, u_1)$ so the zero check is unneeded. The price of a harvested biomass unit is p and

$$C(u_1) = c_1 u_1 + c_2 u_1^2$$

is the total cost of the harvesting effort given that $c_2 > 0$ so that $C(u_1)$ is a genuine quadratic.

The effort control constraints are again assumed to be

$$0 \le U_1^{(\min)} \le U_1(t) \le U_1^{(\max)} < \infty, \tag{11.11}$$

while the objective is to seek the maximum, expected current value

$$v^*(\mathbf{x}, t) = \max_{U_1} \left[\overline{V}[\mathbf{X}, U_1](\mathbf{x}, t) \right]$$

and the optimal feedback effort control

$$u_1^*(\mathbf{x},t) = \underset{U_1}{\operatorname{argmax}} \left[\overline{V}[\mathbf{X}, U_1](\mathbf{x},t) \right]$$

for $0 \le t < t_f$. Again, the present values $v^*(\mathbf{x},t)$ are transformed present values $\mathcal{V}^*(\mathbf{x},t)$,

$$v^*(\mathbf{x},t) = \exp(-\delta t)\mathcal{V}^*(\mathbf{x},t).$$

The present value Hamiltonian $\mathcal{H}(\mathbf{x},u_1,t)$ related to the current value Hamiltonian $\mathcal{H}(\mathbf{x},u_1,t)$ is

$$\begin{aligned}
\mathcal{H}(\mathbf{x},u_1,t) &\equiv e^{+\delta t}\mathcal{H}(\mathbf{x},u_1,t) \\
&= (p_0 + p_1 q_1 u_1 x_1)x_2 - c_1 u_1 - c_2 u_1^2 \\
&\quad +(r_1 x_1(1 - x_1/K_1) - q_1 u_1 x_1)\mathcal{V}^*_{x_1}(\mathbf{x},t) + r_2 x_2 \mathcal{V}^*_{x_2}(\mathbf{x},t) \\
&\quad +\frac{1}{2}\sigma_1^2 x_1^2 \mathcal{V}^*_{x_1,x_1}(\mathbf{x},t) + \frac{1}{2}\sigma_2^2 x_2^2 \mathcal{V}^*_{x_2,x_2}(\mathbf{x},t) \\
&\quad + \sum_{i=1}^{n_1} \lambda_{i,1}(t) \left(\mathcal{V}^*(1+\nu_{i,1})x_1, x_2, t) - \mathcal{V}^*(\mathbf{x},t) \right) \\
&\quad + \sum_{i=1}^{n_2} \lambda_{i,2}(t) \left(\mathcal{V}^*(x_1, (1+\nu_{i,2})x_2, t) - \mathcal{V}^*(\mathbf{x},t) \right) .
\end{aligned}$$

Upon cancelling out the discount factor $\exp(-\delta t)$ and separating out the control dependence from the Hamiltonian, the HJB equation is

$$\begin{aligned}
0 = {} & \mathcal{V}^*_t(\mathbf{x},t) - \delta \mathcal{V}^*(\mathbf{x},t) + p_0 x_2 + r_1 x_1(1 - x_1/K_1)\mathcal{V}^*_{x_1}(\mathbf{x},t) + r_2 x_2 \mathcal{V}^*_{x_2}(\mathbf{x},t) \\
& + \frac{1}{2}\sigma_1^2 x_1^2 \mathcal{V}^*_{x_1,x_1}(\mathbf{x},t) + \frac{1}{2}\sigma_2^2 x_2^2 \mathcal{V}^*_{x_2,x_2}(\mathbf{x},t) \\
& + \sum_{i=1}^{n_1} \lambda_{i,1}(t) \left(\mathcal{V}^*((1+\nu_{i,1})x_1, x_2, t) - \mathcal{V}^*(\mathbf{x},t) \right) \\
& + \sum_{i=1}^{n_2} \lambda_{i,2}(t) \left(\mathcal{V}^*(x_1, (1+\nu_{i,2})x_2, t) - \mathcal{V}^*(\mathbf{x},t) \right) \\
& + \widehat{\mathcal{S}}^*(\mathbf{x},t)
\end{aligned} \tag{11.12}$$

where the control switching term has the quadratic form:

$$\widehat{\mathcal{S}}(\mathbf{x},u_1,t) \equiv p_1 q_1 u_1 x_1 x_2 - c_1 u_1 - c_2 u_1^2 - q_1 u_1 x_1 \mathcal{V}^*_{x_1}(\mathbf{x},t),$$

including only the control dependent terms. The interior critical point of $\widehat{\mathcal{S}}(x,u,t)$ with respect to the control u is the regular optimal control,

$$u_1^{(\mathrm{reg})}(\mathbf{x},t) = \frac{0.5}{c_2} \left(\left(p_1 x_2 - \mathcal{V}^*_{x_1}(\mathbf{x},t) \right) q_1 x_1 - c_1 \right), \tag{11.13}$$

since $c_2 > 0$, with the regular control being easily computed in terms of the gradient $\mathcal{V}_x^*(\mathbf{x}, t)$ due to the quadratic cost assumption. As in the case of many applications, the control is constrained like in (11.11), so the constrained optimal control is the composite **bang-regular-bang** control function,

$$u_1^*(\mathbf{x}, t) = \left\{ \begin{array}{ll} U_1^{(\min)}, & u_1^{(\mathrm{reg})}(\mathbf{x}, t) \le U_1^{(\min)} \\ u_1^{(\mathrm{reg})}(\mathbf{x}, t), & U_1^{(\min)} \le u_1^{(\mathrm{reg})}(\mathbf{x}, t) \le U_1^{(\max)} \\ U_1^{(\max)}, & U_1^{(\max)} \le u_1^{(\mathrm{reg})}(\mathbf{x}, t) \end{array} \right\}. \tag{11.14}$$

The temporal side condition for the backward HJBE (11.12) is the final condition $\mathcal{V}^*(\mathbf{x}, t_f) = 0$ in absence of any terminal conditions for the first quadrant of state space and the natural corner condition

$$\mathcal{V}^*(0, 0, t) = -\left(c_1 + c_2 U_1^{(\min)}\right) U_1^{(\min)} (1 - \exp(-\delta(t_f - t))/\delta$$

at the origin $(0, 0)$ for $0 < t < t_f$, since $U_1^{(\min)} \ge 0$. On the edge $(x_1, 0)$ for $x_1 > 0$, the boundary condition is similar to solving the pure jump optimal resource HJBE of Subsection 11.1.1 except that there is an additional diffusion term. On the edge $(0, x_2)$ for $x_2 > 0$, the boundary condition also involves solving an even less similar HJBE since in this case the deterministic inflationary growth is exponential rather than logistic.

In [115], data of the International Pacific Halibut Commission (IPHC) annual reports [148] are used since the catch and price data were readily available over a long period of time. Other data came from Clark [56]. The hybrid exptrapolated-predictor-corrector Crank-Nicolson method similar to that described in the previous section and in Section 8.1was used. The major result was that large inflationary increases had a very strong effect on the optimal return but not on the optimal effort.

Another multidimensional optimal harvesting problem can be found in the Lake Michigan salmon-alewife predator-prey model of Hanson in [105], where the alewife suffered large scale die-offs every several years. The model was also mixed economically, since the salmon are fished recreationally while the alewife were fished in a commercial fishery, now disbanded. Multidimentional visualization and parallel processing for renewable resources are developed by developed Practico et al. [228] and Hanson et al. in [112].

11.2 Stochastic Biomedical Applications

Variability plays an important role in medicine. Discussing critical care, Buchman [45] emphasizes that variability is **normal** for the individual patient and that illness is often accompanied by loss of individual variability. For instance, Boker et al. [36] find a variable ventilator improved lung function during surgery and recovery more than a controlled constant ventilator. Priplata et al. [231] find that input noise enhances balance, particulary for the elderly. Ashkenazy et al. [14] present a stochastic model to portray the variation in an individual's gait showing that

variability changes with maturation and aging. Moss et al. [211] find increased sensitivity in detecting threshold levels with stochastic noise for stochastic resonance to occur for nonlinear neural systems during information processing.

Swan [261] presents many applications of optimal control to biomedicine in his book, but the emphasis is on deterministic compartment or ODE models. One chapter is on cancer therapy control and another is on drug administration control. Murray's [213, 214] two volumes on models of mathematical biology has information on cancer and other models, but no real optimal control models.

According to Steel [255], and Goldie and Coldman [99] stochastic effects play a important role in the stages of development of cancer, the subsequent growth and the invasiveness of tumors or the more liquid lymphomas. Mutations can be induced by environmental chemical agents or ionizing radiation, while spontaneous mutations are more rare, usually without obvious cause [99].

11.2.1 Diffusion Approximation of Tumor Growth and Tumor Doubling Time Application

Tumor Growth Branching Process

Sometimes approximating a discrete stochastic process by a diffusion process can be useful. Hanson and Tier [117] present an example for branching process for modeling the growth tumor cells. This discrete model is then approximated as a diffusion process for the purposes of analysis and computation.

Let B_i be the branching process, in the ith generation for $i = 1, 2, 3, \cdots$, such that there are three possible transitions in the time interval $(t, t+\Delta t)$ for generation i,

$$B_i = \left\{ \begin{array}{ll} 0, & \text{if cell death} \\ 1, & \text{if no cell change} \\ 2, & \text{if cell division} \end{array} \right\} , \tag{11.15}$$

similar to a birth-death model, but with a middle state of no change. This yields a total cancer cell count change from $N(t)$ at time t to

$$N(t + \Delta t) = \sum_{i=1}^{N(t)} B_i \, , \tag{11.16}$$

with the cell count change in $(t, t + \Delta t)$ being

$$\Delta N(t) = \sum_{i=1}^{N(t)} B_i - N(t) = \sum_{i=1}^{N(t)} (B_i - 1) \, .$$

The B_i are assumed to be independent, identically distributed (IID) random variables with basic conditional moments that are dependent on $N(t)$, i.e., density dependent,

$$\mathrm{E}[B_i \mid N(t) = n] = m(n; \Delta t)$$

and

$$\mathrm{Var}[B_i \mid N(t) = n] = v(n; \Delta t)\,.$$

The higher moments

$$\mathrm{E}[(B_i - m(N(t), \Delta t)^k \mid N(t) = n] = m_k(n; \Delta t)\,,$$

will also be needed to demonstrate that they will be small for $k \geq 3$.

Thus, the basic conditional moments of the tumor cell count change $\Delta N(t)$ are

$$\mathrm{E}[\Delta N(t) \mid N(t) = n] = \sum_{i=1}^{n} \mathrm{E}[B_i \mid N(t) = n] - n = n(m(n; \Delta t) - 1)$$

and

$$\begin{aligned}
\mathrm{Var}[\Delta N(t) \mid N(t) = n] &= \mathrm{E}[(\Delta N(t) - n(m(n; \Delta t) - 1))^2 \mid N(t) = n] \\
&= \mathrm{E}[(\sum_{i=1}^{n} (B_i - m(n; \Delta t))^2 \mid N(t) = n] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{E}[(B_i - m(n; \Delta t))(B_j - m(n; \Delta t)) \mid N(t) = n] \\
&= = \sum_{i=1}^{n} E[(B_i - m(n; \Delta t))^2 \mid N(t) = n] = nv(n; \Delta t)\,,
\end{aligned}$$

where the usual diagonalization technique has been used to apply the IID property of the B_i.

Diffusion Approximation of the Tumor Branching Process

Using some additional assumptions, a diffusion approximation will be constructed. Suppose T is some reference generation time such as the threshold for detection so

$$\tau = t/T$$

is a new scaled time and let a new scaled stochastic process be

$$X(\tau) = N(t)/T,$$

since the tumor will grow roughly as the number of generations. In order, for the model to be consistent with these scalings, the basic moments have to be refined so that the changes in $X(\tau)$ are small for small changes in τ. *The basic idea of the diffusion approximation is that it will not work well unless the order of the state changes are the same order as the time changes,* i.e., $\Delta X(\tau) = \mathrm{O}(\Delta\tau)$. Hence, let the infinitesimal mean be of the *near-replacement* form,

$$m(N(t), \Delta t) = 1 + (m_1(X(\tau)) + \mathrm{o}(1))\Delta\tau \quad \text{as} \quad \Delta\tau \to 0\,,$$

where $m_1(x)$ is a function to be specified, and let the infinitesimal variance be of the form

$$v(N(t), \Delta t) = (v_1(X(\tau)) + \mathrm{o}(1))T\Delta\tau \ \text{ as } \ \Delta\tau \to 0\,,$$

where $v_1(x)$ is a function to be specified. In addition, the higher moments should satisfy the form,

$$m_k(N(t), \Delta t) = \mathrm{o}(\Delta\tau) \ \text{ as } \ \Delta\tau \to 0\,.$$

First for the diffusion approximation, the infinitesimal moments of $X(\tau)$, with $\Delta X(\tau) = \Delta N(t)/T$ are computed as in (7.64-7.65),

$$\begin{aligned}\mu(x) &= \lim_{\Delta\tau\to 0} \frac{\mathrm{E}[\Delta X(\tau) \mid X(\tau) = x]}{\Delta\tau} = \lim_{\Delta\tau\to 0} \frac{\mathrm{E}[\Delta N(T\tau)/T \mid N(T\tau) = Tx]}{\Delta\tau} \\ &= \lim_{\Delta\tau\to 0} \frac{x(m(Tx, T\Delta\tau) - 1)}{\Delta\tau} = \lim_{\Delta\tau\to 0} [x(m_1(x) + \mathrm{o}(1))] = x m_1(x) \quad (11.17)\end{aligned}$$

and

$$\begin{aligned}\sigma^2(x) = \sigma_{1,1}(x) &= \lim_{\Delta\tau\to 0} \frac{\mathrm{Var}[\Delta X(\tau) \mid X(\tau) = x]}{\Delta\tau} \\ &= \lim_{\Delta\tau\to 0} [x(v_1(x) + \mathrm{o}(1))] = x v_1(x). \quad (11.18)\end{aligned}$$

In addition, the higher moment condition (7.66) when $k = 3$ is used (since any $k \geq 3$ can be used) in place of the continuity condition (7.63) due to the Chebyshev inequality (7.67). Hence,

$$\begin{aligned}\lim_{\Delta\tau\to 0} \frac{\mathrm{E}[|\Delta X(\tau) - x(m(Tx, T\Delta\tau) - 1)|^3 \mid X(\tau) = x]}{\Delta\tau} &= \lim_{\Delta\tau\to 0} \frac{n \cdot m_3(Tx, T\Delta\tau)}{T^3\Delta\tau} \\ &= \lim_{\Delta\tau\to 0} \frac{x \cdot \mathrm{o}(\Delta\tau)}{T^2\Delta\tau} = 0,\end{aligned}$$

completing the verification of the diffusion approximation and going substantially beyond the justification in [117].

For this particular application, the deterministic growth is chosen to be the **Gompertz growth model** [255, 99]

$$\mu(x) = x m_1(x) = \mu_1 x \ln(K/x)\,, \quad (11.19)$$

where μ_1 is a constant growth coefficient and K is the carrying capacity or saturation level. Note that the Gompertz growth is singular as $x \to 0^+$, in that its derivative is unbounded as $x \to 0^+$, since $d(\mu_1 x \ln(k/x))/dx = -\mu_1 \ln(ex/K) \to +\infty$. However, the Gompertz model is often used in analyzing cancer experiments, although other models are also used, such as the simpler exponential growth model in shorter clinical trials [99]. In addition, the infinitesimal variance is taken to be purely linear, i.e.,

$$\sigma^2(x) = x v_1(x) = \sigma_1 x\,,$$

where $\sigma_1 > 0$ is a constant.

In summary, the backward operator in this time homogeneous case is

$$\mathcal{B}_{x_0}[u](x_0) = \frac{1}{2}\sigma_1 x u''(x_0) + \mu_1 x_0 \ln(K/x_0) u'(x_0)\,. \quad (11.20)$$

Expected Tumor Doubling Time

The interest here is in the tumor doubling time, so suppose the tumor start is at the observed size c and then find the time it takes the tumor to double in size to $X(t) = 2c$. However, due to the stochastic nature of cancer, the tumor could become extinct, $X(t) = 0$ before it doubles in size. Hence, the proper problem is one of conditional probabilities, with the condition that the tumor doubles before it becomes extinct.

First consider the exit time at $2c$ starting at the general size $x_0 > 0$ at time t_0,

$$\tau_e^{(2c)}(x_0, t_0) = \inf_t [\ t \mid X(t) = 2c, X(s) \in (0^+, 2c), t_0 \le s < t, X(t_0) = x_0]\,, \quad (11.21)$$

so the backward formulation of Subsection 7.7.1 can be used with variable x_0, here with $b = 2c$. Again let the exit time distribution function be

$$\Phi_{\tau_e^{(2c)}(x_0,t_0)}(t) = \text{Prob}[\tau_e^{(2c)}(x_0, t_0) < t]$$

with corresponding density $\phi_{\tau_e^{(2c)}(x_0,t_0)}(t)$ and let the ultimate probability of exit at $X(t) = 2c$ be

$$\Phi_e^{(2c)}(x_0, t_0) = \int_0^\infty \phi_{\tau_e^{(2c)}(x_0,t_0)}(t)dt\,.$$

Consequently, the final answer will be the expected doubling time

$$\Phi_{\text{dbl}}(c) = \Phi_e^{(2c)}(c, 0)\,,$$

eventually using the initial values $x_0 = c$ and $t_0 = 0$.

Now let $u = u_0(x_0) = \Phi_e^{(2c)}(x_0, 0)$ and this satisfies the homogeneous backward equation

$$\mathcal{B}_{x_0}[u_0](x_0) = \frac{1}{2}\sigma_1 x u_0''(x_0) + \mu_1 x_0 \ln(K/x_0)u_0'(x_0) = 0\,, \quad (11.22)$$

from (11.20) in particular and (7.59) in general, but with boundary conditions,

$$u_0(0^+) = 0 \ \text{ and } \ u_0(2c) = 1,$$

since an exit at $X(0) = 0^+$ is excluded under the conditioning and an exit at $X(t) = 2c$ is a certain conditional exit. Eq. (11.22) is integrable in u and $x_0 > 0$ by using an integrating factor or its inverse called the **Wronskian** (also called the diffusion **scale density**),

$$\begin{aligned} W(x_0) &\equiv \exp\left(-2\int^{x_0} \frac{\mu(x)}{\sigma^2(x)}dx\right) = \exp\left(-2\frac{\mu_1}{\sigma_1}\int^{x_0} \ln\left(\frac{K}{x}\right)dx\right) \\ &= \exp\left(-2\frac{\mu_1 x_0}{\sigma_1}\ln\left(\frac{K}{ex_0}\right)\right) = (\beta_1 x_0)^{\gamma_1 x_0} \end{aligned} \quad (11.23)$$

here for the Gompertz model, where $\gamma_1 = 2\mu_1/\sigma_1$ and $\beta_1 = e/K > 0$. Thus, (11.22) simplifies to

$$(u_0'/W)'(x_0) = 0.$$

Thus, after two integrations and boundary condition substitutions lead to the solution of the boundary value problem,

$$\Phi_e^{(2c)}(x_0, 0) = u_0(x_0) = \frac{\int_{0^+}^{x_0} W(x)dx}{\int_{0^+}^{2c} W(y)dy}. \tag{11.24}$$

Since as $x \to 0^+$, $W(x) = (\beta_1 x)^{\gamma_1 x} \sim 1 + \gamma_1 x \ln(\beta_1 x)$ and then

$$\int_{0^+}^{x} dy W(y) \sim x + 0.5\gamma_1 x^2(\ln(\beta_1 x) - 0.5),$$

$W(x)$ is integrable as $x \to 0^+$ so that (11.24) is well defined, all other points on $(0, 2c)$ being obviously regular or non-singular points. Thus, setting $x_0 = c$ as the initial size gives the ultimate probability of a tumor doubling in size, $\Phi_e^{(2c)}(c, 0)$. More results by way of numerical and asymptotic approximations are given in [117].

The expected doubling time from (7.61) is

$$T_e^{(2c)}(c) = M_e^{(2c)}(c)/\Phi_e^{(2c)}(c, 0), \tag{11.25}$$

normalizing the first moment from (7.60), which here is

$$M_e^{(2c)}(x_0) \equiv \int_0^{+\infty} t\phi_{\tau_e^{(2b)}(x_0,0)}(t)dt,$$

for general initial size x_0 and satisfying the backward equation from (7.62)

$$\mathcal{B}_{x_0}\left[M_e^{(2c)}\right](x_0) = -\Phi_e^{(2c)}(x_0, 0)\,. \tag{11.26}$$

The backward equation for the moment is easier to solve than the one derived for the expected time quotient (11.25) since the quotient leads to a much more complicated equation. The boundary conditions are homogeneous,

$$M_e^{(2c)}(0^+) = 0 \quad \text{and} \quad M_e^{(2c)}(2c) = 0,$$

but for different reasons, the first because 0^+ is the excluded exit and the second because it means an instant exit.

The solution can again utilize the Wronskian as a reciprocal integrating factor, such that

$$(u_0'/W)'(x_0) = -2V(x_0)u_0(x_0)\,,$$

where

$$V(x) \equiv \frac{1}{\sigma^2(x)W(x)} = \frac{1}{\sigma_1 x(\beta_1 x)^{\gamma_1}}\,,$$

here for the Gompertz model, is called the **speed density**. As $x \to 0+$,

$$V(x) \sim \frac{1}{\sigma_1 x}(1 - \gamma_1 x \ln(\beta_1 x)),$$

so that for $0 < \epsilon \ll x \ll 1$,

$$\int_{\epsilon}^{x} dy V(y) \sim \sigma^{-1}(\ln(x/\epsilon) + \gamma_1 \epsilon \ln(\beta_1 \epsilon) + 1))$$

and is not integrable as $\epsilon \to 0^+$. The integrability of both $W(x)$ and $V(x)$, as well as that of some other functions, plays role in the Feller classification of boundaries for the Kolmogorov equations in one-dimension [31, 162]. Since a boundary is called a **regular boundary** if both $W(x)$ and $V(x)$ are integrable as the boundary point is approached from the interior of the domain, then 0^+ is a non-regular or **singular boundary** [162].

After two integrations, substitution of the boundary conditions to eliminate constants of integration and some manipulation of the integral forms, the solution of (11.26) can be written in the form

$$\begin{aligned} u_1(x_0) = 2(1 - u_0(x_0)) \int_{0^+}^{x_0} dy W(y) \int_{y}^{2c} dz V(z) u_0(z) \\ -2u_0(x_0) \int_{x_0}^{2c} dy W(y) \int_{y}^{2c} dz V(z) u_0(z) \,, \end{aligned} \tag{11.27}$$

provided the integrals exist. Letting $x_0 = c$, the expected doubling time is given by the formula in (11.25) or more simply by.

$$T_e^{(2c)}(c) = u_1(c)/u_0(c) \,.$$

The multiple integral form of the solution (11.27) is too complicated to analyze further here, but additional numerical and asymptotic results are given in the paper of Hanson and Tier [117], including deterministic results. The application in [117] is based upon Fortner plasmacytoma data of Simpson-Herren and Lloyd [249]. The presentation here is somewhat different since it needed to be consistent with the notation and analytical formulation of this text.

Related formulation and results for other optimal stopping problem are some extinction problems for stochastic populations. They are examined for both diffusion in [116] and Poisson noise in [119, 121].

11.2.2 Optimal Drug Delivery to Brain PDE Model

In many applications, the control problem is formulated in terms of partial differential equations (PDEs), not ordinary differential equations (ODEs), since the problem depends on spatial variations and not just time variations. The ODE driven control problem is usually called **lumped parameter control**, sometimes arising from **compartmental models** lumping the spatial variables so that a PDE is not used, while the PDE driven control model is called **distributed parameter control**. The parameters in this latter case refer to the spatial variables in the background of the control problem. The mathematical background to this problem can be found in Section A.5 or in Gunzberger [101] in much more detail for flow problems.

Cancer drug delivery to eliminate or reduce tumors is usually based upon expensive sets of experiments using animal and later human subjects to determine a fixed dose size and dose period to fit general patient, tumor and drug characteristics. Brain tumors can be very invasive and deadly, especially gliomas [262, 214]. When possible, the most of the mass of the tumor is removed (also called resectioned), but drug chemotherapy or radiotherapy is used in an attempt to kill any remaining cancer cells, including mobile metastases [80]. Gliomas can also be very diffusive [214], so reaction-diffusion equations may be used to model the drug delivery to the brain [262, 214, 92]. However, these reaction-diffusion investigations are only studies of the behavior of the solutions. No control of the drug delivery is involved. In this subsection, the paper of Chakrabarty and Hanson [48] on the control of reaction diffusion equations for optimal drug delivery to the brain is briefly summarized.

Optimal Control Problem for Drug Delivery Reaction-Diffusion Equations

Consider a reaction-diffusion model of a three-state system consisting of tumor cells, normal cells and cancer drug concentration in a brain. Let $y_1(\mathbf{x},t)$ be the density of remaining tumor cells, $y_2(\mathbf{x},t)$ be the density of normal cells and $y_3(\mathbf{x},t)$ be the concentration of the drug at time t in time horizon $[0,t_f]$ and position $\mathbf{x}$ in the brain domain $\mathcal{D}_x$. Let $\mathbf{y}(\mathbf{x},t) = [y_3(\mathbf{x},t)]_{3\times 1}$ be the global state vector.

The tumor cell density satisfies the coupled reaction-diffusion equation

$$\frac{\partial y_1}{\partial t}(\mathbf{x},t) = D_1\nabla_x^2[y_1](\mathbf{x},t) + a_1y_1g_1(y_1) - (\alpha_{1,2}y_2 + \kappa_{1,3}y_3)y_1 \qquad (11.28)$$

and the normal cells satisfy a similar equation

$$\frac{\partial y_2}{\partial t}(\mathbf{x},t) = D_2\nabla_x^2[y_3](\mathbf{x},t) + a_2y_2g_2(y_2) - (\alpha_{2,1}y_1 + \kappa_{2,3}y_3)y_2\,, \qquad (11.29)$$

where D_i is the diffusion coefficient for the ith state, the $a_iy_ig_i(y_i)$ is the growth law for the ith state, the interaction coefficient $\alpha_{i,j} > 0$ signifies a constant death rate of tissue of state i due to tissue state j and the coefficient $\kappa_{i,3} > 0$ denotes a constant death rate due to the drug. For concreteness, the growth terms are taken to be logistic, i.e., $a_iy_ig_i(y_i) = a_iy_i(1 - y_i/K_i)$, where a_i is a constant intrinsic growth coefficient and $K_i > 0$ is a constant carrying-capacity or saturation level. Thus, there can be a strong interaction between the tumor and normal tissues, but the drug interaction is uni-directional. The drug concentration $y_3(\mathbf{x},t)$ diffuses, gets absorbed and is controlled according to this reaction diffusion equation,

$$\frac{\partial y_3}{\partial t}(\mathbf{x},t) = D_3\nabla_x^2[y_3](\mathbf{x},t) + a_3y_3g_3(y_3) + u(\mathbf{x},t)\,, \qquad (11.30)$$

where $a_3y_3g_3(y_3)$ is the drug absorption loss term and $u(\mathbf{x},t)$ is the drug input control variable. For simplicity, the absorption term is taken to be exponential decay, so $a_3y_3g_3(y_3) = a_3y_3$, where $a_3 < 0$ is the negative of the absorption coefficient and is assumed constant.

The vector reaction-diffusion PDE form merging (11.28,11.29,11.30) corresponding to (A.138) is

$$\frac{\partial \mathbf{y}}{\partial t}(\mathbf{x},t) = D\nabla_x^2[\mathbf{y}](\mathbf{x},t)+\mathbf{B}(\mathbf{y}(\mathbf{x},t),\mathbf{x},t)+A\mathbf{u}(\mathbf{x},t)\,, \tag{11.31}$$

where $D = [D_i\delta_{i,j}]_{3\times 3}$ is the diffusion coefficient,

$$\begin{aligned}\mathbf{B}(\mathbf{y}(\mathbf{x},t),\mathbf{x},t) = &\left(a_1y_1(1-y_1) - (\alpha_{1,2}y_2+\kappa_{1,3}y_3)y_2\right)\mathbf{e}_1\mathbf{e}_1^\top\\ &+\left(a_2y_2(1-y_2) - (\alpha_{2,1}y_1+\kappa_{2,3}y_3)y_2\right)\mathbf{e}_2\mathbf{e}_2^\top\\ &+a_3y_1\mathbf{e}_3\mathbf{e}_3^\top\end{aligned} \tag{11.32}$$

is the bilinear reaction term with unit vectors $\mathbf{e}_k = [\delta_{i,k}]_{3\times 1}$ for $k = 1{:}3$, $A = \mathbf{e}_3\mathbf{e}_3^\top$ is the unit drug control coefficient and the drift term does not appear since $C \equiv 0$ here. The initial conditions for the vector PDE (11.31) is the vector

$$\mathbf{y}(\mathbf{x},0) = \mathbf{y}_0(\mathbf{x}),\ \text{for } \mathbf{x}\in\mathcal{D}_x \tag{11.33}$$

and the boundary condition is a no-flux condition,

$$-(\widehat{\mathbf{n}}^\top\nabla_x)[\mathbf{y}](\mathbf{x},t) = \mathbf{0}\,, \tag{11.34}$$

where $\widehat{\mathbf{n}} = \widehat{\mathbf{n}}(\mathbf{x},t)$ is the outward normal to the boundary $\partial\mathcal{D}_x$.

An objective in space-time is the minimization of the quadratic costs form,

$$\begin{aligned}V[\mathbf{y},\mathbf{u}] = &\frac{1}{2}\int_{t_0}^{t_f}dt\int_{\mathcal{D}_x}d\mathbf{x}\left(\mathbf{y}^\top Q\mathbf{y}+(\mathbf{u}-\mathbf{u}_0)^\top R(\mathbf{u}-\mathbf{u}_0)\right)(\mathbf{x},t)\\ &+\frac{1}{2}\int_{\mathcal{D}_x}d\mathbf{x}\left(\mathbf{y}^\top S\mathbf{y}\right)(\mathbf{x},t_f)\,,\end{aligned} \tag{11.35}$$

which is a slight variation in the control of the form (A.139), where the quadratic coefficients are $R = r_3\mathbf{e}_3\mathbf{e}_3^\top$ for the tumor burden cost, $S = s_1\mathbf{e}_1\mathbf{e}_1^\top$ for the drug delivery costs and $Q = q_1\mathbf{e}_1\mathbf{e}_1^\top + q_3\mathbf{e}_3\mathbf{e}_3^\top$ for the terminal costs, while the target threshold control value is $\mathbf{u}_0 = u_{0,3}\mathbf{e}_3$.

Hamiltonian Variational Formulation

The optimization problem above has three sets of constraints: the dynamics (11.31), the initial condition (11.33) and the boundary condition (11.34), so requires three Lagrange multipliers $\boldsymbol{\lambda}(\mathbf{x},t)$, $\boldsymbol{\mu}(\mathbf{x},t)$ and $\boldsymbol{\nu}(\mathbf{x})$ (without t since $t = 0$ for the initial condition), respectively, to form the **pseudo-Hamiltonian** as in (A.140),

$$\begin{aligned}\mathcal{H}(\mathbf{y},\mathbf{u},\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu}) = V[\mathbf{y},\mathbf{u}] &+ \int_{t_0}^{t_f}dt\int_{\mathcal{D}_x}d\mathbf{x}\,\boldsymbol{\lambda}^\top\big(\mathbf{y}_t - D\nabla_x^2[\mathbf{y}] - \mathbf{B} - A\mathbf{u}\big)(\mathbf{x},t)\\ &+\int_{t_0}^{t_f}dt\int_{\partial\mathcal{D}_x}d\boldsymbol{\Gamma}\boldsymbol{\mu}^\top\big(-\left(\widehat{\mathbf{n}}^\top\nabla_x\right)[\mathbf{y}]\big)(\mathbf{x},t)\\ &+\int_{\mathcal{D}_x}d\mathbf{x}\,\boldsymbol{\nu}^\top\big(\mathbf{y}(\mathbf{x},0^+)-\mathbf{y}_0(\mathbf{x})\big)\ .\end{aligned} \tag{11.36}$$

The main idea is that the Lagrange multipliers extend the three-vector state space to an extended six-vector state space

$$\mathbf{z}(\mathbf{x},t) \equiv \{\mathbf{y}(\mathbf{x},t), \mathbf{u}(\mathbf{x},t), \boldsymbol{\lambda}(\mathbf{x},t), \boldsymbol{\mu}(\mathbf{x},t), \boldsymbol{\nu}(\mathbf{x})\}$$

to make the variations $\delta\mathbf{z}(\mathbf{x},t)$ about $zbf^*(\mathbf{x},t)$ in the extended objective systematic. Hence,

$$\mathcal{H}(\mathbf{z}^*(\mathbf{x},t) + \delta\mathbf{z}(\mathbf{x},t)) = \mathcal{H}(\mathbf{z}^*(\mathbf{x},t)) + \delta\mathcal{H}(\mathbf{z}^*(\mathbf{x},t), \delta\mathbf{z}(\mathbf{x},t)) + \mathrm{O}(|\delta\mathbf{z}|^2(\mathbf{x},t))\,,$$

assuming that $\mathbf{z}^*(\mathbf{x},t)$ exists and is a unique optimal solution under sufficient differentiability assumptions on $\mathcal{H}(\mathbf{z}(\mathbf{x},t))$. Critical to these assumptions is that the perturbation of the nonlinear reaction term $\mathbf{B}(\mathbf{y},\mathbf{x},t)$ has a quadratic approximation, but that is trivial for this application since $\mathbf{B}$ is quadratic in $\mathbf{y}$.

Skipping the details contained in Subsection A.5.2, something very similar to the first variation $\delta\mathcal{H}(\mathbf{z}^*(\mathbf{x},t), \delta\mathbf{z}(\mathbf{x},t))$ in (A.142) is found. Setting the coefficients of $\delta\boldsymbol{\lambda}^\top(\mathbf{x},t)$, $\delta\boldsymbol{\nu}^\top(\mathbf{x})$ and $\delta\boldsymbol{\mu}^\top(\mathbf{x},t)$ (only for $\mathbf{x} \in \mathcal{D}_x$), respectively, to zero confirms that the PDE (11.31), initial condition (11.33) and boundary condition (11.34) hold with the optimal state $\mathbf{y}^*(\mathbf{x},t)$ replacing for the state $\mathbf{y}(\mathbf{x},t)$ of the original problem.

The final-boundary value PDE problem for the optimal adjoint state $\boldsymbol{\lambda}^*(\mathbf{x},t)$ comes from setting the coefficients for $\delta\mathbf{y}^\top(\mathbf{x},t_f)$, $\delta\mathbf{y}^\top(\mathbf{x},t_f)$ and $\delta\mathbf{y}^\top(\mathbf{x},t)$ (only for $\mathbf{x} \in \mathcal{D}_x$), respectively, to zero, producing

$$\left(\boldsymbol{\lambda}_t^* + \nabla_x^2[D\boldsymbol{\lambda}^*] - \nabla_y[\mathbf{B}^\top]^*\boldsymbol{\lambda}^* - Q\mathbf{y}^*\right)(\mathbf{x},t) = \mathbf{0}\,, \quad \mathbf{x} \in \mathcal{D}_x, \quad t \in [0,t_f)\,, \tag{11.37}$$

with final condition,

$$(\boldsymbol{\lambda}^* + S\mathbf{y}^*)(\mathbf{x},t_f) = \mathbf{0}, \quad \mathbf{x} \in \mathcal{D}_x\,, \tag{11.38}$$

and boundary condition

$$\left(\widehat{\mathbf{n}}^\top\nabla_x\right)[D\boldsymbol{\lambda}^*](\mathbf{x},t) = \mathbf{0}\,, \quad \mathbf{x} \in \partial\mathcal{D}_x, \quad t \in (0,t_f)\,, \tag{11.39}$$

which is the corresponding no-flux condition in backward form.

Setting the coefficient of $\delta\mathbf{u}(\mathbf{x},t)$ to zero leads to

$$R(\mathbf{u}^*(\mathbf{x},t) - \mathbf{u}_0^*(\mathbf{x},t)) = A^\top\boldsymbol{\lambda}(\mathbf{x},t)\,,$$

which reduces to

$$u_3^*(\mathbf{x},t) = u_{0,3}^*(\mathbf{x},t) + \lambda_3^*(\mathbf{x},t)/r_3\,, \quad \mathbf{x} \in \mathcal{D}_x, \quad t \in [0,t_f)\,, \tag{11.40}$$

There are other optimality conditions that interrelate the Lagrange multipliers,

$$\boldsymbol{\nu}^*(\mathbf{x}) = \boldsymbol{\lambda}^*(\mathbf{x},0^+) \text{ for } \mathbf{x} \in \mathcal{D}_x$$

and

$$\boldsymbol{\mu}^*(\mathbf{x},t) = D\boldsymbol{\lambda}^*(\mathbf{x},t)\,, \quad \mathbf{x} \in \partial\mathcal{D}_x, \quad t \in [0,t_f)\,,$$

which will not be needed in the computations.

Forward-Backward Computational Iterations

The presence of nonlinear reaction terms in the forward state equation (11.31) using $\mathbf{y}^*(\mathbf{x},t)$ with $\mathbf{u}^*(\mathbf{x},t)$ and in the corresponding backward co-state equation (11.37) for $\boldsymbol{\lambda}^*(\mathbf{x},t)$ make computational methods essential. The computational method of Chakrabarty and Hanson [48, 49, 50] employs a forward state integration of (11.31) and a backward integration of (11.37) with sufficient iterations until the norm of the iteration difference is small enough. The forward equation (11.31) is independent of the co-state $\boldsymbol{\lambda}^*(\mathbf{x},t)$ but depends on the optimal control $\mathbf{u}^*(\mathbf{x},t)$ which is a critical objective to be determined, so a starting guess for $\mathbf{u}^*(\mathbf{x},t)$ is needed to start the forward integration, until a backward itegration generates a better guess using (11.40). On the other hand, the backward equation (11.37) depends strongly on the state distribution $\mathbf{y}^*(\mathbf{x},t)$ as well as on its final values from (11.38), so that iterations, each consisting of a **double-shot** of both a forward interation followed by a backward iteration, are required for reasonable accuracy. This double shot method is similar to the **opposite directions** multiple shooting method of Hackbusch [103] for parabolic equations. Gunzberger [101] calls many such methods **one-shot** methods and give more rigorous justification of them.

In order to keep the computational presentation manageable, let the forward and backward PDEs be represented in the more compact notation:

$$\begin{aligned}\mathbf{y}_t^*(\mathbf{x},t) &= \boldsymbol{\mathcal{F}}(\mathbf{x},t,\mathbf{y}^*(\mathbf{x},t),\mathbf{u}^*(\mathbf{x},t)),\\ \mathbf{0} &= \boldsymbol{\lambda}_t^*(\mathbf{x},t)+\boldsymbol{\mathcal{G}}(\mathbf{x},t,\boldsymbol{\lambda}^*(\mathbf{x},t),\mathbf{y}^*(\mathbf{x},t)),\end{aligned}$$

with general vector functions $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{G}}$ for the forward and backward equations, respectively. Let the space vector $\mathbf{x}$ be replaced by the discrete representation,

$$\mathbf{x_j} \equiv [x_{j_i,1}+(j_i-1)\cdot\Delta x_i]_{3\times 1}$$

where Δx_i is the step size in the ith direction, $\mathbf{j}=[j_i]_{3\times 1}$ where, $j_i = 1{:}M_i$ nodes per direction, $i = 1{:}3$. Let the time t be replaced by the forward discretization

$$t_k \equiv k\Delta t,$$

for $k = 0{:}K$ time steps where Δt is the forward time step size, $t_0 = 0$ and $t_K = t_f$. The backward discrete time will be of the form $t_k^{(b)} \equiv t_f - k\Delta t = (K-k)\Delta t = t_{K-k}$. The corresponding discretization of the dependent vectors will be

$$\mathbf{y}(\mathbf{x_j},t_k) \simeq \mathbf{Y}_{\mathbf{j},k}\,,\boldsymbol{\lambda}(\mathbf{x_j},t_k)\simeq\boldsymbol{\Lambda}_{\mathbf{j},k} \;\text{ and }\; \mathbf{u}(\mathbf{x_j},t_k)\simeq\mathbf{U}_{\mathbf{j},k}\,.$$

The numerical procedure used is the Crank-Nicolson method for second order accuracy in both space and time, but modified with additional extrapolation, prediction and correction techniques to accommodate nonlinear terms and multi-dimensions. The forward and backward discrete versions are written,

$$\mathbf{Y}_{\mathbf{j},k+1}^{(\gamma+1,\ell)} = \mathbf{Y}_{\mathbf{j},k}^{(\ell)} + \Delta t\boldsymbol{\mathcal{F}}_{\mathbf{j},k+0.5}^{(\gamma,\ell)}\,, \tag{11.41}$$

$$\boldsymbol{\Lambda}_{\mathbf{j},k-1}^{(\gamma+1,\ell)} = \boldsymbol{\Lambda}_{\mathbf{j},k}^{(\ell)} + \Delta t\boldsymbol{\mathcal{G}}_{\mathbf{j},k-0.5}^{(\gamma,\ell)}\,, \tag{11.42}$$

for $\gamma = 0\!:\!n_c$ corrections ($\gamma = 0$ is the prediction step) in each time step k until a relative stopping criterion for corrections in the tumor cell state component $Y^{(\gamma+1,\ell)}_{1,\mathbf{j},k+1}$ is satisfied,

$$\left|\left|Y^{(\gamma+1,\ell)}_{1,\mathbf{j},k+1} - Y^{(\gamma,\ell)}_{1,\mathbf{j},k+1}\right|\right| < \mathrm{tol}_y \left|\left|Y^{(\gamma,\ell)}_{1,\mathbf{j},k+1}\right|\right| \tag{11.43}$$

for every state index $\mathbf{j}$, for $k = 0 : K-1$ and during all double shot iterations $\ell = 1\!:\!L$, provided $||Y^{(\gamma,\ell)}_{1,\mathbf{j},k+1}|| \neq 0$. The general notation means that

$$\boldsymbol{\mathcal{F}}^{(\gamma,\ell)}_{\mathbf{j},k+0.5} = \boldsymbol{\mathcal{F}}\left(\mathbf{x}_\mathbf{j}, t_{k+0.5}, \mathbf{Y}^{(\gamma,\ell)}_{\mathbf{j},k+0.5}, \mathbf{U}^{(\gamma,\ell)}_{\mathbf{j},k+0.5}\right)$$

and similarly for $\boldsymbol{\mathcal{G}}^{(\gamma,\ell)}_{\mathbf{j},k-0.5}$. The relative tolerance in $Y^{(\gamma,\ell)}_{1,\mathbf{j},k}$ is tol_y. The Crank-Nicolson midpoint values are ordinarily approximated by the average,

$$\mathbf{Y}^{(\gamma,\ell)}_{\mathbf{j},k+0.5} \simeq 0.5\left(\mathbf{Y}^{(\gamma,\ell)}_{\mathbf{j},k+1} + \mathbf{Y}^{(\ell)}_{\mathbf{j},k}\right)$$

for $k = 0\!:\!K-1$ and

$$\mathbf{\Lambda}^{(\gamma,\ell)}_{\mathbf{j},k-0.5} \simeq 0.5\left(\mathbf{\Lambda}^{(\gamma,\ell)}_{\mathbf{j},k-1} + \mathbf{\Lambda}^{(\ell)}_{\mathbf{j},k}\right),$$

for $k = K : -1 : 1$, where $\mathbf{Y}^{(\ell)}_{\mathbf{j},k}$ and $\mathbf{\Lambda}^{(\ell)}_{\mathbf{j},k}$ are the final corrections for each time step k given shot ℓ, consistent with the second order Crank-Nicolson accuracy and implicitness reduction. A similar form is used for $\mathbf{U}^{(\gamma,\ell)}_{\mathbf{j},k+0.5}$. Second order central finite differences are used for all derivatives and based upon $\mathbf{Y}^{(\gamma,\ell)}_{\mathbf{j},k+0.5}$ or $\mathbf{\Lambda}^{(\gamma,\ell)}_{\mathbf{j},k-0.5}$.

The final stopping criterion for the convergence of the double shot iterations $\ell = 2\!:\!L$ is the pair of norms,

$$\left|\left|U^{(\ell)}_{3,\mathbf{j},k} - U^{(\ell-1)}_{3,\mathbf{j},k}\right|\right| < \mathrm{tol}_u \left|\left|U^{(\ell-1)}_{3,\mathbf{j},k}\right|\right| \quad \text{and} \quad \left|\left|Y^{(\ell)}_{1,\mathbf{j},k} - Y^{(\ell-1)}_{1,\mathbf{j},k}\right|\right| < \mathrm{tol}_y \left|\left|Y^{(\ell-1)}_{1,\mathbf{j},k}\right|\right|, \tag{11.44}$$

where the norm is over all $\mathbf{j}$ and k, for $\ell = 2\!:\!L$ until satisfied, provided $||U^{(\ell-1)}_{3,\mathbf{j},k}|| \neq 0$ and $||Y^{(\ell-1)}_{1,\mathbf{j},k}|| \neq 0$, where $\mathrm{tol}_u > 0$ and $\mathrm{tol}_y > 0$ are some specified tolerances.

The treatment of the bilinear reaction term (11.32) requires careful consideration to accommodate the usual linear framework of the Crank-Nicolson method. Since this term has the pure bilinear form,

$$\mathbf{B}(\mathbf{y}, \mathbf{x}, t) = \widehat{B}(\mathbf{y})\mathbf{y}\,,$$

in this application, this quasi-linear approximation is very appropriate

$$\widehat{B}\left(\mathbf{Y}^{(\gamma,\ell)}_{\mathbf{j},k+0.5}\right)\mathbf{Y}^{(\gamma,\ell)}_{\mathbf{j},k+0.5} \simeq \widehat{B}\left(\mathbf{Y}^{(\gamma-1,\ell)}_{\mathbf{j},k+0.5}\right)\mathbf{Y}^{(\gamma,\ell)}_{\mathbf{j},k+0.5}\,,$$

in the forward equation for corrections $\gamma \geq 1$ and time steps $k \geq 1$.

Another special treatment needed is that of the no-flux boundary condition since central differences are inappropriate at the boundary, but backward and for-

ward differences of the same second order accuracy work very well, e.g.,

$$\mathbf{0} = -\left((\widehat{\mathbf{n}}^\top \nabla_x)[\mathbf{Y}^*]\right)_{\mathbf{j},k}^{(\gamma,\ell)} \simeq -\frac{\left(3\mathbf{Y}_{\mathbf{j},k}^{(\gamma,\ell)} - 4\mathbf{Y}_{\mathbf{j}-\widehat{\mathbf{n}},k}^{(\ell)} + \mathbf{Y}_{\mathbf{j}-2\widehat{\mathbf{n}},k}^{(\gamma,\ell)}\right)}{(2|\widehat{\mathbf{n}}^\top \Delta\mathbf{x}|)},$$

$$\mathbf{0} = \left((\widehat{\mathbf{n}}^\top \nabla_x)[(\mathbf{\Lambda})^*]\right)_{\mathbf{j},k}^{(\gamma,\ell)} \simeq +\frac{\left(3\mathbf{\Lambda}_{\mathbf{j},k}^{(\gamma,\ell)} - 4\mathbf{\Lambda}_{\mathbf{j}-\widehat{\mathbf{n}},k}^{(\gamma,\ell)} + \mathbf{\Lambda}_{\mathbf{j}-2\widehat{\mathbf{n}},k}^{(\gamma,\ell)}\right)}{(2|\widehat{\mathbf{n}}^\top \Delta\mathbf{x}|)},$$

respectively, where $\widehat{\mathbf{n}} \equiv \widehat{\mathbf{n}}_{\mathbf{j},k}$, $\Delta\mathbf{x} = [\Delta x_i]_{3\times 1} > \mathbf{0}$ and, e.g.,

$$\mathbf{Y}_{\mathbf{j}-\widehat{\mathbf{n}},k}^{(\gamma,\ell)} = \mathbf{Y}^{(\gamma,\ell)}(\mathbf{x}_\mathbf{j} - |\widehat{\mathbf{n}}^\top \Delta\mathbf{x}|\widehat{\mathbf{n}}, t_k).$$

A sample output of the computations in Fig. 11.1 shows significant decrease in tumor size in one space dimension for a five day drug treatment trial. For information on the parameters used see Chakrabarty and Hanson [48]. For the corresponding two-dimensional space model of drug delivery see [49].

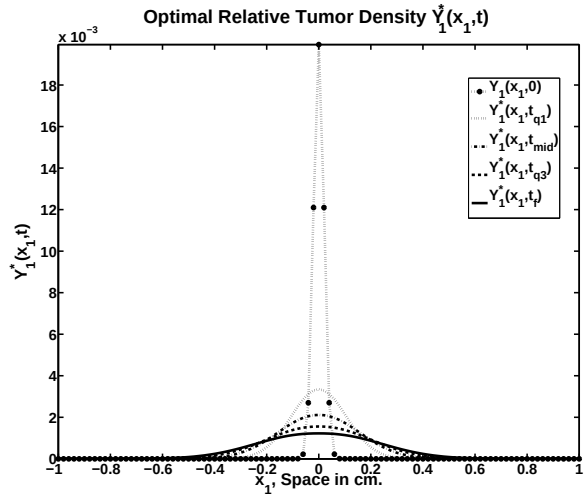


Figure 11.1. *Optimal tumor density $Y_1^*(x_1, t)$ in the one-dimensional case with time as a parameter rounded at quartile values $\{0, t_{q_1} = t_f/4, t_{\text{mid}} = t_f/2, t_{q_3} = 3t_f/4, t_f\}$, where $t_f = 5$ days. The total tumor density integral is reduced by 29% in the 5-day simulated drug treatment trial.*

Suggested References for Further Reading

- Anderson and Sutinen, 1984 [9].
- Ashkenazy et al., 2002 [14].
- Bharucha-Reid, 1960 [31].

- Boker et al., 2004 [36].
- Buchman, 2004 [45].
- Chakrabarty and Hanson, 2005a [48], 2005b [49] and 2006 [50].
- Clark, 1976 and 1990 [56].
- Clark and Lamberson, 1982 [57].
- Engelhard, 2000 [80].
- Goel and Dyn, 1974 [98].
- Goldie and Coldman, 1998 [99].
- Gunzberger, 2003 [101].
- Hackbusch, 1978 [103].
- Hanson, 1987 [105] and 1996 [108].
- Hanson et al., 1993 [112].
- Hanson and Ryan, 1988 [113] and 1998 [115].
- Hanson and Tier, 1981 [116] and 1982 [117].
- Hanson and Tuckwell, 1978 [118], 1981 [119], 1983 [120] and 1997 [121].
- Hennemuth et al., 1980 [138].
- Karlin and Taylor, 1975 [161, I] and 1981 [162, II].
- Kamien and Schwartz, 1981 [158].
- Kushner and Dupuis, 2001 [179].
- Ludwig, 1974 [187] and [188].
- Mangel, 1985 [193].
- Moss et al., 2004 [211].
- Murray, 2002 [213, I] and 2003 [214, II].
- Nisbet and Gurney, 1982 [219].
- Practico et al., 1992 [228].
- Priplata et al., 2002 [231].
- Ryan and Hanson, 1958 [241] and 1986 [242].
- Simpson-Herren and Lloyd, 1970 [249].

- Steel, 1977 [255].
- Swan, 1984 [261].
- Swanson, 1999 [262].
- Taylor and Karlin, 1998 [265].
- Tuckwell, 1989 [269] and 1995 [270].

Chapter 12

Applied Guide to Abstract Theory of Stochastic Processes:

Mathematicians are like Frenchmen: Whatever you say to them they translate into their own language and forthwith it is something entirely different.
—Johann Wolfgang von Goethe.

Since the mathematicians have invaded the theory of relativity, I do not understand it myself anymore.
—Albert Einstein (1879-1955), `http://en.wikiquote.org/wiki/Mathematics`.

Martingale (1589): Any of several systems of betting in which a player increases his stake, usually by doubling each time he loses a bet.
—Digital Webster, definition 3, 1992.

Martingales are treated because of their great importance, but they are not used as a tool in this book.
—William (Willy) Feller (1906-1970), p. 209 in [84].

The concept of martingales is due to P. Lévy, but it was J. L. Doob who realized its unexpected potential and developed the theory.
—William (Willy) Feller (1906-1970), p. 210 in [84].

Our view of Brownian motion never focused too closely on the underlying measure space, and, by and large, we have profited from keeping a respectful distance.
—J. Michael Steele, p. 218 in [256].

This chapter briefly introduces more of the abstract analytical methods, such as measure theoretic methods, Martingale methods, Radon-Nikodým derivatives, Girsanov's theorem, Itô processes, Lévy processes, characteristic functions and exponents, Lévy-Klintchine formula, jump-diffusion process comparisons and other topics from the applied point of view as a bridge to more abstract methods.

The purpose of this chapter is to supply some insightful and useful background to make the more abstract literature on stochastic processes and control more accessible to general students in applied mathematics, statistics, computer science, applied science and engineering. The overall approach in this book is designed to start from the common calculus and analysis constructs of entry level graduate students in these applied areas by evolving these constructs to those of applied stochastic processes and control, much as genes have evolved by small but powerful changes. However, students still need to understand the important results that come from using more abstract methods.

The applied motivation is to solve problems with a combination of analytical and computational methods. These problems may have great complexity in terms on nonlinearities in the state and other dependencies. It is necessary to train both students and researchers from a broad range of areas in science and engineering to solve large scale problems. In the abstract approach the emphasis is not necessarily to solve applied problems, but to prove existence, uniqueness and convergence, often in very abstract language. However, sometimes the conditions of the proofs are too restrictive, so as to exclude many complex and large scale applications. Proofs as such are not given in this chapter, but some formal applied derivations are given and readers can refer to the list of references at the end of the chapter for more rigorous treatments.

12.1 Very Basic Probability Measure Background

In order to keep things simple and concise, it will be necessary to compromise on completeness, but keep sufficient detail for a coherent story. The notation will be somewhat different from the usual, if there is such a thing as usual notation, so that we can avoid conflict with the stochastic process notation where possible. The symbols are also selected so that they are related to what the quantity signifies, where possible.

12.1.1 Mathematical Measure Theory Basics

The starting point will be some notions of measure theory and its algebra, called σ-algebra. Measure theory provides an abstract generalization of integration theory including expectations, and distributions that are based on counts, intervals, areas, volumes and mass to that of general sets. The ultimate goal is **probability measure**, but the presentation begins with the foundations in the more general mathematical measure theory.

Measure σ-Algebra Definition:

Let $\mathcal{U}$ be a nonempty set called the **universe**, but really is only the principal set of interest. Let $\mathbf{\Sigma}$ be a collection of subsets on $\mathcal{U}$.

Definition 12.1. *$\mathbf{\Sigma}$ is a **σ-algebra** if*

- *$\emptyset \in \mathbf{\Sigma}$, i.e., the empty set $\emptyset$ is included.*
- *$\mathcal{U} \in \mathbf{\Sigma}$, i.e., the universe $\mathcal{U}$ is included.*
- *The set $\mathcal{S} \in \mathbf{\Sigma} \Longrightarrow \mathcal{S}^c \in \mathbf{\Sigma}$, i.e., its complement $\mathcal{S}^c$ with respect to $\mathcal{U}$ is included too, i.e., verifying that $\mathcal{S} \cup \mathcal{S}^c = \mathcal{U}$.*
- *If $\{\mathcal{S}_i \in \mathbf{\Sigma} : i = 1:n\}$ is a sequence of subsets, then the union $\bigcup_{i=1}^{n} \mathcal{S}_i \in \mathbf{\Sigma}$, i.e., additive closure under unions.*
- *If so, then $\{\mathcal{U}, \mathbf{\Sigma}\}$ is called a **measurable space**.*

Often the symbol Ω is used for the general universe $\mathcal{U}$ and the symbol $\mathcal{F}$ is used for the σ-algebra $\mathbf{\Sigma}$. Recall that the **union** of two sets

$$\mathcal{S}_1 \cup \mathcal{S}_2 = \{\text{points } X : \ X \in \mathcal{S}_1 \text{ OR } X \in \mathcal{S}_2\},$$

the logical OR playing an important role when translated to probability measures.

A **Borel set** $\mathbf{\Sigma} = \mathcal{B} = \mathcal{B}(\mathbb{R}^{n_x})$ is the σ-algebra of open sets on $\mathcal{U} = \mathbb{R}^{n_x}$, so $\mathcal{B}(\mathbb{R}^{n_x})$ automatically contains all closed sets of $\mathbb{R}^{n_x}$ by complementarity.

Measure Definition:

Definition 12.2. ***The measure $\mathcal{M}$ is a function on the measurable space** $\{\mathcal{U}, \mathbf{\Sigma}\}$ **that maps** $\mathbf{\Sigma} \longrightarrow [0, \infty)$, such that*

- *$\mathcal{M}(\emptyset) = 0$, i.e., the empty set $\emptyset$ has **measure zero**.*
- *If for any subset $\mathcal{S} \in \mathbf{\Sigma}$, then $\mathcal{M}(\mathcal{S}) \geq 0$, i.e., **nonnegativity**, as in mass.*
- *If $\{\mathcal{S}_i \in \mathbf{\Sigma} : i = 1, 2, \ldots,\}$ is any countable sequence of **disjoint subsets** (i.e., $\mathcal{S}_i \bigcap \mathcal{S}_j = \emptyset$, $i \neq j$, the intersection is empty), then the measure of the union is the sum of the measures,*

$$\mathcal{M}\left(\bigcup_{i=1}^{\infty} \mathcal{S}_i\right) = \sum_{i=1}^{\infty} \mathcal{M}(\mathcal{S}_i), \tag{12.1}$$

*i.e., **countable additivity**, as in preserving mass under partitioning.*

The triplet $\{\mathcal{U}, \mathbf{\Sigma}, \mathcal{M}\}$ is called a **measure space**. Often the symbol μ is used for the general measure symbol $\mathcal{M}$ used here, but the former conflicts with the use of μ as the mean or drift in this book. Recall that the **intersection** of two sets

$$\mathcal{S}_1 \cap \mathcal{S}_2 = \{\text{points } X: \ X \in \mathcal{S}_1 \text{ AND } X \in \mathcal{S}_2\},$$

the logical AND playing an important role when translated to probability measures.

The nonnegativity measure property $\mathcal{M}(\mathcal{S}) \geq 0$ means that **positive measure** has been defined. Positive meaures, among other things, facilitate convergence proofs, i.e., monotone convergence. However, if for any subset $\mathcal{S} \in \mathbf{\Sigma}$ and $\mathcal{M}(\mathcal{S}) \leq 0$, then $\mathcal{M}(\mathcal{S})$ would be a **negative measure** and negative measure may be needed for some applications in spite of the added awkwardness of the proofs.

Lebesgue Measure Introduction:

If the set $\mathcal{S}$ is measurable, the $\mathcal{M}(\mathcal{S})$ is called the **total mass of the set**, e.g., if $\mathcal{S}$ is an interval $[a, b]$ then it is the length $(b - a)$, if a rectangle $[a, b] \times [c, d]$ then it is its area $(b - a) \cdot (d - c)$, or if a cube $[a, b] \times [a, b] \times [a, b]$ then it is its volume $(b - a)^3$. The closed intervals $[a, b]$, open intervals (a, b) and semi-open intervals $[a, b)$ or $(a, b]$, have the same measure or mass or length of $(b - a)$, since they differ only by **points of zero measure**.

In general, a **Lebesgue measure** is a measure on an n_x dimensional space of real vectors, so the universe is $\mathcal{U} = \mathbb{R}^{n_x}$, a representative set is a hypercube

$$\mathcal{S} = (\mathbf{a}, \mathbf{b}) \equiv (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_{n_x}, b_{n_x}),$$

such that $-\infty < a_i < b_i < +\infty$ and the measure has the form

$$\mathcal{M}(\mathcal{S}) = \prod_{i=1}^{n_x} (b_i - a_i).$$

Alternatively,

$$\mathcal{M}(\mathcal{S}) = \int_{\mathcal{S}} d\mathbf{x}.$$

Lebesgue measure is a special case of Borel measure specialized to $\mathbb{R}^{n_x}$.

Often, the infinitesimal hypercube measure between vector positions from $\mathbf{x}$ to $\mathbf{x} + d\mathbf{x}$ is abbreviated as

$$\mathcal{M}(\mathbf{dx}) = \mathcal{M}((\mathbf{x}, \mathbf{x} + d\mathbf{x})),$$

for compact notation, letting $\mathbf{dx}$ represent the vector-interval set $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$. This also recognizes the translation invariance of the measure of a generalized interval $(\mathbf{x}, \mathbf{x} + d\mathbf{x})$, since its generalized length $\prod_{i=1}^{n_x} dx_i$ is independent of the interval start at $\mathbf{x}$.

Dirac Measures:

Another measure that complements the Lebesgue measure is the **Dirac measure** δ_x, for some point in $\mathcal{U}$, having the properties that for some set $\mathcal{S} \subseteq \mathcal{U}$,

$$\delta_x(\mathcal{S}) = \left\{ \begin{array}{ll} 1, & x \in \mathcal{S} \\ 0, & x \notin \mathcal{S} \end{array} \right\}. \tag{12.2}$$

This is the set version of the **Dirac delta function** and apparently the same basic definition as the indicator function $\mathbf{1}_{x \in \mathcal{S}}$, except without the measure infrastructure.

Counting Measures:

For Poisson processes and other discrete applications, there are also counting measures, i.e., when

$$\mathcal{M}(\mathcal{S}) = N(\mathcal{S}) \equiv \textit{ number of elements in set } \mathcal{S}. \tag{12.3}$$

This includes the points of zero measure that **do count**.

Some Properties of Measures:

- The measure space $\{\mathcal{U}, \mathbf{\Sigma}, \mathcal{M}\}$ is **finite** if $\mathcal{M}(\mathcal{U}) < \infty$ and real.
- The measure space $\{\mathcal{U}, \mathbf{\Sigma}, \mathcal{M}\}$ is $\boldsymbol{\sigma}$**-finite** if there exists a countable sequence of measurable sets $\{\mathcal{S}_i \in \mathbf{\Sigma} : i = 1, 2, \ldots,\}$ such that $\mathcal{M}(\mathcal{S}_i) < \infty$ and real for all i, i.e., sets of finite measure, and

  $$\mathcal{U} = \bigcup_{i=1}^{\infty} \mathcal{S}_i,$$

  the union of a countable number of sets of finite measure. Note that σ-finite is not necessarily finite, since the set of real intervals $[i, i+1]$, have unit measure which is finite (a Lebesgue measure), but their union is the real line, $\mathcal{U} = \mathbb{R}^1$, which is infinite, so $\mathcal{U}$ is σ-finite while not finite.
- The measure $\mathcal{M}$ is a **montonic function** since if measurable sets $\mathcal{S}_1$ and $\mathcal{S}_2$ ordered $\mathcal{S}_1 \subseteq \mathcal{S}_2$ then $\mathcal{M}(\mathcal{S}_1) \leq \mathcal{M}(\mathcal{S}_2)$.
- If $\{\mathcal{S}_i \in \mathbf{\Sigma} : i = 1, 2, \ldots,\}$ is any countable sequence of subsets that are **not necessarily disjoint**, then the measure of the union is only bounded by the sum of the measures,

  $$\mathcal{M}\left(\bigcup_{i=1}^{\infty} \mathcal{S}_i \right) \leq \sum_{i=1}^{\infty} \mathcal{M}(\mathcal{S}_i),$$

  unlike the lack of redundancies of disjoint sets given in (12.1).

- If $\{\mathcal{S}_i \in \mathbf{\Sigma} : i = 1, 2, \ldots,\}$ is any countable sequence of subsets that are **forward nested** so that $\mathcal{S}_i \subseteq \mathcal{S}_{i+1}$, then the limit of the **measure of the union** has the limiting measure,

$$\mathcal{M}\left(\bigcup_{i=1}^{\infty} \mathcal{S}_i\right) = \lim_{n\to\infty} \mathcal{M}(\mathcal{S}_n),$$

  noting that $\mathcal{M}(\mathcal{S}_i \cup \mathcal{S}_{i+1}) = \mathcal{M}(\mathcal{S}_{i+1})$.

- If $\{\mathcal{S}_i \in \mathbf{\Sigma} : i = 1, 2, \ldots,\}$ is any countable sequence of subsets that are **backward nested** so that $\mathcal{S}_{i+1} \subseteq \mathcal{S}_i$, then the limit of the **measure of the intersection** has the limiting measure,

$$\mathcal{M}\left(\bigcap_{i=1}^{\infty} \mathcal{S}_i\right) = \lim_{n\to\infty} \mathcal{M}(\mathcal{S}_n),$$

  noting that $\mathcal{M}(\mathcal{S}_i \cap \mathcal{S}_{i+1}) = \mathcal{M}(\mathcal{S}_{i+1})$.

- A **null set** $\mathcal{N} \in \mathbf{\Sigma}$ is a measurable set such that $\mathcal{M}(\mathcal{N}) = 0$, a **negligible set** is a subset of a null set and a measure $\mathcal{M}$ is **complete** if every negligible set is measurable. A σ-algebra $\mathbf{\Sigma}$ can always be completed by adding any missing null sets to it.

- A property P holds **almost everywhere (a. e.)** if the set of elements $\mathcal{S}$ in $\mathbf{\Sigma}$ for which the property does not hold is a null set, i.e., $\mathcal{S} = \mathcal{N}$ is a set with measure zero such that $\mathcal{M}(\mathcal{N}) = 0$.

- Given the measure space $\{\mathcal{U}, \mathbf{\Sigma}, \mathcal{M}_1\}$, another measure $\mathcal{M}_2$ on the measurable space $\{\mathcal{U}, \mathbf{\Sigma}\}$ is **absolutely continuous** with respect to $\mathcal{M}_1$ if for any measurable set $\mathcal{S} \in \mathbf{\Sigma}$

$$\mathcal{M}_1(\mathcal{S}) = 0 \Longrightarrow \mathcal{M}_2(\mathcal{S}) = 0,$$

  Absolute continuity is written symbolically as $\mathcal{M}_2(\mathcal{S}) \prec \mathcal{M}_1(\mathcal{S})$ (or as $\mathcal{M}_2(\mathcal{S}) \ll \mathcal{M}_1(\mathcal{S})$, but this conflicts with asymptotic notation). This property permits defining the ratio $\mathcal{M}_2(\mathcal{S})/\mathcal{M}_1(\mathcal{S})$ for comparison between two measures of a set.

  If $\mathcal{M}_2(\mathcal{S}) \prec \mathcal{M}_1(\mathcal{S})$ and $\mathcal{M}_1(\mathcal{S}) \prec \mathcal{M}_2(\mathcal{S})$, i.e., both are mutually absolutely continuous with respect to the other, then the measures $\mathcal{M}_1$ and $\mathcal{M}_2$ are said to be **equivalent** ($\mathcal{M}_1(\mathcal{S}) \overset{\text{a.c.}}{\equiv} \mathcal{M}_1(\mathcal{S})$). As Cont and Tankov [59] suggest, the term **equivalence** is perhaps misleading and should be called something like **comparable**.

Many of these properties are needed for proofs of existence and convergence, as well as for constructing stochastic processes.

Measurable Functions:

A prerequisite that a function f is integrable is that f is a **measurable function**.

Definition 12.3. *Given two measurable spaces, $(\mathcal{U}_1, \mathbf{\Sigma}_1)$ and $(\mathcal{U}_2, \mathbf{\Sigma}_2)$, a mapping of the **function** f from $\mathcal{U}_1$ to $\mathcal{U}_2$ is **measurable** with respect to $(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2)$ if the **inverse (preimage)** $f^{-1}(\mathcal{S}_2) \in \mathbf{\Sigma}_1$ for all $\mathcal{S}_2 \in \mathbf{\Sigma}_2$, i.e., there is a $\mathcal{S}_1 \in \mathbf{\Sigma}_1$ such that $f(\mathcal{S}_1) = \mathcal{S}_2$.*

Just as in Riemann integration for general Riemann integrable functions, the integral is built up from the limit of finite Riemann sums, the integral with respect to a measurable function is built-up from sums of step functions called a **simple function**.

Definition 12.4.

- *A **simple function** is a finite linear combination of set indicator functions $\{\mathbf{1}_{x\in\mathcal{S}_i}\}$ of measurable sets $\mathcal{S}_i$ for $i = 1\!:\!n$ on a measurable space $(\mathcal{U}, \mathbf{\Sigma})$, with real coefficients (could also be complex) c_i, having the form*

$$f(x) = \sum_{i=1}^{n} c_i \mathbf{1}_{x\in\mathcal{S}_i},$$

where $x \in \mathcal{U}$.

- *The **integral with respect to the measure $\mathcal{M}$** for such a simple function is*

$$\mathcal{I}_{\mathcal{M}}[f] = \sum_{i=1}^{n} c_i \mathcal{M}(\mathcal{S}_i),$$

provided all the measures $\mathcal{M}(\mathcal{S}_i)$ are finite, i.e., providing the analogy to the Riemann sums.

- *For a general, positive measurable function f, **integrability** can be extended to f by **comparison to simple measurable functions** on $\mathcal{U}$, such as*

$$\mathcal{I}_{\mathcal{M}}[f] = \sup_{g} \left\{ \mathcal{I}_{\mathcal{M}}[g] :\ g(x) = \sum_{i=1}^{n} c_i \mathbf{1}_{x\in\mathcal{S}_i},\ g(x) \le f(x),\ x \in \mathcal{U} \right\},$$

*provided $\mathcal{I}_{\mathcal{M}}[f]$ is finite. For functions that change sign, i.e., signed functions, the **positive-negative decomposition** $f(x) = f_+(x) - f_-(x)$ with the $f_\pm(x) \equiv (|f|(x) \pm f(x))/2$ for $x \in \mathcal{U}$, such that*

$$\mathcal{I}_{\mathcal{M}}[f] = \mathcal{I}_{\mathcal{M}}[f_+] - \mathcal{I}_{\mathcal{M}}[f_-],$$

provided the $\mathcal{I}_{\mathcal{M}}[f_\pm]$ are finite. (The positive-negative decomposition is used in Chapt. 8 for numerical up-winding to ensure stability.)

- *If $\mathcal{M}$ is a Lebesgue measure, then the Lebesgue of the measure function f on $\mathcal{S} \in \mathcal{U}$ can be written,*

$$\mathcal{I}_{\mathcal{M}}[f] = \int_{\mathcal{S}} f(x)\mathcal{M}(dx) = \int_{\mathcal{U}} \mathbf{1}_{x\in\mathcal{S}}\ f(x)\mathcal{M}(dx),$$

where recall dx symbolizes the set $(x, x+dx)$.

- ***Monotone Convergence Theorem:***
Given the measure space $(\mathcal{U}, \mathbf{\Sigma}, \mathcal{M})$, if $\{f_n(x),\ f_n(x) \geq 0 \text{ for } n = 1, 2, \ldots\}$ is a countable sequence of 1-dimensional (non-negative) measurable functions on $\mathcal{U}$ that is a. e. monotone increasing and converging pointwise to $f(x)$ a. e., then

$$\lim_{n\to\infty} \int_{\mathcal{U}} f_n(x)\mathcal{M}(dx) = \int_{\mathcal{U}} f_n(x)\mathcal{M}(dx).$$

This basic convergence theorem leads to several others.

12.1.2 Change of Measure: Radon-Nikodým Theorem and Derivative:

The abstract analog of the change of variables, chain rule and Jacobian techniques for Riemann or Riemann-Stieltjes integral is the change of measures and the Radon-Nikodým derivative.

Theorem 12.5. ***Radon-Nikodým Change of Measures:***
Given the measure space $\{\mathcal{U}, \mathbf{\Sigma}, \mathcal{M}_1\}$ with σ-finite measure $\mathcal{M}_1$, if $\mathcal{M}_2$ is a finite measure that is absolutely continuous with respect to $\mathcal{M}_1$ ($\mathcal{M}_2 \prec \mathcal{M}_1$) then there exists a measurable real function $\mathbb{D}(x) > 0$ for $x \in \mathcal{U}$ such that for each measurable set $\mathcal{S} \in \mathbf{\Sigma}$

$$\mathcal{M}_2(\mathcal{S}) = \mathcal{I}_{\mathcal{M}_1}[\mathbb{D}\,\mathbf{1}_{*\in\mathcal{S}}] = \int_{\mathcal{U}} \mathbb{D}(x)\mathbf{1}_{x\in\mathcal{S}} d\mathcal{M}_1(x) = \int_{\mathcal{S}} \mathbb{D}(x) d\mathcal{M}_1(x), \quad (12.4)$$

where $d\mathcal{M}_i(x) = \mathcal{M}_i(dx)$ is equivalent notation for $i = 1{:}2$. The function $\mathbb{D}$ is the ***Radon-Nikodým derivative*** *of $\mathcal{M}_2$ with respect to $\mathcal{M}_1$, i.e.,*

$$\mathbb{D}(x) = \frac{d\mathcal{M}_2}{d\mathcal{M}_1}(x) \quad \text{or} \quad \mathbb{D}(\mathcal{S}) = \frac{d\mathcal{M}_2}{d\mathcal{M}_1}(\mathcal{S}). \quad (12.5)$$

Further, if η is integrable with respect to the measure $\mathcal{M}_2$, then

$$\mathcal{I}_{\mathcal{M}_2}[\eta] = \int_{\mathcal{U}} \eta(x) d\mathcal{M}_2(x) = \int_{\mathcal{U}} \eta(x) \frac{d\mathcal{M}_2(x)}{d\mathcal{M}_1(x)} d\mathcal{M}_1(x)$$

$$= \mathcal{I}_{\mathcal{M}_2}[\eta\,\mathbb{D}] = \int_{\mathcal{U}} \eta(x)\mathbb{D}(x)\mathcal{M}_1(x),$$

i.e., using the Radon-Nikodým derivative in a measure-theoretic chain rule.

Thus, the Radon-Nikodým derivative is the analog of the Jacobian of the transformation (9.56) in an integral change of variables and leads to the absolutely continuous measure chain rule, symbolically substituting for $\mathbb{D}$,

$$d\mathcal{M}_2 = \frac{d\mathcal{M}_2}{d\mathcal{M}_1} d\mathcal{M}_1.$$

If $d\mathcal{M}_2$ and $d\mathcal{M}_1$ are mutually absolutely continuous, i.e., equivalent ($\mathcal{M}_1(\mathcal{S}) \overset{\text{a.c.}}{\equiv} \mathcal{M}_1(\mathcal{S})$), the Radon-Nikodým derivatives are mutual reciprocals,

$$\frac{d\mathcal{M}_1}{d\mathcal{M}_2} = 1 \Big/ \frac{d\mathcal{M}_2}{d\mathcal{M}_1} \, ,$$

formally justified by common null sets.

See the probability measure Examples 12.13 illustrations of applied-oriented calculations for Radon-Nikodým derivatives in Subsect. 12.2.1.

12.1.3 Probability Measure Basics

Since the probability distribution function for the real random variable $\mathbf{X}$ on the real set $\mathcal{S} \subseteq \mathbb{R}^{n_x}$ has the property that

$$\Phi_{\mathbf{X}}(\mathcal{S}) = \text{Prob}[\mathbf{X} \in \mathcal{S}] \in [0, 1],$$

it is a natural candidate for a measure and the density $\phi_{\mathbf{X}}(\mathbf{x})$ could play the role of the Radon-Nikodým derivative. According to convention, we reset the universe as $\mathcal{U} = \Omega$, the σ-algebra as $\mathbf{\Sigma} = \mathcal{F}$ and the measure as $\mathcal{M} = \mathbb{P}$. For the jump part of jump-diffusions, counting or jump measures will also be needed.

Definition 12.6. ***Probability Measure:***
A ***probability space*** $(\Omega, \mathcal{F}, \mathbb{P})$ *is a measure space with elements* $\omega \in \Omega$ *called* ***sample points*** *of random events in the* ***sample space*** Ω*, elements* $\mathcal{F}_i \in \mathcal{F}$ *called random* ***events*** *and any* ***probability measure*** $\mathbb{P}$ *on the measurable space* $(\Omega, \mathcal{F})$ *has total mass of one, i.e.,* $\mathbb{P}(\Omega) = 1$.

Summarizing the Kolmogorov axioms [33] of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$*:*

- $\mathbb{P}(\emptyset) = 0$ *and* $\mathbb{P}(\Omega) = 1$.
- $\mathbb{P}(\mathcal{S}) \geq 0$ *for all* $\mathcal{S} \in \Omega$.
- $\mathbb{P}(\cup_{i=1}^{\infty} \mathcal{S}_i) = \cup_{i=1}^{\infty} \mathbb{P}(\mathcal{S}_i)$*, assuming the* $\{\mathcal{S}_i\}$ *are disjoint and countable, i.e., there is* ***countable additivity****, so that if* $\mathcal{S} \cup \mathcal{S}^c = \Omega$*, then the complementarity property also holds,* $\mathbb{P}(\mathcal{S}^c) = \mathbb{P}(\Omega) + \mathbb{P}(\mathcal{S})$.
- *If* $\mathcal{S}_2 \subseteq \mathcal{S}_2$ *and* $\mathbb{P}(\mathcal{S}_1) = 0$*, then* $\mathbb{P}(\mathcal{S}_2) = 0$*, i.e., the probability space is* ***complete****.*

Some additional properties and nomenclature:

- *The $\omega \in \Omega$ are also called scenarios as well as **outcomes**, the **underlying** or **background random variables**, e.g., like the mark variable of the compound Poisson process or Poisson random measure.*

- *An event set $\mathcal{S}$ with probability $\mathbb{P}(\mathcal{S}) = 1$ is said to happen **almost surely** (**a.s.**) or **with probability one** (**w.p.o.**), equivalent to **almost everywhere** (**a.e.**) for mathematical measures. If an event $\mathcal{S}$ has probability $\mathbb{P}(\mathcal{S}) = 0$, the event is said to be **impossible**.*

- *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then a (real) **random variable** $\mathbf{X}(\omega)$ is a measurable mapping from Ω to $\mathbb{R}^{n_x}$ such that the inverse (preimage) $\mathbf{X}^{-1}(\mathcal{S}) = \{\omega \in \Omega : \mathbf{X}(\omega) \in \mathcal{S}\}$ is $\mathcal{F}$-measurable for Borel (open) sets $\mathcal{S} \in \mathcal{B}(\mathbb{R}^{n_x})$, i.e., $\mathbf{X}(\omega)$ is the **realization** $\mathbf{X}$ upon event ω. If f is a (real) measurable function, then $f(\mathbf{X}(\omega))$ will also be a random variable.*

- *If the problem involves only a single probability measure $\mathbb{P}$ for the single random variable ω, then we can write in more usual notation,*

$$X \equiv \omega, \quad \text{Prob}[X \in \mathcal{S}] = \Pr[X \in \mathcal{S}] \equiv \mathbb{P}(\mathcal{S}),$$

  i.e., the probability measure is the distribution $\Phi_\omega(\mathcal{S}) = \mathbb{P}(\mathcal{S})$ for $\mathcal{S} \subseteq \Omega$.

- *In general, if $\mathbf{X} = \mathbf{X}(\omega) \in \mathbb{R}^{n_x}$ for $\omega \in \Omega$, then let $\omega \in \mathcal{S}_\omega \subseteq \Omega$, $\mathbf{X}(\omega) \in \mathcal{S}_\mathbf{X} = \mathbf{X}(\mathcal{S}_\omega)$ and assuming the preimage $\mathcal{S}_\omega = \mathbf{X}^{-1}(\mathcal{S}_\mathbf{X})$ exists, then the **distribution** of $\mathbf{X}$ is the probability measure*

$$\Phi_\mathbf{X}(\mathcal{S}_\mathbf{X}) = \mathbb{P}(\mathbf{X}^{-1}(\mathcal{S}_\mathbf{X})),$$

  so $\Phi_\mathbf{X}(\mathbf{x}) = \mathbb{P}(\{\omega \ni \mathbf{X} \leq \mathbf{x}\})$, the inequality $(\mathbf{X} \leq \mathbf{x})$ meant element-wise.

- *The **expectation** for a measurable real function f of $X \in \mathbb{R}^{n_x}$ with $\omega \in \Omega$ is then*

$$\mathrm{E}[f(\mathbf{X})] = \int_\Omega f(\mathbf{X}(\omega))\mathbb{P}(d\omega) = \int_\Omega f(\mathbf{X}(\omega))d\mathbb{P}(\omega) = \int_{\mathbb{R}^{n_x}} f(\mathbf{x})\Phi_\mathbf{X}(d\mathbf{x}),$$

  provided f is absolutely integrable,

$$\int_\Omega |f(\mathbf{X}(\omega))|\mathbb{P}(d\omega) < \infty,$$

  noting that the $d\omega$ argument of $\mathbb{P}$ is an abbreviation for the interval $(\omega, \omega + d\omega)$ and that $d\mathbb{P}(\omega)$ and $\mathbb{P}(d\omega)$ will be assumed to be equivalent notation.

- ***Almost Sure Equivalence:** Let $\mathbf{X}_1(\omega)$ and $\mathbf{X}_2(\omega)$ be two random variables for $\omega \in \Omega$, then $\mathbf{X_1} \overset{\text{a.s.}}{=} \mathbf{X_2}$ if*

$$\mathbb{P}(\{\omega \in \Omega, \mathbf{X}_1(\omega) = \mathbf{X}_2(\omega)\}) = 1.$$

- ***Equivalence in Distribution:*** *Let* $\mathbf{X}_1(\omega)$ *and* $\mathbf{X}_2(\omega)$ *be two random variables for* $\omega \in \Omega$. *If the distribution satisfy*

$$\Phi_{\mathbf{X}_1} = \Phi_{\mathbf{X}_2},$$

then $\mathbf{X}_1(\omega)$ *and* $\mathbf{X}_2(\omega)$ *are called* ***equal in distribution*** *and we write*

$$\mathbf{X_1} \stackrel{\text{dist}}{=} \mathbf{X_2}.$$

(*Also called* ***equal in law*** *or* ***identically distributed****; the notation* $\mathbf{X_1} \stackrel{\text{d}}{=} \mathbf{X_2}$ *is also used.*)

- *The set of* n *random variables* $\{X_i\}$ *are* ***independent*** *with respect to the measurable sets* $\mathcal{S}_i$ *for* $i = 1{:}n$ *if the probability of the union is the product of the probabilities,*

$$\mathbb{P}\left(\bigcup_{i=1}^{n}\{X_i \in \mathcal{S}_i\}\right) = \prod_{i=1}^{n}\mathbb{P}(\{X_i \in \mathcal{S}_i\}),$$

where the underlying random variable ω *has been suppressed. A more concrete and useful form as distribution in the vector* $\mathbf{X} = [X_i]_{n\times 1}$ *is*

$$\Phi_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}\left(\bigcup_{i=1}^{n}\{X_i \le x_i\}\right) = \prod_{i=1}^{n}\mathbb{P}(\{X_i \le x_i\}) = \prod_{i=1}^{n}\Phi_{X_i}(x_i).$$

An immediate corollary is the multiplication rule for the expectation of a set of independent random variables,

$$\mathrm{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n}\mathrm{E}[X_i],$$

assuming finite expectations, $\mathrm{E}[|X_i|] < \infty$ *for* $i = 1{:}n$.

For more background information, see Applebaum [12], Billingsley [32], Bingham and Kiesel [33], Cont and Tankov [59]. Cyganowski, Kloeden and Ombach [66], Øksendal [222] and Øksendal and Sulem [223].

Much of the further results, such as conditional expectations, follow the applied path in this book, except that matters like that of positivity and changes in sign have to be treated with care to account for particular abstract constructs and conditions that are designed to facilitate proofs rather than the wide variety of problem applications.

12.1.4 Stochastic Processes in Continuous Time on Filtered Probability Spaces

Since the emphasis of this book is on jump-diffusions, stochastic processes in continuous time are treated and the relatively simpler, but not simple, discrete time

stochastic processes are omitted (see Pliska's [225] book or Bingham and Kiesel's [33, Chapt. 3] chapter devoted to discrete time processes). The main additional difficulty treating stochastic processes in continuous time is extending the notion of a single probability space to a family of probability spaces over the continuous time variable t which often has infinite range.

Definition 12.7. ***Filtered Probability Space:***

- *Based upon a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, a* ***filtration*** *is a family of increasing* σ*-algebras*

$$\mathbb{F} = \{\mathcal{F}_t : \ t \geq 0; \ \mathcal{F}_s \subseteq \mathcal{F}_t, \ 0 \leq s \leq t < \infty\}$$

  and the extended space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ *is called a* ***filtered probability space****. The sub-*σ*-algebra* $\mathcal{F}_t$ *represents the known information of the system on* $(0, t]$ *at time* t*.*

- *The usual filtration conditions (with jump-diffusions in mind) are*
  - *The initial sub-*σ*-algebra* $\mathcal{F}_0$ *the* $\mathbb{P}$*-null-sets of* $\mathcal{F}$*.*
  - *The filtration* $\mathbb{F}$ *is* ***right-continuous with left limits (RCLL or cádlág in French)****, i.e.,* $\mathcal{F}_t = \mathcal{F}_{t^+} = \lim_{\epsilon \to 0^+} \mathcal{F}_{t+\epsilon}$ *for the RC part and* $\mathcal{F}_{t^-} = \lim_{\epsilon \to 0^+} \mathcal{F}_{t-\epsilon}$ *for the LL part exists. The* ***jump*** *in the sub-*σ*-algebra at time* t *is* $[\mathcal{F}]_t = \mathcal{F}_{t^+} - \mathcal{F}_{t^-}$*. If only continuous processes such as diffusions are under consideration, then* ***right continuity (RC)*** *is sufficient.*

Definition 12.8. ***Stochastic Process:***

- *Given the filtered probability space* $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$*, a* ***stochastic process in continuous time*** $\mathbf{X} = \{\mathbf{X}(t) : \ t \geq 0\}$ *and* X *is* $\mathcal{F}_t$***-adapted*** *to the filtration* $\mathbb{F}$ *if* $\mathbf{X}(t)$ *is* $\mathcal{F}_t$*-measurable* $(\mathbf{X}(t) \in \mathcal{F}_t)$ *for each* t*.*

- *The* ***natural filtration*** *for the stochastic process* $\mathbf{X}(t)$ *can be written as*

$$\mathcal{F}_{t,\mathbf{X}} = \widehat{\sigma}(\mathbf{X}(s), \ 0 \leq s \leq t),$$

  with $\widehat{\sigma}$ *signifying the* σ*-field of* $\mathbf{X}(t)$*, or more loosely the information or history of the process* $\mathbf{X}(t)$ *up until time* t*.*

- *Including the dependence on the underlying random variable,* $\omega \in \Omega$*, the* $\mathbf{X}(t; \omega)$ *defines a random function of time, called the* ***sample path*** *and is a mapping from* $[0, t] \times \Omega$ *to* $\mathbb{R}^{n_x}$*. Usually,* $\mathbf{X}(t; \omega)$ *is denoted by* $\mathbf{X}_t(\omega)$ *or just* X_t*, however in this book real subscripts are reserved to denote partial derivatives, except for algebraic quantities like* $\mathcal{F}_t$ *that are not genuine functions.*

- *If* X *is adapted, i.e.,* $\mathcal{F}_t$***-adapted*** *to* $\mathbb{F}$*, for* $t \geq 0$*, then the conditional expectation satisfies*

$$\mathrm{E}[X(t) \mid \mathcal{F}_t] \stackrel{\text{a.s.}}{=} X(t),$$

*since $X(t)$ is known from $\mathcal{F}_t$ (recall the symbol $\overset{\text{a.s.}}{=}$ denotes **equals almost surely**). Saying that X or $X(t)$ is $\mathcal{F}_t$-**adapted** to $\mathbb{F}$ means the same as saying that $X(t)$ is **nonanticipating**.*

- *Two stochastic processes X_1 and X_2 are the **same with respect to a set of finite-dimensional distributions** if for some positive integer n and discrete time points $\{t_i : \ i = 1:n\}$, the random vectors $\mathbf{X}_j = [X_{i,j}]_{n\times 1}$ for $j = 1:2$ have the same n-dimensional distribution, corresponding to the stochastic processes X_j for $j = 1:2$, respectively.*

12.1.5 Martingales in Continuous Time

Martingales are processes with the property that the best predictor of the process future value is the present value given present knowledge, i.e., it represents a fair game of gambling, rather than a favorable or unfavorable one.

Definition 12.9. ***MartingaleProperties in Continuous Time:***

- *Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ and $\mathcal{F}_t$-adapted process $X(t)$ on $[0,T]$, $T < \infty$, then $X(t)$ is a **martingale** if*

$$\mathrm{E}[X(t) \mid \mathcal{F}_s] \overset{\text{a.s.}}{=} X(s), \quad t > s \geq 0, \tag{12.6}$$

provided $X(t)$ is absolutely integrable, $\mathrm{E}[|X(t)|] < \infty$ on $[0,T]$, i.e., the best predictor of $X(t)$ with respect to the filter $\mathcal{F}_s$ is $X(s)$.

- *If instead of (12.6),*

$$\mathrm{E}[X(t) \mid \mathcal{F}_s] \overset{\text{a.s.}}{\leq} X(s), \quad t > s \geq 0,$$

*then $X(t)$ is a **supermartingale**,*
but if

$$\mathrm{E}[X(t) \mid \mathcal{F}_s] \overset{\text{a.s.}}{\geq} X(s), \quad t > s \geq 0,$$

*then $X(t)$ is a **submartingale**. (The submartingale corresponds to the favorable game and the supermartingale corresponds to the unfavorable game, provided $X(t) - X(s)$ represents the gain.)*

- *Two martingales $\mathcal{M}_1(t)$ and $\mathcal{M}_2(t)$ which are also equivalent or mutually absolutely continuous measures, i.e., $\mathcal{M}_1(t) \overset{\text{a.c.}}{\equiv} \mathcal{M}_2(t)$, are called **equivalent martingale measures (EMM)** and they play an important role in mathematical finance.*

Examples 12.10. ***Diffusion, Jump and other Martingales:***
For this set of examples, the time interval $[0,T]$ as well as the coefficients will be finite, so there is no question that the stochastic processes will be absolutely integrable.

1. *Let $X(t)$ be a $\{\mu_0, \sigma_0\}$-constant coefficient,* ***diffusion process*** *with SDE,*

$$dX(t) = \mu_0 dt + \sigma_0 dW(t),$$

and it is of interest to know for what values of μ_0 is $X(t)$ a martingale, a supermartingale or a submartingale.

The solution by integrating over $[s,t]$ is

$$X(t) = X(s) + \mu_0(t-s) + \sigma_0(W(t) - W(s)),$$

noting that $(W(t)-W(s)) \stackrel{\text{dist}}{=} W(t-s)$ by stationary property and is independent of $W(s)$ and $\mathrm{E}[W(t-s)|W(s)] = 0$, so with $\mathcal{F}_t = \widehat{\sigma}(X(r),\ 0 \le r \le t)$, the natural filtration for $X(t)$,

$$\mathrm{E}[X(t) \mid \mathcal{F}_s] = X(s) + \mu_0(t-s),$$

$0 \le s < t$. Hence, $X(t)$ is a martingale if $\mu_0 = 0$ (the case of the zero-mean infinitesimal diffusion, denoted by $d\widehat{X}(t) = \sigma_0 dW(t)$), a supermartingale if $\mu_0 < 0$ or a submartingale if $\mu_0 > 0$. Alternatively, the translated process

$$\widetilde{X}(t) \equiv X(t) - \mu_0 t \stackrel{\text{dist}}{=} \widetilde{X}(s) + \sigma_0 W(t-s)$$

is a martingale.

2. *Let $X(t)$ be a $\{\mu_0, \sigma_0\}$-constant coefficient,* ***geometric diffusion process****,*

$$dX(t) = X(t)(\mu_0 dt + \sigma_0 dW(t)),$$

which has the Itô calculus solution,

$$X(t) \stackrel{\text{dist}}{=} X(s)\exp((\mu_0 - \sigma_0^2/2)(t-s) + \sigma_0 W(t-s)),$$

so

$$\mathrm{E}[X(t) \mid \mathcal{F}_s] = X(s)\exp(\mu_0(t-s)),$$

$0 \le s < t$. Again, $X(t)$ is a martingale if $\mu_0 = 0$, a supermartingale if $\mu_0 < 0$ or a submartingale if $\mu_0 > 0$. Alternatively, the scaled process

$$\widetilde{X}(t) \equiv \exp(-\mu_0 t)X(t) \stackrel{\text{dist}}{=} \widetilde{X}(s)\exp(\sigma_0(W(t-s) - \sigma_0(t-s)/2)$$

is martingale, or more specifically an ***exponential martingale*** *[22] and the scaling corresponds to the Girsanov transformation of $W(t)$ that will be discussed in Subsect. 12.2.2.*

3. *Let $X(t)$ be a* ***simple Poisson process*** *$P(t)$ with additional drift and constant coefficients, $\{\mu_0, \nu_0, \lambda_0\}$,*

$$dX(t) = \mu_0 dt + \nu_0 dP(t),$$

where $\mathrm{E}[dP(t)] = \lambda_0 dt = \mathrm{Var}[dP(t)]$. *The solution is*

$$X(t) = X(s) + \mu_0(t-s) + \nu_0(P(t) - P(s)),$$

where $(P(t) - P(s)) \stackrel{\text{dist}}{=} P(t-s)$ *and the conditional expectation is*

$$\mathrm{E}[X(t) \mid \mathcal{F}_s] = X(s) + (\mu_0 + \lambda_0\nu_0)(t-s),$$

so $X(t)$ *is a martingale if* $\mu_0 = -\lambda_0\nu_0$ *(the zero-mean infinitesimal jump process, denoted by* $d\widehat{X}(t) = \nu_0 d\widehat{P}(t)$*, using the zero mean Poisson* $d\widehat{P}(t) \equiv dP(t) - \lambda_0 dt$*), a supermartingale if* $\mu_0 < -\lambda_0\nu_0$ *or a submartingale if* $\mu_0 > -\lambda_0\nu_0$. *Alternatively, the translated process*

$$\widetilde{X}(t) \equiv X(t) - (\mu_0 + \lambda_0\nu_0)t \stackrel{\text{dist}}{=} \widetilde{X}(s) + \nu_0\widehat{P}(t-s)$$

is a martingale.

4. *Let* $X(t)$ *be a* ***compound Poisson process*** *with additional drift and constant coefficients,* $\{\mu_0, \nu_0, \lambda_0, \mu_Q\}$,

$$dX(t) = \mu_0 dt + \nu_0 \sum_{i=1}^{dP(t)} Q_i,$$

where $\mathrm{E}[dP(t)] = \lambda_0 dt = \mathrm{Var}[dP(t)]$ *and the* Q_i *are IID random marks with mean* μ_Q *and variance* σ_Q^2 *which will not be needed (also note that the zero-one law has not been applied to* $dP(t)$*). The solution is*

$$X(t) \stackrel{\text{dist}}{=} X(s) + \mu_0(t-s) + \nu_0 \sum_{i=1}^{P(t-s)} Q_i$$

and the conditional expectation, by iterated conditional expectations between the Poisson counting process and the marks, is

$$\mathrm{E}[X(t) \mid \mathcal{F}_s] = X(s) + (\mu_0 + \lambda_0\nu_0\mu_Q)(t-s),$$

so $X(t)$ *is a martingale if* $\mu_0 = -\lambda_0\nu_0\mu_Q$ *(the zero-mean infinitesimal compound Poisson,* $d\widehat{X}(t) = \nu_0\mu_Q d\widehat{P}(t) + \nu_0 \sum_{i=1}^{dP(t)} \widehat{Q}_i$ *where* $\widehat{P}(t) \equiv (P(t) - \lambda_0 t)$ *and* $\widehat{Q}_i \equiv (Q_i - \mu_Q)$*), a supermartingale if* $\mu_0 < -\lambda_0\nu_0\mu_Q$ *or a submartingale if* $\mu_0 > -\lambda_0\nu_0\mu_Q$. *The alternative process*

$$\widetilde{X}(t) = X(t) - (\mu_0 + \lambda_0\nu_0\mu_Q)t = \widetilde{X}(s) + \nu_0\mu_Q\widehat{P}(t-s) + \nu_0 \sum_{i=1}^{P(t-s)} \widehat{Q}_i$$

is a martingale, such that the difference $\widetilde{X}(t) - \widetilde{X}(s)$ *is a linear combination of zero-mean random processes, or variables as in the case of* $\widehat{Q}_i$*, counting only the jumps in* $(s, t]$.

5. *As an exercise the reader can find the similar martingale properties as a function of the additional drift for the* ***geometric jump diffusion*** *problem with constant coefficients,*

$$dX(t) = X(t)\left(\mu_0 dt + \sigma_0 dW(t) + \nu_0 \sum_{i=1}^{dP(t)} (\exp(Q_i) - 1)\right),$$

where again the marks are IID with mean μ_Q and variance σ_Q^2, with the amplitude in the log-ready exponential form.

6. *The simplest, but trivial, example is the* ***constant process*** *$X(t) = c_0$ for $t \geq 0$, i.e., $dX(t) = 0$, so $X(t)$ is a martingale since $\mathrm{E}[X(t)|\mathcal{F}_s] = c_0 = X(s)$ for $s < t$.*

7. *Another example is the* ***closed martingale*** *that is constructed from an absolutely integrable random variable Y, independent of t, on the filtered probability space, such that a stochastic process is defined as*

$$X(t) \equiv \mathrm{E}[Y \mid \mathcal{F}_t], \quad t \geq 0.$$

Thus, by the ***tower law*** *([22, p. 34], [209, Rule 6, p. 72]),*

$$\mathrm{E}[X(t) \mid \mathcal{F}_s] = \mathrm{E}[\mathrm{E}[Y \mid \mathcal{F}_t] \mid \mathcal{F}_s] = \mathrm{E}[Y \mid \mathcal{F}_s] \equiv X(s),$$

for $s < t$, since the conditioning on $\mathcal{F}_t$ followed by the conditioning on $\mathcal{F}_s$ is the same as the original conditioning on $\mathcal{F}_s$, i.e., dependence is on the smaller of the conditioning filters.

12.1.6 Jump-Diffusion Martingale Representation:

For hedging in financial applications, martingale representations are heavily relied upon. There are many versions of martingale representation in the literature. Some have useful and elementary presentations. Many are restricted to diffusions except for a mention of jump processes. A selected sample is given by the references: Baxter and Rennie [22], Duffie [74], Glasserman [96], Øksendal [222] and Steele [256]. Here, a form of the martingale repesentation theorem is given for marked-jump-diffusion processes following Applebaum [12] and, particularly, Runggaldier [239]. Their formulation, after Jacod and Shiryaev [152], and Kunita and Watanabe [172], respectively, uses Poisson random measure $\mathcal{P}(\mathbf{dt}, \mathbf{dq})$ defined beginning in (5.1) on page 132 and whose integrals are related to compound Poisson processes (5.6) on mark-sample-space $\mathcal{Q}$ by

$$\int_{\mathcal{Q}} h(t,q)\mathcal{P}(\mathbf{dt}, \mathbf{dq}) = \sum_{i=1}^{dP(t)} h(T_i^-, Q_i), \tag{12.7}$$

without using the zero-one law for $dP(t)$, where the T_i^- are the pre-jump-times and the Q_i are the IID sampled marks, but often found in martingale form by using the centered or mean-zero Poisson random measure,

$$\widetilde{\mathcal{P}}(\mathbf{dt},\mathbf{dq}) \equiv \mathcal{P}(\mathbf{dt},\mathbf{dq}) - \mathrm{E}[\mathcal{P}(\mathbf{dt},\mathbf{dq})] = \mathcal{P}(\mathbf{dt},\mathbf{dq}) - \phi_Q(q;t)dq\lambda(t)dt,$$

where $\Phi_Q(dq;t) = \phi_Q(q;t)dq$ is jump-amplitude probability measure and $\lambda(t)$ is the Poisson jump-rate. The mean-zero relationship corresponding to the original relationship (12.7) is then

$$\int_{\mathcal{Q}} h(t,q)\widetilde{\mathcal{P}}(\mathbf{dt},\mathbf{dq}) = \sum_{i=1}^{dP(t)} h(T_i^-,Q_i) - \mathrm{E}_Q[h(t,Q)]\lambda(t)dt, \tag{12.8}$$

where $\mathrm{E}_Q[h(t,Q)] = \int_{\mathcal{Q}} h(t,q)\phi_Q(q;t)dq$.

Theorem 12.11. ***Marked-Jump-Diffusion Martingale Representation Theorem:***
Given the Wiener process $W(t)$ and compound Poisson triplet

$$\{dP(t), \lambda(t), \phi_Q(q;t)\}$$

or else a Poisson random measure $\mathcal{P}(\mathbf{dt},\mathbf{dq})$ on the sigma-field

$$\mathbb{F} = \mathcal{F}_t^{(W,P,Q)} = \widehat{\sigma}\{W(s), P(t), \mathcal{S}_Q,\ \mathcal{S}_N : 0 \le s \le t,\ \mathcal{S}_Q \in \mathcal{Q},\ \mathcal{S}_N \in \mathcal{N}_1\},$$

$\mathcal{N}_1$ is the collection of null-sets of $\mathbb{P}$. Then, any $(\mathbb{P},\mathbb{F})$-martingale $\mathcal{M}(t)$ has the representions

$$\begin{aligned}
\mathcal{M}(t) = &\ \mathcal{M}(0) + \int_0^t \Gamma^{(\mathrm{D})}(s)dW(s) + \int_0^t \int_{\mathcal{Q}} \Gamma^{(\mathrm{MJ})}(s,q)\widetilde{\mathcal{P}}(\mathbf{ds},\mathbf{dq}) \\
= &\ \mathcal{M}(0) + \int_0^t \Gamma^{(\mathrm{D})}(t)dW(s) + \sum_{i=1}^{P(t)} \Gamma^{(\mathrm{MJ})}\left(T_i^-,Q_i\right) \\
&\ -\mathrm{E}_Q\Big[\Gamma^{(\mathrm{MJ})}(t,Q)\Big]\,\Lambda(t),
\end{aligned} \tag{12.9}$$

where $\Gamma^{(\mathrm{D})}(t)$ is a predictable (measurable with respect to $\mathbb{P}$), square-integrable process, while $\Gamma^{(\mathrm{MJ})}(t,q)$ is a $\mathcal{F}_t^{(W,P,Q)}$-predictable, $\mathcal{Q}$-marked process, such that

$$\mathrm{E}_Q\Big[\Gamma^{(\mathrm{MJ})}(t,Q)\Big] = \int_{\mathcal{Q}} \Gamma^{(\mathrm{MJ})}(t,q)\phi_Q(q;t)dq < \infty$$

and $\Lambda(t) \equiv \int_0^t \lambda(s)ds$ is the mean jump count.

The martingale representation theorem is used in the following Subsect. 12.2.2 for two versions of Girsanov's stochastic process transformation theorem, one for the diffusion process alone, i.e., without the Poisson terms in (12.9), and another for marked-jump-diffusion processes using the full form in (12.9).

The martingale approach may be a favored approach to solving SDE problems, but Heath and Schweizer [135] show the equivalence of the martingale and PDE approaches for a number of financial applications. The Feynmann-Kac formula (see (7.71) in the item on p. 219 here or the appendix of Duffie [74, Appendix E.] for more background) is used to solve the corrresponding PDE problem that is derived from the SDE.

12.2 Change in Probability Measure: Radon-Nikodým Derivatives and Girsanov's Theorem

12.2.1 Radon-Nikodým Theorem and Derivative for Change of Probability Measure:

Here, a version of the Radon-Nikodým Theorem 12.5 and derivative is formulated especially for probability measures and expectations. The abstract analog of the change of distribution corresponding to a change in random variables presented in Eq. (B.5) for the distribution and (B.6) for the density on p. B4 in preliminaries Chapt. B.

Theorem 12.12. ***Radon-Nikodým Change of Probability Measures:*** *Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ with σ-finite measure $\mathbb{P}_1$, if $\mathbb{P}_2$ is a finite measure that is mutually absolutely continuous with $\mathbb{P}_1$ (equivalent, $\mathbb{P}_2 \overset{\text{a.c.}}{\equiv} \mathbb{P}_1$) then there exists a positive measurable real function*

$$\mathbb{D}(x) = \frac{d\mathbb{P}_2}{d\mathbb{P}_1}(x) \quad \text{or} \quad \mathbb{D}(\mathcal{S}) = \frac{d\mathbb{P}_2}{d\mathbb{P}_1}(\mathcal{S}). \tag{12.10}$$

called the ***Radon-Nikodým derivative*** *of $\mathbb{P}_2$ with respect to $\mathbb{P}_1$, for $x \in \Omega$ such that for each measurable set $\mathcal{S} \in \mathcal{F}$*

$$\mathbb{P}_2(\mathcal{S}) = \mathrm{E}_{\mathbb{P}_1}[\mathbb{D}(X)\mathbf{1}_{X\in\mathcal{S}}] = \int_\Omega \mathbb{D}(x)\mathbf{1}_{x\in\mathcal{S}} d\mathbb{P}_1(x) = \int_{\mathcal{S}} \mathbb{D}(x) d\mathbb{P}_1(x), \tag{12.11}$$

where $d\mathbb{P}_i(x) = \mathbb{P}_i(dx)$ is equivalent notation for $i = 1{:}2$.

Further, if η is absolutely integrable with respect to the measure $\mathbb{P}_2$, then

$$\begin{aligned}\mathrm{E}_{\mathbb{P}_2}[\eta(X)] &= \int_\Omega \eta(x) d\mathbb{P}_2(x) = \int_\Omega \eta(x) \frac{d\mathbb{P}_2(x)}{d\mathbb{P}_1(x)} d\mathbb{P}_1(x) \\ &= \mathrm{E}_{\mathbb{P}_2}[\eta(X)\mathbb{D}(X)] = \int_\Omega \eta(x)\mathbb{D}(x)\mathbb{P}_1(x),\end{aligned}$$

i.e., using the Radon-Nikodým derivative in a measure-theoretic chain rule.

Thus, the Radon-Nikodým derivative is the analog of the Jacobian of the transformation (9.56) in an integral change of variables and leads to the absolutely

continuous measure chain rule, symbolically substituting for g,

$$d\mathbb{P}_2 = \frac{d\mathbb{P}_2}{d\mathbb{P}_1} d\mathbb{P}_1.$$

If $d\mathbb{P}_2$ and $d\mathbb{P}_1$ are mutually absolutely continuous, i.e., equivalent ($\mathbb{P}_1(\mathcal{S}) \overset{\text{a.c.}}{\equiv} \mathbb{P}_1(\mathcal{S})$), the Radon-Nikodým derivatives are mutual reciprocals,

$$\frac{d\mathbb{P}_1}{d\mathbb{P}_2} = 1 \Big/ \frac{d\mathbb{P}_2}{d\mathbb{P}_1} ,$$

formally justified by common null sets.

Examples 12.13. ***Radon-Nikodým Derivative Calculations:***

- ***Normal distributions:***
  Suppose a transformation from a standard normal distribution with density

  $$\phi_1(x) = \exp\left(-x^2/2\right) \Big/ \sqrt{2\pi}$$

  to a mean-μ, variance-σ^2 normal distribution with density,

  $$\phi_2(x) = \exp\left(-(x-\mu)^2 / \left(2\sigma^2\right)\right) \Big/ \sqrt{2\pi\sigma^2} \, .$$

  The change in measure coincides with a change of drift and a change of scale. Thus, $\mathbb{P}_1(x) = \Phi_1(x)$ is the first probability measure and the second is

  $$\mathbb{P}_2(x) = \Phi_2(x) = \int_{-\infty}^{\infty} \mathbb{D}(y)\phi_1(y)dy = \int_{-\infty}^{\infty} \mathbb{D}(y)\phi_1(y)dy,$$

  or $\phi_2(x) = \mathbb{D}(x)\phi_1(x)$ upon differentiating according to the fundamental theorem of integral calculus and the Radon-Nikodým derivative is

  $$\mathbb{D}(x) = \frac{d\mathbb{P}_2(x)}{d\mathbb{P}_1(x)} = \frac{d\Phi_2(x)}{d\Phi_1(x)} = \frac{\phi_2(x)}{\phi_1(x)} = \frac{\exp\left(-(x-\mu)^2/\left(2\sigma^2\right)\right) \Big/ \sqrt{2\pi\sigma^2}}{\exp\left(-x^2/2\right)/\sqrt{2\pi}}$$

  $$= \frac{1}{\sigma} \exp\left(- \frac{\left(1-\sigma^2\right)x^2 - 2\mu x + \mu^2}{2\sigma^2} \right). \tag{12.12}$$

  Hence, under measure $\mathbb{P}_1$ the random variable X has mean 0 and variance 1, but under measure $\mathbb{P}_2$ the random variable X has mean μ and variance σ^2.

  If $\sigma = 1$, then there is only a change of drift and the Radon-Nikodým derivative is simpler:

  $$\mathbb{D}(x) = \frac{d\mathbb{P}_2(x)}{d\mathbb{P}_1(x)} = \exp\left(\frac{\mu(2x-\mu)}{2} \right).$$

The more general form (12.12), formally justified here, can be transformed to the form in a proposition of Cont and Tankov [59, p. 306, Prop. 9.7] for two diffusion or Brownian motion processes, both denoted by $X = X(T)$, with parameters $\mu \to \mu_j T$ for the drifts and $\sigma^2 \to \sigma_1^2 T = \sigma^2 T = \sigma_2^2 T$ for a common variance on $(\Omega, \mathcal{F}_T, \mathbb{P}_j, \mathbb{F})$ for $j = 1:2$. Hence, using the fact the Radon-Nikodým derivative is the ratio of the two densities,

$$\begin{aligned}\mathbb{D}(X(T),T) &= \frac{d\mathbb{P}_2(X(T),T)}{d\mathbb{P}_1(X(T),T)} \\ &= \frac{\exp\left(-(X(T)-\mu_2 T)^2/\left(2\sigma_2^2 T\right)\right)\Big/\sqrt{2\pi\sigma_2^2 T}}{\exp\left(-(X(T)-\mu_1 T)^2/\left(2\sigma_1^2 T\right)\right)\Big/\sqrt{2\pi\sigma_1^2 T}} \\ &= \exp\left(\frac{2(\mu_2-\mu_1)X(T)-(\mu_2^2-\mu_1^2)T}{2\sigma^2}\right). \qquad (12.13)\end{aligned}$$

This corrects an error in [59, p. 306, Prop. 9.7]. They also convert this to the Cameron-Martin theorem form, by letting $X(T) = \mu_1 T + \sigma W_1(T)$, in the notation here, so

$$\mathbb{D}(T) = \frac{d\mathbb{P}_2(T)}{d\mathbb{P}_1(T)} = \exp\left(\frac{2(\mu_2-\mu_1)\sigma W_1(T)-(\mu_2-\mu_1)^2 T}{2\sigma^2}\right), \quad (12.14)$$

which is correct in [59, p. 306, following Prop. 9.7].

- ***Sets of Independent Random Variables:***
Let $\mathbf{X} = [X_i]_{n\times 1}$ be a set on n independent random variables with vector mean $\boldsymbol{\mu}^{(1)} = \left[\mu_i^{(1)}\right]_{n\times 1}$ and variance vector $\mathbf{V}^{(1)} = \left[\sigma_i^{(1)}\right]_{n\times 1}$, with product density

$$\phi^{(1)}(\mathbf{x}) = \prod_{i=1}^{n} \phi_i^{(1)}(x_i),$$

due to the independence property. The relationship between the measure, the distribution

$$\Phi^{(1)}(\mathbf{x}) = \mathrm{Prob}_{\mathbb{P}_1}[\mathbf{X} \le \mathbf{x}]$$

and the density can be written formally as

$$\frac{d\mathbb{P}_1(\mathbf{x})}{d\mathbf{x}} = \left(\prod_{i=1}^{n} \frac{\partial}{\partial x_i}\right) \Phi^{(1)}(\mathbf{x}) = \phi^{(1)}(\mathbf{x}),$$

where $\mathbf{X} \le \mathbf{x}$ means $X_i \le x_i$ for $i = 1{:}n$ and $d\mathbf{x} = \prod_{i=1}^{n} dx_i$ is the infinitesimal n-dimensional Euclidean measure, not a vector differential.

Let there be a function $\mathbb{D}(\mathbf{x})$ that generates a second distribution or measure,

$$\Phi^{(2)}(\mathbf{x}) = \text{Prob}_{\mathbb{P}_2}[\mathbf{X} \le \mathbf{x}] = \left(\prod_{i=1}^{n} \int_{-\infty}^{x_i} dy_i \phi_i^{(2)}(y_i) \right)$$

$$= \left(\prod_{i=1}^{n} \int_{-\infty}^{x_i} dy_i \phi_i^{(1)}(y_i) \right) \mathbb{D}(\mathbf{y}),$$

so

$$\frac{d\mathbb{P}_2(\mathbf{x})}{d\mathbf{x}} = \left(\prod_{i=1}^{n} \frac{\partial}{\partial x_i} \right) \Phi^{(2)}(\mathbf{x}) = \prod_{i=1}^{n} \int_{-\infty}^{x_i} dy_i \phi_i^{(2)}(x_i) = \phi_i^{(2)}(\mathbf{x})$$

$$= \mathbb{D}(\mathbf{x}) \prod_{i=1}^{n} \int_{-\infty}^{x_i} dy_i \phi_i^{(1)}(x_i) = \mathbb{D}(\mathbf{x}) \phi^{(1)}(\mathbf{x}).$$

Solving produces

$$\mathbb{D}(\mathbf{x}) = \frac{d\mathbb{P}_2(\mathbf{x})}{d\mathbb{P}_1(\mathbf{x})} = \frac{\phi_i^{(2)}(\mathbf{x})}{\phi^{(1)}(\mathbf{x})} = \prod_{i=1}^{n} \frac{\phi_i^{(2)}(x_i)}{\phi_i^{(1)}(x_i)}. \tag{12.15}$$

This result is important for stochastic processes $X(t)$ for $t \in [0, T]$, since a Radon-Nikodým derivative cannot be computed for a random variable over an infinite-dimensional interval, but it is possible to sample $X(t)$ at sample times $t_i = (i-1)T/n$ using $X_i = X(t_i)$ for $i = 1:n$, assuming the process of interest has independent increments.

As a more concrete example, suppose that the X_i have a standard normal distribution, i.e., IID with $\mu_i^{(1)} = 0$ and $\left(\sigma_i^{(1)}\right)^2 = 1$, and a nonstandard distribution is sought with mean $\mu_i^{(2)} = \mu_i$ and $\left(\sigma_i^{(2)}\right)^2 = \sigma_i^2$, then using (12.12),

$$\mathbb{D}(\mathbf{x}) = \frac{d\mathbb{P}_2(\mathbf{x})}{d\mathbb{P}_1(\mathbf{x})} = \frac{1}{\prod_{j=1}^{n} \sigma_j} \exp\left(-\sum_{i=1}^{n} \left(\frac{(1-\sigma_i^2)x_i^2 - 2\mu_i x_i + \mu_i^2}{2\sigma_i^2} \right) \right). \tag{12.16}$$

This example is similar to one in Glasserman [96], except there the $\sigma_i \equiv 1$.

- ***Poisson Distribution, a Discrete Analogy:***
  Next consider a Poisson cumulative distribution with parameter Λ_1 for the discrete variable N_1,

$$\Phi_n^{(1)} = \text{Prob}[N_1 < n] = e^{-\Lambda_1} \sum_{k=0}^{n} \frac{\Lambda_1^n}{n!}$$

which has increment (discrete derivative analog)

$$\Delta\Phi_{n-1}^{(1)} \equiv \Phi_n^{(1)} - \Phi_{n-1}^{(1)} = e^{-\Lambda_1} \frac{\Lambda_1^n}{n!},$$

the numerical forward difference notation, corresponding to a discrete density and consistent with Itô rules. The change of measure from variable N_1 with parameter Λ_1 to variable N_2 with parameter Λ_2 is given by

$$\Phi_n^{(2)} = \text{Prob}[N_2 < n] = e^{-\Lambda_2} \sum_{k=0}^{n} \frac{\Lambda_2^n}{n!} = e^{-\Lambda_1} \sum_{k=0}^{n} \mathbb{D}_n \frac{\Lambda_1^n}{n!},$$

with the Radon-Nikodým discrete derivative satisfying

$$\Delta\Phi_{n-1}^{(2)} = e^{-\Lambda_2} \frac{\Lambda_2^n}{n!} = \mathbb{D}_n e^{-\Lambda_1} \frac{\Lambda_1^n}{n!},$$

and solving yields

$$\mathbb{D}_n = \frac{\Delta\mathbb{P}_2(n-1)}{\Delta\mathbb{P}_1(n-1)} = \frac{\Delta\Phi_{n-1}^{(2)}}{\Delta\Phi_{n-1}^{(2)}} = e^{\Lambda_1} - \Lambda_2 \ \left(\frac{\Lambda 2}{\Lambda 1}\right)^n$$

$$= e^{\Lambda_1} - \Lambda_2 + n \ln(\Lambda_2/\Lambda_1). \tag{12.17}$$

Thus, with the change in measure from $\mathbb{P}_1$ to $\mathbb{P}_2$, the mean or average jump count changes from Λ_1 to Λ_2.

- ***Poisson Distribution with Fixed Size Jumps:***
Now, consider a Poisson distribution for discrete variable N_1 with parameter Λ_1 and constant jump size $\nu_1 \neq 0$, so

$$X = \nu_1 N_1.$$

Given the primary measure

$$\mathbb{P}_1(x) = \text{Prob}[X \leq x] = \text{Prob}[N_1 \leq x/\nu_1] = e^{-\Lambda_1} \sum_{k=0}^{\infty} \frac{\Lambda_1^k}{k!} \mathbf{1}_{\{k \leq x/\nu_1\}},$$

a change in measure with parameters $\{\Lambda_2, \nu_2\}$ is sought such that

$$\mathbb{P}_2(x) = \text{Prob}[X \leq x] = \text{Prob}[N_2 \leq x/\nu_2] = e^{-\Lambda_2} \sum_{k=0}^{\infty} \frac{\Lambda_2^{k_2}}{k_2!} \mathbf{1}_{\{k_2 \leq x/\nu_2\}}$$

$$= e^{-\Lambda_1} \sum_{k=0}^{\infty} \frac{\Lambda_1^{k_1}}{k_1!} \mathbf{1}_{\{k_1 \leq x/\nu_1\}} \mathbb{D}_{k_1}.$$

In lieu of a proper derivative for the indicator functions $\mathbf{1}_{\{k_j \leq x/\nu_j\}}$ for $j = 1{:}2$, consider the increment at $x = (n-1)\nu_2$,

$$\Delta\mathbb{P}_2((n-1)\nu_2) = \mathbb{P}_2((n-1)\nu_2 + \Delta x) - \mathbb{P}_2((n-1)\nu_2)$$

$$= e^{-\Lambda_2} \sum_{k=0}^{\infty} \frac{\Lambda_2^{k_2}}{k_2!} \mathbf{1}_{\{n-1 < k_2 \leq n-1+\Delta x/\nu_2\}}$$

$$= e^{-\Lambda_1} \sum_{k=0}^{\infty} \frac{\Lambda_1^{k_1}}{k_1!} \mathbf{1}_{\{(n-1)\nu_2/\nu_1 < k_1 \leq (n-1)\nu_2/\nu_1 + \Delta x/\nu_1\}} \mathbb{D}_{k_1}.$$

Aside from the coupling of the potential Radon-Nikodým discrete derivatives $\mathbb{D}_{k_1}$*, as Cont and Tankov [59, Prop. 9.5] state that the two measures will not be equivalent since their null sets will in general not coincide unless the jump sizes are the same,* $\nu_2 = \nu_1$*.*

Thus, with $\nu_2 = \nu_1$ *and* $\Delta x = \nu_1$ *for a semi-open unit step* $(n-1, n]$*, the new measure increment becomes*

$$\begin{aligned}\Delta\mathbb{P}_2((n-1)\nu_1) &= \mathbb{P}_2(n\nu_1) - \mathbb{P}_2((n-1)\nu_1) \\ &= e^{-\Lambda_2}\sum_{k=0}^{\infty}\frac{\Lambda_2^{k_2}}{k_2!}\mathbf{1}_{\{n-1<k_2\leq n\}} = e^{-\Lambda_2}\frac{\Lambda_2^n}{n!} \\ &= e^{-\Lambda_1}\sum_{k=0}^{\infty}\frac{\Lambda_1^{k_1}}{k_1!}\mathbf{1}_{\{n-1<k_1\leq n\}}\mathbb{D}_{k_1} = e^{-\Lambda_1}\frac{\Lambda_1^n}{n!}\mathbb{D}_n,\end{aligned}$$

so obtaining the same Radon-Nikodým discrete derivative as in the previous unit step example (12.17)

$$\mathbb{D}_n = \frac{\Delta\mathbb{P}_2(n-1)}{\Delta\mathbb{P}_1(n-1)} = e^{\Lambda_1 - \Lambda_2 + n\ln(\Lambda_2/\Lambda_1)}. \tag{12.18}$$

Note that although the original measures $\mathbb{P}_j(n\nu_j)$ *are RCLL as they should be, inherited from the indicators* $\mathbf{1}_{\{k\leq n\}}$*, the increment* $\Delta\mathbb{P}_1((n-1)\nu_1)$ *is* ***LCRL*** (***left continuous, right limits***) *due to the indicator increments* $\mathbf{1}_{\{n-1<k_2\leq n\}}$*, but they precisely allow the selection of just the* n*th jump term in the Poisson distribution sum since the indicator increments are closed at* n *and open at* $n-1$*.*

This Poisson distribution example is an applied justification of the proposition in Cont and Tankov [59, Prop. 9.5] for two Poisson processes $n = N_j = P(T) = N(T)$ *with parameters* $\Lambda_j = \lambda_j T$ *on* $(\Omega, \mathcal{F}_T, \mathbb{P}_j, \mathbb{F})$ *for* $j = 1{:}2$*, i.e.,*

$$\frac{\Delta\mathbb{P}_2(N(T)-1)}{\Delta\mathbb{P}_1(N(T)-1)} = e^{(\lambda_1-\lambda_2)T + N(T)\ln(\lambda_2/\lambda_1)}, \tag{12.19}$$

but only for the same size, $\nu_2 = \nu_1$*, which has an explicit form as given here.*

12.2.2 Change in Measure for Stochastic Processes: Girsanov's Theorem

There are many versions of Girsanov's theorem for changing a probability measure to change the drift of a stochastic diffusion process and some of these variants are not very distinguishable from the Radon-Nikodým theorem. Here, a modification of Runggaldier's [239] (see also Brémaud [43] for even more details) version will be followed since it has been found to be the most useful, the Radon-Nikodým derivative being relatively easy to calculate and comes with an extension to jump-diffusions. The application of this theorem is determining the measure change for

a relative change $\gamma(t)$ of the drift from $\mu_1(t)$ to a drift $\mu_2(t)$ appropriate for the problem of interest, e.g., the change of the drift coefficient $\mu_1(t) = \mu$ in the Black-Scholes [34] method to the current market rate $\mu_2(t) = r$.

Diffusion Girsanov Transformations

Let the reference $\mathbb{P}_1$-SDE for a state diffusion process $X(t)$ be

$$dX(t) = \mu_1(t)dt + \sigma(t)dW_1(t) \tag{12.20}$$

with time-dependent coefficients $\{\mu_1(t), \sigma(t)\}$, whose integrabilities are implied by the following Girsanov diffusion theorem, on a finite time interval $[0, T]$ on the filtered probability space $(\Omega, \mathcal{F}_t, \mathbb{P}_1, \mathbb{F})$ with $W_1(t)$ being a $\mathbb{P}_1$-Wiener process. In addition, let the target $\mathbb{P}_2$-SDE objective for this state diffusion process $X(t)$ be

$$dX(t) = \mu_2(t)dt + \sigma(t)dW_2(t) \tag{12.21}$$

with the same volatility $\sigma(t)$ but changed to drift $\mu_2(t)$, integrability also implied, on the finite time interval $[0, T]$ on the filtered probability space $(\Omega, \mathcal{F}_t, \mathbb{P}_2, \mathbb{F})$ with $W_1(t)$ being a corresponding $\mathbb{P}_2$-Wiener process.

Theorem 12.14. ***Girsanov's Theorem for Changing the Probability Measure of a Diffusion Process to Change the Drift:***
Let $(\Omega, \mathcal{F}_t, \mathbb{P}_1, \mathbb{F})$ be a filtered probability space with $\mathbb{F} = \cup_t \mathcal{F}_t$, symbolically over t. Let $\gamma(t)$ be a square integrable predictable (measurable with respect to $\mathbb{P}_1$, i.e., knowable given $\mathcal{F}_t$) drift process

$$\int_0^t \gamma^2(s)ds < \infty$$

for all $t \in [0, T]$. Then, the Radon-Nikodým derivative $\mathbb{D}(t)$ at time t for the process $X(t)$ is given by the martingale representatiion (12.9),

$$d\mathbb{D}(t) = \mathbb{D}(t)\gamma(t)dW_1(t), \quad \mathbb{D}(0) \overset{\text{w.p.o.}}{=} 1, \tag{12.22}$$

supposing that $\mathrm{E}_{\mathbb{P}_1}[\mathbb{D}(t)] = 1$ and there exists a second probability measure $\mathbb{P}_2$ on $\mathbb{F}$ that is equivalent to $\mathbb{P}_1$ (mutually absolutely continuous, $\mathbb{P}_2 \overset{\text{a.c.}}{\equiv} \mathbb{P}_1$), such that

$$d\mathbb{P}_2 = \mathbb{D}(t)d\mathbb{P}_1$$

and

$$dW_2(t) = dW_1(t) - \gamma(t)dt, \tag{12.23}$$

where $W_1(t)$ is a $\mathbb{P}_1$-Wiener process as in (12.20) while $W_2(t)$ is a $\mathbb{P}_2$-Wiener process as in (12.21).

The Radon-Nikodým derivative is explicitly given by

$$\mathbb{D}(t) = \frac{d\mathbb{P}_2(t)}{d\mathbb{P}_1(t)} = \exp\left(\int_0^t \gamma(s)\left(dW_1(s) - \frac{1}{2}\gamma(s)ds\right)\right) \tag{12.24}$$

and the relative drift change is

$$\gamma(t) = \frac{\mu_2(t) - \mu_1(t)}{\sigma(t)}. \tag{12.25}$$

If the filtration

$$\mathbb{F} = \mathcal{F}_t^{(W_1)} = \widehat{\sigma}\{W_1(s),\ \mathcal{S}_N : 0 \le s \le t,\ \mathcal{S}_N \in \mathcal{N}_1\},$$

$\mathcal{N}_1$ *is the collection of null-sets of* $\mathbb{P}_1$*, then conversely every probability measure* $\mathbb{P}_2 \overset{\text{a.c.}}{\equiv} \mathbb{P}_1$ *has the same Radon-Nikodým derivative structure.*

Substituting (12.23) the Wiener process shift into the original SDE,

$$dX(t) = \mu_1(t)dt + \sigma(t)dW_1(t) = (\mu_1(t) + \sigma(t)\gamma(t))dt + \sigma(t)dW_2(t),$$

so comparing to the second SDE $\mu_2(t) = \mu_1(t) + \sigma(t)\gamma(t)$ and (12.25) for $\gamma(t)$ is immediate, given common volatilities $\sigma_1(t) = \sigma(t) = \sigma_2(t)$.

Upon applying the Itô stochastic chain rule to solve the $\mathbb{D}$-SDE (12.22),

$$d\ln(\mathbb{D}(t)) \overset{\text{dt}}{=} \frac{\gamma\mathbb{D}dW_1}{\mathbb{D}} - \frac{(\gamma\mathbb{D})^2dt}{2\mathbb{D}^2} = \gamma\left(dW_1 - \frac{1}{2}\gamma dt\right),$$

integrating with $\mathbb{D}(0) = 1$ and inverting the logarithm, the answer for $\mathbb{D}(t)$ in (12.24) follows. Note that the assumption of common volatility is essential for obtaining the simple linear SDE in $\mathbb{D}(t)$ given in (12.22), since from just one of the independent example terms i in (12.16) it is seen that there is a quadratic term in x of the ith exponent unless $\sigma_i^{(2)} = \sigma_i = \sigma_i^{(1)} = 1$, the common σ in this example. Hence, this Girsanov theorem is quite simple and special. The crudely derived constant coefficient case in (12.14), as an example for Radon-Nikodým derivatives, can be properly recovered from the Girsanov form (12.24) by setting $t = T$ and replacing the time-dependent coefficients by constants, i.e., $\mu_j(s) \to \mu_j$ for $j = 1:2$ and $\sigma(s) \to \sigma$.

Note that the relative drift shift (12.25), being state independent, is also the same for the linear diffusion case,

$$dX(t) = X(t)\left(\mu_1(t)dt + \sigma(t)dW_1(t)\right), \tag{12.26}$$

which is important for applications in finance. This is a linear SDE for **geometric Brownian motion (GBM) or multiplicative diffusion noise** of the Black-Scholes-Merton[34, 201] option pricing model, while the reference SDE 12.26 for the Theorem 12.14 is for **arithmetic Brownian motion (ABM) or additive diffusion noise** of the historic 1900 Bachelier [16] model. The multiplicative model is better for compounded effects, while the additive model is better for strictly cumulative effects. It is well-known that the multiplicative model can be transformed into an additive one by the logarithmic transformation using Itô rules,

$$d\ln(X(t)) = \left(\mu_1 - \sigma^2(t)/2\right)dt + \sigma(t)dW_1(t). \tag{12.27}$$

Since the diffusion coefficient shift, $\sigma^2(t)/2$, of the drift would be the same for the GBM target model (2) as for the GBM reference model (1), it is clear that the diffusion Girsanov transformation of the drift will be the same as for the ABM model, i.e.,

$$\gamma^{(\mathrm{GBM})}(t) = \frac{\mu_2(t) - \mu_1(t)}{\sigma(t)}. \tag{12.28}$$

Marked-Jump-Diffusion Girsanov Transformations

Now consider the case of marked-jump-diffusions or compound-jump-diffusion. Let the reference $\mathbb{P}_1$-SDE for a state marked-jump-diffusion process $X(t)$ be

$$dX(t) = \mu_1(t)dt + \sigma(t)dW_1(t) + \int_{\mathcal{Q}_1} h_1(t,q)\mathcal{P}_1(\mathbf{dt},\mathbf{dq}), \tag{12.29}$$

with $\mathbb{P}_1$-Wiener process $W_1(t)$, $\mathbb{P}_1$-Poisson process $P_1(t)$, $\mathrm{E}_{\mathbb{P}_1}[dP_1(t)] = \lambda_1(t)dt$ defines the jump-rate, integrable time-dependent coefficients $\{\mu_1(t), \sigma(t), \lambda_1(t)\}$, (time, mark)-dependent jump-amplitude $h_1(t,q)$, whose integrability is implied by the following theorem, $\mathbb{P}_1$-Poisson jump-times T_i and IID sample marks Q_i distributed with density $\phi_Q^{(1)}(q;t)$ on the filtered probability space $(\Omega, \mathcal{F}_t, \mathbb{P}_1, \mathbb{F})$ with $W_1(t)$ over on a finite time-interval $[0,T]$. Several forms of the Poisson measure integrals,

$$\int_{\mathcal{Q}_1} h_1(t,q)\mathcal{P}_1(\mathbf{dt},\mathbf{dq}) = \sum_{k=1}^{dP_1(t)} h_1(T_k^-, Q_k) \stackrel{\mathrm{dt}}{=} h_1(t,Q)dP_1(t),$$

will be used here, sometimes one form being more convenient than the other.

In addition, let the target $\mathbb{P}_2$-SDE objective for this state marked-jump-diffusion process $X(t)$ be

$$dX(t) = \mu_2(t)dt + \sigma(t)dW_2(t) + \int_{\mathcal{Q}_2} h_2(t,q)\mathcal{P}_2(\mathbf{dt},\mathbf{dq}), \tag{12.30}$$

with $\mathbb{P}_2$-Wiener process $W_2(t)$, $\mathbb{P}_2$-Poisson process $P_2(t)$, $\mathrm{E}_{\mathbb{P}_2}[dP_2(t)] = \lambda_2(t)dt$ defines the jump-rate, the same volatility $\sigma(t)$ but changed to drift $\mu_2(t)$ and changed jump-rate $\lambda_2(t)$, (time, mark)-dependent jump-amplitude $h_2(t,q)$, integrability also implied, $\mathbb{P}_2$-Poisson jump-times T_i and IID sample marks Q_i distributed with density $\phi_Q^{(2)}(q;t)$ on the finite time interval $[0,T]$ on the filtered probability space $(\Omega, \mathcal{F}_t, \mathbb{P}_2, \mathbb{F})$.

The following theorem follows Runggaldier [239, Theorem 2.4], but is also presented more in the notation of this book.

Theorem 12.15. ***Girsanov's Theorem for Changing the Probability Measure of a Jump-Diffusion Process to Change the Drift, the Jump-Rate and Mark-Density:***

Let $(\Omega, \mathcal{F}_t, \mathbb{P}_1, \mathbb{F})$ a filtered probability space on the finite time-interval $[0,T]$ with the mark space $\mathcal{Q} = \mathbb{R}$. and the (jump-rate, mark-density)-characteristics

$$(\lambda_1(t), \Phi_Q(\mathbf{dq}; t) \stackrel{\text{gen}}{=} (\lambda_1(t), \phi_Q(q;t)dq).$$

Let $\gamma^{(\mathrm{D})}(t)$ be the square integrable diffusion drift change given in (12.25)

$$\gamma^{(\mathrm{D})}(t) = \frac{\mu_2(t) - \mu_1(t)}{\sigma(t)} \tag{12.31}$$

of Theorem 12.14. Let $\gamma^{(\mathrm{J})}(t)$ be a nonnegative, $\mathcal{F}_t$-predictable jump-rate user-defined scaling process such that

$$\Lambda_2(t) \equiv \int_0^t \lambda_2(s)ds = \int_0^t \gamma^{(\mathrm{J})}(s)\lambda_1(s)ds < \infty$$

for all $t \in [0,T]$, i.e., the transformed mean jump count is finite, and let $\gamma^{(\mathrm{M})}(q;t)$ be a nonnegative, $\mathcal{F}_t$-predictable, $\mathcal{Q}$-space dependent mark-distribution user-defined scaling process such that

$$\int_{\mathcal{Q}_2} \phi_Q^{(2)}(q;t)dq = \int_{\mathcal{Q}_1} \gamma^{(\mathrm{M})}(q;t)\phi_Q^{(1)}(q;t)dq = 1,$$

i.e., transformed mark-space probability is conserved.

Let

$$\mathbb{D}(t) = \mathbb{D}^{(\mathrm{D})}(t)\mathbb{D}^{(\mathrm{MJ})}(t),$$

where the diffusion martingale representation factor $\mathbb{D}^{(\mathrm{D})}(t)$ is given in (12.24) with stochastic differential in (12.22) and the marked-jump factor $\mathbb{D}^{(\mathrm{MJ})}(t)$ is given by the marked-jump martingale representation (12.9),

$$d\mathbb{D}^{(\mathrm{MJ})}(t) = \mathbb{D}^{(\mathrm{MJ})}(t) \int_{\mathcal{Q}_1} \left(\gamma^{(\mathrm{J})}(t)\gamma^{(\mathrm{M})}(q;t) - 1\right) \widehat{\mathcal{P}}(\mathbf{dt}, \mathbf{dq}), \tag{12.32}$$

subject to the side condition

$$\mathrm{E}_{\mathbb{P}_1}[\mathbb{D}^{(\mathrm{MJ})}] = 1,$$

where

$$\widehat{\mathcal{P}}(\mathbf{dt}, \mathbf{dq}) \equiv \mathcal{P}(\mathbf{dt}, \mathbf{dq}) - \mathrm{E}[\mathcal{P}(\mathbf{dt}, \mathbf{dq})] = \mathcal{P}(\mathbf{dt}, \mathbf{dq}) - \left(\gamma^{(\mathrm{J})}(t) - 1\right) \lambda_1(t)dt,$$

so the solution to the SDE in (12.32) is

$$\begin{aligned} \mathbb{D}^{(\mathrm{MJ})}(t) = &\quad \exp\left(\int_0^t \left(\int_{\mathcal{Q}_1} \ln\left(\gamma^{(\mathrm{J})}(s)\gamma^{(\mathrm{M})}(q;s)\right) - \left(\gamma^{(\mathrm{J})}(s) - 1\right)\lambda_1(s)ds\right)\right) \\ = &\quad \exp\left(\int_0^t - \left(\gamma^{(\mathrm{J})}(s) - 1\right)\lambda_1(s)ds\right) \prod_{k=1}^{P_1(t)} \gamma^{(\mathrm{J})}(T_k^-)\gamma^{(\mathrm{M})}(Q_k; T_k^-). \end{aligned} \tag{12.33}$$

The transformed quantities are

$$dW_2(t) = dW_1(t) - \gamma^{(\mathrm{D})}(t)dt, \tag{12.34}$$

$$\lambda_2(t) = \gamma^{(\mathrm{J})}(t)\lambda_1(t) \tag{12.35}$$

$$\phi_Q^{(2)}(q;t) = \gamma^{(\mathrm{M})}(q;t)\phi_Q^{(1)}(q;t). \tag{12.36}$$

Thus, the explicit form of the marked-jump-diffusion Radon-Nikodým derivative is

$$\begin{aligned}\mathbb{D}(t) = & \frac{d\mathbb{P}_2(t)}{d\mathbb{P}_1(t)} \\ = & \exp\left(\int_0^t \left(\gamma^{(\mathrm{D})}(s)\left(dW_1(s) - \gamma^{(\mathrm{D})}(s)ds/2\right) - \left(\gamma^{(\mathrm{J})}(s) - 1\right)\lambda_1(s)ds\right)\right) \\ & \cdot \prod_{i=1}^{P_1(t)} \gamma^{(\mathrm{J})}(T_k^-)\gamma^{(\mathrm{M})}(Q_k;T_k^-).\end{aligned} \tag{12.37}$$

If the filtration

$$\mathbb{F} = \mathcal{F}_t^{(W_1,P_1,Q)} = \widehat{\sigma}\{W_1(s), P_1(t), \mathcal{S}_Q,\ \mathcal{S}_N : 0 \le s \le t,\ \mathcal{S}_Q \in \mathcal{Q}_1,\ \mathcal{S}_N \in \mathcal{N}_1\},$$

$\mathcal{N}_1$ is the collection of null-sets of $\mathbb{P}_1$, then conversely every probability measure $\mathbb{P}_2 \overset{\text{a.c.}}{\equiv} \mathbb{P}_1$ has the same Radon-Nikodým derivative structure.

Note that the Wiener process W_1 is independent of the marked Poisson process double (P_1, Q), but the mark random variables Q are only conditionally independent of P_1 and that condition is that there exists a jump of the state X in time, so the factoring $\mathbb{D}(t) = \mathbb{D}^{(\mathrm{D})}(t)\mathbb{D}^{(\mathrm{MJ})}(t)$ into only two parts makes sense. Also, using the product form of Itô's stochastic chain rule,

$$\begin{aligned}d\mathbb{D}(t) &= \mathbb{D}^{(\mathrm{MJ})}(t)d\mathbb{D}^{(\mathrm{D})}(t) + \mathbb{D}^{(\mathrm{D})}(t)d\mathbb{D}^{(\mathrm{MJ})}(t) + d\mathbb{D}^{(\mathrm{D})}(t)d\mathbb{D}^{(\mathrm{MJ})}(t) \\ &\overset{\mathrm{dt}}{=} \mathbb{D}^{(\mathrm{MJ})}(t)d\mathbb{D}^{(\mathrm{D})}(t) + \mathbb{D}^{(\mathrm{D})}(t)d\mathbb{D}^{(\mathrm{MJ})}(t) \\ &= \mathbb{D}(t)\left(\gamma^{(\mathrm{D})}(t)dW_1(t) + \int_{\mathcal{Q}_1}\left(\gamma^{(\mathrm{J})}(t)\gamma^{(\mathrm{M})}(q;t) - 1\right)\widehat{\mathcal{P}}_1(\mathbf{dt},\mathbf{dq})\right).\end{aligned}$$

Since (12.32) for $\mathbb{D}^{(\mathrm{MJ})}$ is linear, in formal dt-precision notation,

$$d\ln\left(\mathbb{D}^{(\mathrm{MJ})}(t)\right) = -\left(\gamma^{(\mathrm{J})}(t) - 1\right)\lambda_1(t)dt + \ln\left(\gamma^{(\mathrm{J})}(t)\gamma^{(\mathrm{M})}(q;t)\right)dP_1(t),$$

since if dP_1 jumps, then the jump is given by

$$\left[\mathbb{D}^{(\mathrm{MJ})}\right](t) = \left(\gamma^{(\mathrm{J})}(t)\gamma^{(\mathrm{M})}(q;t) - 1\right)dP_1(t)$$

and the jump of the logarithm is

$$\begin{aligned}\left[\ln\left(\mathbb{D}^{(\mathrm{MJ})}\right)\right](t) &= \left(\ln\left(\mathbb{D}^{(\mathrm{MJ})}+\left(\gamma^{(\mathrm{J})}(t)\gamma^{(\mathrm{M})}(q;t)-1\right)\mathbb{D}^{(\mathrm{MJ})}\right)-\ln\left(\mathbb{D}^{(\mathrm{MJ})}\right)\right)dP_1(t)\\ &= \ln\left(\gamma^{(\mathrm{J})}(t)\gamma^{(\mathrm{M})}(q;t)-1\right)dP_1(t),\end{aligned}$$

so

$$\begin{aligned}\mathbb{D}^{(\mathrm{MJ})}(t) &= \exp\Big(-\int_0^t\Big(\gamma^{(\mathrm{J})}(s)-1\Big)\lambda_1(s)ds+\sum_{i=1}^{P_1(t)}\ln\Big(\gamma^{(\mathrm{J})}(t)\gamma^{(\mathrm{M})}(q;t)\Big)\Big)\\ &= \exp\Big(-\int_0^t\Big(\Big(\gamma^{(\mathrm{J})}(s)-1\Big)\lambda_1(s)ds+\int_{\mathcal{Q}_1}\ln\left(\gamma^{(\mathrm{J})}(s)\gamma^{(\mathrm{M})}(q;s)\right)\mathcal{P}_1(\mathbf{ds},\mathbf{dq})\Big)\Big).\end{aligned}\tag{12.38}$$

Finally, combining equations (12.22), (12.24), (12.32) and (12.38), along with converting the exponential of a sum to a product yields the result (12.37) for $\mathbb{D}(t)$ for the marked-jump-diffusion change from measure $\mathbb{P}_1$ to $\mathbb{P}_2$ according to the recipe (12.34) to (12.36).

For the geometric or linear marked-jump-diffusion,

$$dX(t)=X(t)\left(\mu_1(t)dt+\sigma(t)dW_1(t)+\sum_{i=1}^{dP_1(t)}h_1(T_i,Q_i)\right),\tag{12.39}$$

the logarithmic change of variable can transform the geometric model to an arithmetic one like (12.29),

$$d\ln(X(t))=\left(\mu_1(t)-\sigma^2(t)/2\right)dt+\sigma(t)dW_1(t)+\sum_{i=1}^{dP_1(t)}\ln\left(h_1(T_i,Q_i)+1\right).\tag{12.40}$$

Again assuming a common volatility $\sigma(t)$, the Itô rule diffusion coefficient shift of the drift coefficient will be common in both target (2) and reference (1) models, while the jump-rate $\lambda_1(t)$ and jump-amplitude distribution is unchanged, then the Girsanov tranformation triplet,

$$\begin{aligned}dW_2(t) &= dW_1(t)-\gamma^{(\mathrm{D})}(t)dt,\\ \lambda_2(t) &= \gamma^{(\mathrm{J})}(t)\lambda_1(t)\\ \phi_Q^{(2)}(q;t) &= \gamma^{(\mathrm{M})}(t)\phi_Q^{(1)}(q;t),\end{aligned}\tag{12.41}$$

will be preserved for the geometric case.

Also, see Øksendal and Sulem's Lévy process book [223] for a combined jump-rate and mark distribution scaling, with some financial examples.

Example 12.16. ***Two State, 2 Noise Model Girsanov Application:***

In order to determine both the diffusive scaling $\gamma^{(\mathrm{D})}(t)$ and the jump scaling $\gamma^{(\mathrm{J})}(t)$, at least two states (assets in financial applications) are needed to handle two sources of random noise. Following a financial example of Runggaldier [239], let $X_1(t)$ and $X_2(t)$ be two states with the same jump-diffusion noise, $W_1(t)$ and $P_1(t)$, but the

jump-amplitude is assumed to be deterministic in magnitude given a Poisson jump-time, so $\gamma^{(\mathrm{M})}(q;t) \equiv 1$ *here since there are no marks in the problem. The SDE dynamics are given by*

$$dX_i(t) = X_i(t)\left(\mu_i(t)dt + \sigma_i(t)dW_1(t) + \nu_i(t)dP_1(t)\right), \tag{12.42}$$

for $i = 1:2$ *and where* $\mathrm{E}[dP_1(t)] = \lambda_1(t)dt$.

Let the 2nd measure transformed noise be given by

$$dW_2(t) = dW_1(t) - \gamma^{(\mathrm{D})}(t)dt, \qquad dP_2(t) = dP_1(t) - \gamma^{(\mathrm{J})}(t)\lambda_1(t)dt.$$

Hence, the dynamics are transformed to

$$dX_i(t) = X_i(t)\left(\left(\mu_i(t) + \gamma^{(\mathrm{D})}(t)\sigma_i(t) + \gamma^{(\mathrm{J})}(t)\nu_i(t)\lambda_1(t)\right)dt + \sigma_i(t)dW_2(t) + \nu_i(t)dP_2(t)\right),$$

such that the common Radon-Nikodým derivative from (12.37)

$$\begin{aligned}\mathbb{D}(t) = \exp\Big(&\int_0^t \left(\gamma^{(\mathrm{D})}(s)\left(dW_1(s) - \gamma^{(\mathrm{D})}(s)ds/2\right) - \left(\gamma^{(\mathrm{J})}(s) - 1\right)\lambda_1(s)ds\right) \\ &+\gamma^{(\mathrm{J})}(t)P_1(t)\Big),\end{aligned} \tag{12.43}$$

depending only on the given, common noise, so yields an equivalent martingale measure $\mathbb{P}_2(t)$ *transformed from* $\mathbb{P}_1(t)$.

For convenience in applications, a scaling of the state $\widetilde{X}_i(t) = X_i(t)/B(t)$ *is introduced using a deterministic process*

$$dB(t) = r(t)B(t)dt,$$

$B(0) > 0$ *and* $r(t) \geq 0$ *that in finance would be called discounting if* $B(t)$ *were a riskless asset like a zero-coupound bond such that* $B(t)$ *is called the numeraire, the most common one. Thus by the chain rule,*

$$\begin{aligned}d\widetilde{X}_i(t) = &\left(d^{(cont)}X_i(t)\right)\Big/ B(t) - X_i(t)dB(t)/B^2(t) + [X_i/B](t)dP_1(t) \\ = &\widetilde{X}_i(t)\left(\left(\mu_i(t) + \gamma^{(\mathrm{D})}(t)\sigma_i(t) + \gamma^{(\mathrm{J})}(t)\nu_i(t)\lambda_1(t) - r(t)\right)dt\right. \\ &\left.+\sigma_i(t)dW_2(t) + \nu_i(t)P_2(t)\right).\end{aligned}$$

Selecting the common diffusion scaling $\gamma^{(\mathrm{D})}(t)$ *for both* $X_i(t)$*'s and consequently get the common jump scaling* $\gamma^{(\mathrm{J})}(t)$*, so*

$$\gamma^{(\mathrm{D})}(t) = \left(r(t) - \mu_i(t) - \gamma^{(\mathrm{J})}(t)\nu_i(t)\lambda_1(t)\right)\Big/ \sigma_i(t),$$

for $i = 1:2$*. Solving simultaneously for the two scalings produces solutions explicit in the given parameters,*

$$\gamma^{(\mathrm{J})}(t) = \frac{\sigma_1(t)(r(t) - \mu_2(t)) - \sigma_2(t)(r(t) - \mu_1(t))}{(\sigma_1(t)\nu_2(t) - \sigma_2(t)\nu_1(t))\lambda_1(t)} \tag{12.44}$$

and

$$\gamma^{(\mathrm{D})}(t) = \frac{\nu_1(t)(\mu_2(t) - r(t)) + \nu_2(t)(r(t) - \mu_1(t))}{(\sigma_1(t)\nu_2(t) - \sigma_2(t)\nu_1(t))}, \tag{12.45}$$

provided $(\sigma_1(t)\nu_2(t) - \sigma_2(t)\nu_1(t)) \neq 0$ *and* $\gamma^{(\mathrm{J})}(t)\lambda_1(t) > 0$. *This produces a unique martingale measure and in finance the measure uniqueness is required for the completeness of the market [239, 133]. The presence of an infinite number of mark IID random variables requires an infinite number of states or assests to exactly show uniqueness of the transformed martingale measure* $\mathbb{P}_2(t)$.

Refer to Runggaldier's jump-diffusion handbook article [239] for more information and examples on the multidimensional case, Poisson random measure formulation and financial applications.

12.3 Itô, Lévy and Jump-Diffusion Comparisons

12.3.1 Itô Processes and Jump-Diffusion Processes

Many authors, Bingham and Kiesel [33] Duffie [74], Glasserman [96], Hull [147], Merton [203], Mikosch [209], Øksendal [222] and others, mostly refer to Brownian motion or Wiener-driven processes with Wiener scaling by a factor $\sigma(t)$ and translated by drift $\mu(t)$,

$$dX(t) = \mu(t)dt + \sigma(t)dW(t), \tag{12.46}$$

at least as basic definition of an **Itô process**. Some such as Glasserman [96], Hull [147], Merton [203] and Mikosch [209] would explicitly allow the composite interpretation of the coefficient functions in the basic definition (12.46) to include dependence on the state $X(t)$, such that $\mu(t) = f(X(t), t)$, $\sigma(t) = g(X(t), t)$ and

$$dX(t) = f(X(t), t)dt + g(X(t), t)dW(t). \tag{12.47}$$

Others extend the basic Itô process (12.46) to include (12.47) by application of the Itô chain rule using a transformation like $\widehat{X}(t) = F(X(t), t)$ to obtain

$$d\widehat{X}(t) = f(X(t), t)dt + g(X(t), t)dW(t),$$

where

$$f(X(t), t) = \left(F_t + \mu(t)F_x + \frac{1}{2}\sigma^2(t)F_{xx}^2)\right)(X(t), t)$$

and

$$g(X(t), t) = \sigma(t)F_x(X(t), t).$$

Thus, state dependent formula (12.47) will be taken as an acceptable **definition of the Itô process**.

However, in his stochastic differential equation classic 1951 memoir [149], Itô also correctly includes jumps in his discussion of simple Markov processes. Itô

referred to simple Markov processes, specified by a stochastic differential equation, which for general Poisson noise with distributed jump-amplitudes might be called stochastic integral differential equations,

$$dX(t) = f(X(t),t)dt + g(X(t),t)dW(t) + \int_{\mathcal{Q}} h(X(t),t,q)\mathcal{P}(dt,dq)\ , \quad (12.48)$$

in our notation, or preferably by a stochastic integral equation,

$$X(t) = X(t_0) + \int_{t_0}^{t}\Big[f(X(s),s)ds + g(X(s),s)dW(s) + \int_{\mathcal{Q}} h(X(s),s,q)\mathcal{P}(ds,dq)\Big]\ , \quad (12.49)$$

again in our notation. Hence, there is a historical basis for calling the jump-diffusion processes that are the focus of this book as **Itô processes**.

Still others, for instance, Tavella and Randall [264] refer to a jump-diffusion processes as a superposition of an Itô process and a Poisson jump process, while Øksendal and Sulem [223] refer to a similar combination as an Itô-Lévy process, but see the next subsection on Lévy processes for the differences between jump-diffusion and Lévy processes. Applebaum [12] and others more precisely call diffusion processes like (12.47) **Itô diffusion processes**.

Although diffusion processes are easier to treat since they have continuous sample paths, jump processes and jump-diffusion processes have discontinuous sample paths so are relatively more difficult to prove theorems for. Some of the most significant changes occur with jumps, such as extreme financial crashes and natural disasters.

Hence, according to the more of less standard Itô process usage (12.47),

$$\textit{Itô processes} \subset \textit{Jump-diffusion processes}. \quad (12.50)$$

12.3.2 Lévy Processes and Jump-Diffusion Processes

Lévy processes are essentially jump-diffusion processes, but extended to processes with infinite jump rates. There have been much recent efforts in the literature studying and applying Lévy processes, such as Carr, Geman, Madan and Yor (CGMY model) [46], Carr and Madan (VG model) [47] and Rydberg (NIG model) [243]. Sometimes the term non-Gaussian processes is used as in Barndorff-Nielsen and Shepherd (GIG model) [20], but may not necessarily mean strict Lévy processes. There also several recent books on Lévy processes such as that of Applebaum [12], as well as others on Lévy processes but with jump processes or jump-diffusions in the titles such as those of Cont and Tankov [59] and Øksendal and Sulem [223]. As with other abstract concepts, there are many different definitions of a Lévy process, and some attempt has been made to merge them within the spirit of this book.

Definition 12.17. ***Basic Lévy Process Conditions:***
A ***Lévy process*** *satisfies the following conditions:*

- ***RCLL stochastic process:*** *$\{\mathbf{X}(t),\ t \geq 0\}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^{n_x}$ (the term **cádlág** means RCLL in French but is used in English probability texts too).*

- ***Initial condition:*** $\mathbf{X}(0) \stackrel{\text{a.s.}}{=} \mathbf{0}$.

- ***Independent increments:*** *for every partition $0 = t_0 < t_1 < t_2 < \cdots < t_{n_t} < \infty$, the increments*

$$\Delta\mathbf{X}(t_j) \equiv \mathbf{X}(t_{j+1}) - \mathbf{X}(t_j), \text{ for } j = 0{:}n-1 \tag{12.51}$$

  are independent.

- ***Stationary increments:*** *Together with independence,*

$$\Delta\mathbf{X}(t_j) \stackrel{\text{dist}}{=} \mathbf{X}(\Delta t_j), \tag{12.52}$$

  where $\Delta t_j \equiv t_{j+1} - t_j$.

- ***Stochastic Continuity:*** *The increments of $\mathbf{X}(t)$ satisfy,*

$$\lim_{\Delta t \to 0} \text{Prob}[|\mathbf{X}(t + \Delta t) - \mathbf{X}(t)| \geq \epsilon] = 0, \quad \forall\ \epsilon > 0 \ \text{ and } \ t \geq 0. \tag{12.53}$$

All but the last condition (12.53) are standard for the processes dealt with here when the **coefficients are constant**, so it is usually sufficient to show stochastic continuity (*note that continuous in probability is not the same as continuous*). However, when the process coefficients are not constant, then the process will in general not be stationary as Lévy condition (12.52) requires. For many real problems the process coefficients, as in financial markets, time-dependence is important (for instance, see Hanson and Westman [126]), so (12.52) will not be valid in these problems. It is clear that the IID Wiener vector process $\mathbf{W}(t)$ or the Wiener driven vector Gaussian process with constant coefficients

$$\mathbf{G}(t) = \boldsymbol{\mu}_0 t + \sigma_0 \mathbf{W}(t)$$

and the Poisson vector process $\mathbf{P}(t)$ with constant jump rates $\boldsymbol{\lambda}(t) = \boldsymbol{\lambda}_0$ will all be Lévy processes, as well as any linear combination that is the simple constant coefficient jump-diffusion n_x-vector process,

$$\mathbf{X}(t) = \boldsymbol{\mu}_0 t + \sigma_0 \mathbf{W}(t) + \nu_0 \mathbf{P}(t),$$

where $\sigma_0 \in \mathbb{R}^{n_x \times n_w}$ and $\nu_0 \in \mathbb{R}^{n_x \times n_p}$ consistent with IID $\mathbf{W}(t) \in \mathbb{R}^{n_w}$ and IID $\mathbf{P}(t) \in \mathbb{R}^{n_p}$. Adding the compound Poisson process to the combination will be discussed in the sequel.

There are some preliminary definitions that are important for further properties of Lévy processes.

Definition 12.18. ***Infinitely Divisible Distribution:***
A probability distribution $\Phi_{\mathbf{X}}$ on $\mathbb{R}^{n_x}$ is infinitely divisible if for each positive integer n there exists of a set of IID random variable $\mathbf{Y}_j$ for $j = 1:n$ such that the sum

$$\mathbf{S}_n = \sum_{j=1}^{n} \mathbf{Y}_j \stackrel{\text{dist}}{=} \mathbf{X},$$

where $\mathbf{X}$ has distribution $\Phi_{\mathbf{X}}$.

Infinitely Divisibility can be related to the central limit theorem and is closely connected to Lévy processes via compound Poisson processes as follows [59].

Proposition 12.19. ***Lévy processes and Infinitely Divisibility:***
Let $\mathbf{X}(t)$ be a Lévy process for $t \geq 0$ on $\mathbb{R}^{n_x}$, then for every t, $X(t)$ has an infinitely divisible distribution. Conversely, if Φ is an infinitely divisible distribution, then there exists a Lévy process $\mathbf{X}(t)$ with the distribution Φ.

The compound Poisson process is included firmly as a Lévy process by the following result proved in Cont and Tankov [59],

Proposition 12.20. ***Compound Poisson Processes as Lévy Processes:***
The process $CP(t)$ for $t \geq 0$ is a ***compound Poisson process****, i.e.,*

$$CP(t) = \sum_{j=1}^{P(t)} Q_i, \tag{12.54}$$

where $P(t)$ is a ***simple Poisson process*** *with constant rate λ_0 and the Q_j are IID random jump-amplitudes with common distribution $\Phi_Y(y)$ such that λ_0 are independent of the Q_i,*

if and only if

$CP(t)$ is a Lévy process and its sample paths are ***piecewise constant functions****.*

Characteristic Functions and Lévy Characteristic Exponents

Definition 12.21. ***Characteristic Function:***
The characteristic function of a random vector $\mathbf{X}$ on $\mathbb{R}^{n_x}$ is the complex-valued function,

$$\mathcal{C}_{\mathbf{X}}(\mathbf{z}) \equiv \mathrm{E}_{\mathbf{X}}\left[\exp\left(i\mathbf{z}^{\top}\mathbf{X}\right)\right] \tag{12.55}$$

for all $\mathbf{z} \in \mathbb{R}^{n_x}$ and i is the imaginary unit.

Clearly, the characteristic function of a **continuous random variable X** is the **Fourier transform** of the density of $\mathbf{X}$, i.e.,

$$\mathcal{C}_{\mathbf{X}}(\mathbf{z}) = \int_{\mathbb{R}^{n_x}} e^{i\mathbf{z}^{\top}\mathbf{x}} \phi_{\mathbf{X}}(\mathbf{x}) d\mathbf{x},$$

while if X is a **discrete scalar random variable** with distribution given by the countable sequence of probabilities $\pi_k = \text{Prob}[X = k]$, then the characteristic function is the discrete Fourier transform,

$$\mathcal{C}_X(z) = \sum_{k=0}^{\infty} e^{izk} \cdot \pi_k.$$

This is the basic random vector definition, but here the interest will be the same definition when the random vector is a function of time t, i.e., a stochastic process $\mathbf{X}(t)$,

$$\mathcal{C}_{\mathbf{X}(t)}(\mathbf{z}) \equiv \text{E}_{\mathbf{X}(t)}\left[\exp\big(i\mathbf{z}^\top \mathbf{X}(t)\big)\right].$$

One of the most important features of a Lévy process is that the characteristic function has relatively simple form [59, 12]

Proposition 12.22. ***Lévy Characteristic Functions and Exponents:*** *If $\mathbf{X}(t)$ is a Lévy process for $t \geq 0$ on $\mathbb{R}^{n_x}$, then there exist a real-valued continuous function $\eta_{\mathbf{X}(t)}(\mathbf{z})$ of the characteristic vector $\mathbf{z} \in \mathbb{R}^{n_x}$, called the* ***characteristic exponent****, such that*

$$\mathcal{C}_{\mathbf{X}(t)}(\mathbf{z}) = \text{E}_{\mathbf{X}(t)}\left[\exp\big(i\mathbf{z}^\top \mathbf{X}(t)\big)\right] = \exp\big(it\eta_{\mathbf{X}(t)}(\mathbf{z})\big). \tag{12.56}$$

However, for nonstationary problems without the Lévy stationarity condition (12.53), then it would be expected that in general the exponent will not be linear in t,

$$\mathcal{C}_{\mathbf{X}(t)}(\mathbf{z}) = \exp\Big(i\overline{\eta}_{\mathbf{X}(t)}(\mathbf{z}; t)\Big).$$

It is well-known that Fourier transforms, and the characteristic function, is mainly useful for constant coefficients, with few exceptions.

Examples 12.23. ***Characteristic Functions and Exponents of Lévy Processes:***

- ***Standard Wiener Process*** $W(t)$ ***on*** $\mathbb{R}$***:***

$$\mathcal{C}_{W(t)}(z) = \text{E}_{W(t)}\left[e^{izW(t)}\right] = \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} e^{izw} e^{-w^2/(2t)} dw$$
$$= e^{-(tz)^2/(2t)} \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} e^{-(w-itz)^2/(2t)} dw = e^{-tz^2/2},$$

using the completing the square technique, so the Lévy characteristic exponent is

$$\eta_{W(t)}(z) = -\frac{1}{2}z^2. \tag{12.57}$$

- ***IID Wiener Vector Process*** $\mathbf{W}(t)$ ***on*** $\mathbb{R}^{n_w}$
  with $\mathrm{Cov}[\mathbf{W}(t), \mathbf{W}^\top(t)] = tI_{n_w}$***:***

$$\mathcal{C}_{\mathbf{W}(t)}(\mathbf{z}) = \mathrm{E}_{\mathbf{W}(t)}\left[e^{i\mathbf{z}^\top \mathbf{W}(t)}\right] = \prod_{j=1}^{n_w} \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{iz_j w_j} e^{-w_j^2/(2t)} dw_j$$

$$= \prod_{j=1}^{n_w} \mathcal{C}_{W_j(t)}(z_j) = \exp\left(-t \sum_{j=1}^{n_x} z_j^2/2\right) = \exp\left(-t|\mathbf{z}|^2/2\right),$$

  so the Lévy characteristic exponent is

$$\eta_{\mathbf{W}(t)}(\mathbf{z}) = -\frac{1}{2}|\mathbf{z}|^2. \tag{12.58}$$

- ***IID Gaussian Vector Process*** $\mathbf{G}(t) = \boldsymbol{\mu}_0 t + \sigma_0 \mathbf{W}(t)$ ***on*** $\mathbb{R}^{n_x}$
  with $\mathrm{Cov}[\mathbf{W}(t), \mathbf{W}^\top(t)] = tI_{n_w}$***, Constant*** $\boldsymbol{\mu}_0 \in \mathbb{R}^{n_x}$
  and Constant $\sigma_0 \in \mathbb{R}^{n_x \times n_w}$***:***

$$\mathcal{C}_{\mathbf{G}(t)}(\mathbf{z}) = \mathrm{E}_{\mathbf{W}(t)}\left[e^{i\mathbf{z}^\top(\boldsymbol{\mu}_0 t + \sigma_0 \mathbf{W}(t))}\right]$$

$$= e^{i\mathbf{z}^\top \boldsymbol{\mu}_0 t} \prod_{k=1}^{n_w} \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} \exp\left(i \sum_{j=1}^{n_w} z_j \sigma_{0,j,k} w_k\right) e^{-w_k^2/(2t)} dw_k$$

$$= \exp\left(it\mathbf{z}^\top \boldsymbol{\mu}_0 - \sum_{j=1}^{n_x} z_j \sum_{\ell=1}^{n_x} z_\ell \sum_{k=1}^{n_w} \sigma_{0,j,k}\sigma_{0,\ell,k}/2\right)$$

$$= \exp\left(it\mathbf{z}^\top \boldsymbol{\mu}_0 - t\mathbf{z}^\top(\sigma_0\sigma_0^\top)\mathbf{z}\right),$$

  so the Lévy characteristic exponent is

$$\eta_{\mathbf{G}(t)}(\mathbf{z}) = i\mathbf{z}^\top \boldsymbol{\mu}_0 - \frac{1}{2}\mathbf{z}^\top(\sigma_0\sigma_0^\top)\mathbf{z}/2. \tag{12.59}$$

- ***Simple Poisson Process*** $P(t)$ ***on*** $\mathbb{R}$ ***with Constant Jump-Rate*** λ_0***:***

$$\mathcal{C}_{P(t)}(z) = \mathrm{E}_{P(t)}\left[e^{izP(t)}\right] = e^{-\lambda_0 t} \sum_{k=0}^{\infty} \frac{(\lambda_0 t)^k}{k!} e^{izk}$$

$$= e^{-\lambda_0 t} \sum_{k=0}^{\infty} \frac{\left(\lambda_0 t e^{iz}\right)^k}{k!} = e^{-\lambda_0 t + \lambda_0 t e^{iz}} = e^{\lambda_0 t\left(e^{iz} - 1\right)},$$

  so the Lévy characteristic exponent is

$$\eta_{P(t)}(z) = \lambda_0(e^{iz} - 1). \tag{12.60}$$

- ***Centered or Martingale Form of Poisson Process*** $\widetilde{P}(t) \equiv P(t) - \lambda_0 t$ ***on*** $\mathbb{R}$ ***with Constant Jump-Rate*** λ_0***:***

$$\mathcal{C}_{\widetilde{P}(t)}(z) = \mathrm{E}_{P(t)}\left[e^{iz(P(t)-\lambda_0 t)}\right] = e^{-\lambda_0 t}\sum_{k=0}^{\infty}\frac{(\lambda_0 t)^k}{k!}e^{iz(k-\lambda_0 t)}$$

$$= e^{-\lambda_0 t(1+iz)}\mathcal{C}_{P(t)}(z) = e^{\lambda_0 t(e^{iz}-1-iz)},$$

so the Lévy characteristic exponent is

$$\eta_{\widetilde{P}(t)}(z) = \lambda_0(e^{iz} - 1 - iz). \tag{12.61}$$

- ***Simple Poisson Vector Process*** $\mathbf{P}(t)$ ***on*** $\mathbb{R}^{n_p}$ ***with Independent Components and Constant Jump-Rate Vector*** $\boldsymbol{\lambda}_0 = [\lambda_{0,j}]_{n_p\times 1}$***:***

$$\mathcal{C}_{\mathbf{P}(t)}(\mathbf{z}) = \mathrm{E}_{\mathbf{P}(t)}\left[\exp\left(i\mathbf{z}^\top\mathbf{P}(t)\right)\right] = \prod_{j=1}^{n_p} e^{-\lambda_{0,j}t}\sum_{k_j=0}^{\infty}\frac{(\lambda_{0,j}t)^{k_j}}{k_j!}e^{iz_j k_j}$$

$$= \prod_{j=1}^{n_p}\exp(\lambda_{0,j}t\,(\exp(iz_j)-1)) = \exp\left(t\sum_{j=1}^{n_p}\lambda_{0,j}\,(\exp(iz_j)-1)\right)$$

$$= \exp\left(tn_p\left(\overline{\lambda_0\exp(iz)} - \overline{\lambda_0}\right)\right),$$

where $\overline{\lambda_0} \equiv \sum_{j=1}^{n_p}\lambda_{0,j}/n_p$ *and* $\overline{\lambda_0\exp(iz)} \equiv \sum_{j=1}^{n_p}\lambda_{0,j}\exp(iz_j)/n_p$*, so the Lévy characteristic exponent is*

$$\eta_{\mathbf{P}(t)}(\mathbf{z}) = n_p\left(\overline{\lambda_0\exp(iz)} - \overline{\lambda_0}\right). \tag{12.62}$$

- ***Simple Compound Poisson Process*** $CP(t) = \sum_{\ell=1}^{P(t)} Q_\ell$ ***on*** $\mathbb{R}$ ***with Constant Jump-Rate*** λ_0 ***and IID Jump-Amplitudes*** Q_ℓ ***with Distribution*** $\Phi_Q(q)$***:***

$$\mathcal{C}_{CP(t)}(z) = \mathrm{E}_{P(t),Q}\left[e^{izX(t)}\right] = e^{-\lambda_0 t}\sum_{k=0}^{\infty}\frac{(\lambda_0 t)^k}{k!}\mathrm{E}_Q\left[\exp\left(iz\sum_{\ell=1}^{k}Q_\ell\right)\right]$$

$$= e^{-\lambda_0 t}\sum_{k=0}^{\infty}\frac{(\lambda_0 t)^k}{k!}\prod_{\ell=1}^{k}\mathrm{E}_Q\left[\exp(izQ_\ell)\right]$$

$$= e^{-\lambda_0 t}\sum_{k=0}^{\infty}\frac{(\lambda_0 t)^k}{k!}\mathrm{E}_Q^k\left[\exp(izQ)\right] = \exp\left(\lambda_0 t\left(\mathrm{E}_Q\left[\exp(izQ)\right]-1\right)\right),$$

using the iterated conditional expectation technique and IID, so the Lévy characteristic exponent, substituting $\mathrm{E}_Q\left[\exp(izQ)\right] = \mathcal{C}_Q(z)$*, is*

$$\eta_{CP(t)}(z) = \lambda_0(\mathcal{C}_Q(z) - 1) \tag{12.63}$$

and the simple Poisson process exponent is recovered if $Q_\ell \overset{\text{w.p.o.}}{=} 1 \ \ \forall \ell \geq 1$.

- ***Vector Compound Poisson Process*** $\mathbf{CP}(t) = \sum_{\ell=1}^{P(t)} \mathbf{Q}_\ell$ ***on*** $\mathbb{R}^{n_x}$ ***with Constant Jump-Rate*** λ_0 ***and IID Vector Jump-Amplitudes*** $\mathbf{Q}_\ell$ ***with Distribution*** $\Phi_{\mathbf{Q}}(\mathbf{q})$***:*** *Note that the* $\mathbf{Q}_\ell$ *are IID as vectors not necessarily as components, thus,*

$$\begin{aligned}
\mathcal{C}_{\mathbf{CP}(t)}(\mathbf{z}) = \mathrm{E}_{P(t),\mathbf{Q}}\left[e^{i\mathbf{z}^\top \mathbf{CP}(t)}\right] &= e^{-\lambda_0 t}\sum_{k=0}^{\infty}\frac{(\lambda_0 t)^k}{k!}\mathrm{E}_{\mathbf{Q}}\left[\exp\left(i\mathbf{z}^\top\sum_{\ell=1}^{k}\mathbf{Q}_\ell\right)\right] \\
&= e^{-\lambda_0 t}\sum_{k=0}^{\infty}\frac{(\lambda_0 t)^k}{k!}\prod_{\ell=1}^{k}\mathrm{E}_{\mathbf{Q}}\left[\exp\left(i\mathbf{z}^\top\mathbf{Q}_\ell\right)\right] \\
&= e^{-\lambda_0 t}\sum_{k=0}^{\infty}\frac{(\lambda_0 t)^k}{k!}\mathrm{E}_{\mathbf{Q}}^{k}\left[\exp\left(i\mathbf{z}^\top\mathbf{Q}_\ell\right)\right] \\
&= \exp\left(\lambda_0 t\left(\mathrm{E}_{\mathbf{Q}}\left[\exp\left(i\mathbf{z}^\top\mathbf{Q}_\ell\right)\right]-1\right)\right),
\end{aligned}$$

using the iterated conditional expectation technique and IID again, so the Lévy characteristic exponent, substituting $\mathrm{E}_{\mathbf{Q}}\left[\exp(i\mathbf{z}^\top\mathbf{Q})\right] = \mathcal{C}_{\mathbf{Q}}(\mathbf{z})$*, is*

$$\eta_{\mathbf{CP}(t)}(z) = \lambda_0(\mathcal{C}_{\mathbf{Q}}(\mathbf{z})-1). \tag{12.64}$$

Lévy-Klintchine Jump-Diffusion Formula

In these examples, the ingredients for the fundamental theorem of the **Lévy-Klintchine representation formula** specialized to jump-diffusion processes has been derived, based on the vector Gaussian process exponent result in (12.59) and the vector compound Poisson process exponent result in (12.64).

Theorem 12.24. ***Lévy-Klintchine Formula for Jump-Diffusion Processes:*** *Let* $\mathbf{X}(t)$ *be the jump-diffusion process on* $\mathbb{R}^{n_x}$ *for* $t \geq 0$*,*

$$\mathbf{X}(t) = \mathbf{X}(0) + \boldsymbol{\mu}_0 t + \sigma_0 \mathbf{W}(t) + \sum_{\ell=1}^{P(t)}\mathbf{Q}_\ell, \tag{12.65}$$

with ***Lévy characteristic triplet*** $(\sigma_0\sigma_0^\top, \lambda_0\Phi_{\mathbf{Q}}(d\mathbf{q})dt, \boldsymbol{\mu}_0)$*, where* $\boldsymbol{\mu}_0 \in \mathbb{R}^{n_x}$ *is a constant,* $\sigma_0 \in \mathbb{R}^{n_x \times n_w}$ *is a constant,* $\mathbf{W}(t) \in \mathbb{R}^{n_w}$ *is a vector Wiener process,* $\mathbf{P}(t) \in \mathbb{R}$ *is a simple Poisson process with constant and finite jump-rate* $\lambda_0 \in \mathbb{R}$ *and compounded with IID vector jump-amplitudes* $\mathbf{Q}_\ell \in \mathbb{R}^{n_x}$ *with distribution* $\Phi_{\mathbf{Q}}(\mathbf{q})$*. The random triplet* $(\mathbf{W}(t), P(t), Q)$ *are independent variables, except that the jump-amplitude* Q *requires the existence of a jump of the Poisson process.*

Then, the characteristic function with $\mathbf{z} \in \mathbb{R}^{n_x}$ *for the initial condition translated process*

$$\mathbf{Y}(t) \equiv \mathbf{X}(t) - \mathbf{X}(0) \tag{12.66}$$

is

$$\mathcal{C}_{\mathbf{Y}(t)}(\mathbf{z}) = \mathrm{E}_{\mathbf{Y}(t)}\big[\exp(i\mathbf{z}^\top \mathbf{Y}(t)\big] = \exp\left(t\eta_{\mathbf{Y}(t)}(\mathbf{z})\right),$$

where the Lévy characteristic exponent is

$$\eta_{\mathbf{Y}(t)}(\mathbf{z}) = i\boldsymbol{\mu}_0 t - \frac{1}{2}\mathbf{z}^\top \sigma_0 \sigma_0^\top \mathbf{z} + \lambda_0 \int_{\mathbb{R}^{n_x}} \left(\exp(i\mathbf{z}^\top \mathbf{q}) - 1\right) \phi_{\mathbf{Q}}(\mathbf{q}) d\mathbf{q}. \quad (12.67)$$

Except for the technical details, the Lévy characteristic exponent result (12.67) follows from (12.59) for $\mathbf{G}(t)$ and from (12.64) for $\mathbf{CP}(t)$ by the independence properties between $\mathbf{G}(t)$ and $(P(t), Q)$ and by iterative conditional expectation between $P(t)$ and Q that is conditioned on the existence of a jump as for (12.64). Thus,

$$\begin{aligned}
\mathcal{C}_{\mathbf{Y}(t)}(\mathbf{z}) &= \mathrm{E}_{\mathbf{Y}(t)}\big[\exp\big(i\mathbf{z}^\top \mathbf{Y}(t)\big)\big] \\
&= \mathrm{E}_{\mathbf{W}(t)}\big[\exp\big(i\mathbf{z}^\top \mathbf{G}(t)\big)\big] \cdot \mathrm{E}_{P(t),\mathbf{Q}}\big[\exp\big(i\mathbf{z}^\top \mathbf{CP}(t)\big)\big] \\
&= \mathcal{C}_{\mathbf{G}(t)}(\mathbf{z}) \cdot \mathcal{C}_{\mathbf{P}(t)}(\mathbf{z}) \\
&= \exp\big(t\eta_{\mathbf{G}(t)}(\mathbf{z})\big) \cdot \exp\big(t\eta_{\mathbf{CP}(t)}(\mathbf{z})\big) \\
&= \exp\big(t\big(\eta_{\mathbf{G}(t)}(\mathbf{z}) + \eta_{\mathbf{CP}(t)}(\mathbf{z})\big)\big)
\end{aligned}$$

so substituting (12.59) and (12.64) and expanding the expectations leads directly to the main result (12.67). It should be noted that embedded in this derivation is the semi-group property [12, 59] of the characteristic function in the case of constant coefficients.

In the case to the geometric or linear jump-diffusion process (5.42) with constant rate coefficients for $X(t) \in \mathbb{R}$ with SDE,

$$dX(t) = X(t)\left(\mu_0 dt + \sigma_0 dW(t) + \sum_{k=1}^{dP(t)} \left(e^{Q_k} - 1\right)\right),$$

the solution is exponential via a logarithmic change of variable technique,

$$X(t) = X(0)\exp\left((\mu_0 - \sigma_0^2/2)t + \sigma_0 W(t) + \sum_{k=1}^{P(t)} Q_k\right), \quad (12.68)$$

with $X(0) > 0$, is obviously not a Lévy process due to the exponential time-dependence, without further transformation:

Corollary 12.25. ***Lévy-Klintchine Transformed Geometric Jump-Diffusion Formula:***

Assuming the hypotheses of Th. 12.24, except that, $n_x = 1$, $n_w = 1$ and that the Lévy characteristic triplet is $(\sigma_0^2, \lambda_0 \Phi_Q(dq)dt, \mu_0 - \sigma_0^2)$, then the characteristic function with $z \in \mathbb{R}$ of the the logarithmic-translated process $Y(t)$.,

$$Y(t) \equiv \ln(X(t)/X(0)), \quad (12.69)$$

corresponding to the geometric process (12.68), is

$$\mathcal{C}_{Y(t)}(z) = \mathrm{E}_{Y(t)}[\exp(izY(t)] = \exp\left(t\eta_{_{Y(t)}}(z)\right),$$

where the Lévy characteristic exponent is

$$\eta_{_{Y(t)}}(z) = i(\mu_0 - \sigma_0^2/2)t - \frac{1}{2}\sigma_0^2 z^2 + \lambda_0 \int_{\mathbb{R}} (\exp(izq) - 1)\, \phi_Q(q) dq. \quad (12.70)$$

Lévy-Klintchine Lévy Process Formula including Infinite Rate Processes

So far the jump-rate λ_0 has been assumed to be constant and either explicitly or implicitly finite in this Subsection on Lévy processes. However, the infinite jump-rates is a distinguishing feature of Lévy processes, so that, in general, it is not valid to write the jump-rate symbol λ_0 in Lévy process formulas. Instead, it is necessary to refer to the number of jumps rather than to the jump-rate.

Recall the definition (B.180) on page B64 of the **jump function** of a process:

$$[\mathbf{X}](t) \equiv \mathbf{X}(t) - \mathbf{X}(t-),$$

written here for RCLL vector processes (caution: in some of the literature $\Delta\mathbf{X}(t)$ is used, but can be confused with the analytic or numerical time increment, $\Delta\mathbf{X}(t) \equiv \mathbf{X}(t + \Delta t) - \mathbf{X}(t)$). At points where $\mathbf{X}(t)$ is continuous, then $[\mathbf{X}](t) = 0$.

Definition 12.26. ***Number of Jumps of a Process, Poisson Random Measure and Lévy Measure:*** *The number of jumps in the open set $\mathcal{S}$, assuming a bounded number of jumps and excluding zero jumps ($0 \notin \mathcal{S}$) on the interval $(0, t]$, is*

$$\mathcal{P}((0,t], \mathcal{S}) = \sum_{s \in (0,t]} \mathbf{1}_{[\mathbf{X}](s) \in \mathcal{S}}. \quad (12.71)$$

Here, $\mathcal{P}((0,t], \mathcal{S})$ is the Poisson random or jump measure [223]. The differential form is denoted by $\mathcal{P}(dt, d\mathbf{q}) = \mathcal{P}((t, t+dt, (\mathbf{q}, \mathbf{q}+d\mathbf{q}])$, as previous used in Chapt. 5. An alternate form [232] uses a sequence of stopping or jump times,

$$T_{k+1}(\mathcal{S}) = \inf\{t \mid t > T_k(\mathcal{S}), [X](t) \in \mathcal{S}\};\ T_0(\mathcal{S}) \equiv 0,$$

such that

$$\mathcal{P}((0,t], \mathcal{S}) = \sum_{k=1}^{\infty} \mathbf{1}_{T_k(\mathcal{S}) \le t}.$$

The zero-mean (centered or Martingale) form is denoted by

$$\widetilde{\mathcal{P}}(dt, \mathbf{q}) = \mathcal{P}(dt, d\mathbf{q}) - \nu^{(L)}(\mathbf{dq})dt, \quad (12.72)$$

*where now $\nu^{(L)}(\mathbf{dq})dt = \mathrm{E}[\mathcal{P}(dt, d\mathbf{q})]$ and $\nu^{(L)}$ is called the **Lévy measure** in general .*

A fundamental tool for separating out the large jumps in the presence of infinite jump-rates is the following decomposition after the concise form of Øksendal and Sulem [223]:

Theorem 12.27. ***Lévy-Itô Decompsition:*** *Let $0 \leq R < \infty$ be a jump-amplitude cutoff, then a Lévy process $\mathbf{X}^{(L)}(t)$ on $\mathbb{R}^{n_x}$ has the decomposition,*

$$\mathbf{X}^{(L)}(t) = \widetilde{\boldsymbol{\mu}}_{0,R} t + \sigma_0 \mathbf{W}(t) + \int_{|\mathbf{q}|<R} \mathbf{q}\widetilde{\mathcal{P}}(t, d\mathbf{q}) + \int_{|\mathbf{q}|\geq R} \mathbf{q}\mathcal{P}(t, d\mathbf{q}), \quad (12.73)$$

where $\mathbf{W}(t) \in \mathbb{R}^{n_w}$ is an independent vector Wiener process, $\widetilde{\boldsymbol{\mu}}_{0,R} \in \mathbb{R}^{n_x}$ is a constant adjusted with R from the original drift $\boldsymbol{\mu}_0 \in \mathbb{R}^{n_x}$, $\sigma_0 \in \mathbb{R}^{n_x \times n_w}$ is a constant .

In particular, the Lévy-Itô decomposition states that the Lévy process is, as is the jump-diffusion, decomposable into a continuous process and a discontinuous process,

$$\begin{aligned}
\mathbf{X}^{(L)}(t) =& \ \mathbf{X}^{(cont)}(t) + \mathbf{X}^{(discont)}(t);\\
\mathbf{X}^{(cont)}(t)) =& \ \widetilde{\boldsymbol{\mu}}_{0,R} t + \sigma_0 \mathbf{W}(t);\\
\mathbf{X}^{(discont)}(t)) =& \ \mathbf{X}^{(L)}(t) - \mathbf{X}^{(cont)}(t).
\end{aligned}$$

One consequence of this Lévy-Itô decomposition is another fundamental result [223, 232]:

Theorem 12.28. ***Lévy-Klintchine Representation Formula for Lévy Processes:*** *Let $\mathbf{X}^{(L)}(t)$ be a Lévy process for $t \geq 0$ with Lévy measure $\nu^{(L)}$ on $\mathbb{R}^{n_x}$, given constants $\widetilde{\boldsymbol{\mu}}_{0,R} \in \mathbb{R}^{n_x}$ and $\sigma_0 \in \mathbb{R}^{n_x \times n_w}$, then the jump-count satisfies*

$$\int_{\mathbb{R}^{n_x}} \min(|\mathbf{q}|^2, R)\nu^{(L)}(d\mathbf{q}) < \infty$$

and the characteristic function on $\mathbf{z} \in \mathbb{R}^{n_x}$ for $\mathbf{X}(t) = \mathbf{X}^{(L)}(t)$ is

$$\mathcal{C}_{\mathbf{X}(t)}(\mathbf{z}) = \mathrm{E}_{\mathbf{X}(t)}\left[\exp\left(i\mathbf{z}^\top \mathbf{X}(t)\right)\right] = \exp\left(t\eta_{\mathbf{X}(t)}(\mathbf{z})\right),$$

where the Lévy characteristic exponent is

$$\begin{aligned}
\eta_{\mathbf{X}(t)}(\mathbf{z}) =& \ i\widetilde{\boldsymbol{\mu}}_{0,R}\ t - \tfrac{1}{2}\mathbf{z}^\top \sigma_0 \sigma_0^\top \mathbf{z}\\
&+ \textstyle\int_{|\mathbf{q}|<R} \left(\exp\left(i\mathbf{z}^\top \mathbf{q}\right) - 1 - i\mathbf{z}^\top \mathbf{q}\right) \nu^{(L)}(d\mathbf{q})\\
&+ \textstyle\int_{|\mathbf{q}|\geq R} \left(\exp\left(i\mathbf{z}^\top \mathbf{q}\right) - 1\right) \nu^{(L)}(d\mathbf{q}).
\end{aligned} \quad (12.74)$$

Conversely, given constants $\widetilde{\boldsymbol{\mu}}_{0,R} \in \mathbb{R}^{n_x}$ and $\sigma_0 \in \mathbb{R}^{n_x \times n_w}$, along with the Lévy measure $\nu^{(L)}$ on $\mathbb{R}^{n_x}$ such that the jump-count satisfies

$$\int_{\mathbb{R}^{n_x}} \min(|\mathbf{q}|^2, R)\nu^{(L)}(d\mathbf{q}) < \infty,$$

then there exists a Lévy process $\mathbf{X}^{(L)}$ *that is unique in distribution such that the Lévy characteristic is (12.74) for* $\mathbf{z} \in \mathbb{R}^{n_x}$.

Note that the extra linear term $i\mathbf{z}^\top \mathbf{q}$ in the first or inner integral of (12.74) is related to the zero-mean Poisson $\widetilde{P}(t)$ form iz found in (12.61) but not in (12.60) for $P(t)$.

Although jump processes time-dependent coefficients, like drift and volatility coefficients, do not strictly satisfy the stationary increment condition (12.52) for a Lévy process, Øksendal and Sulem [223] define **Lévy-driven processes** which satisfy the Lévy-Itô decomposition formula (12.73), but not the constant coefficient condition. For example, analogous to the **Wiener-driven Itô process** (12.47), there is the **Lévy-driven Itô-Lévy process** [223, Th. 1.14, p. 6] on $\mathbb{R}$ with time-random coefficients,

$$\begin{aligned} dX(t) = \; & \widetilde{\mu}_{0,R}(t;\omega)dt + \sigma_0(t;\omega)dW(t) \\ & + \textstyle\int_{|q|<R} h(t,q;\omega)\widetilde{\mathcal{P}}(dt,dq) + \int_{|q|\geq R} h(t,q;\omega)\mathcal{P}(dt,dq), \end{aligned} \tag{12.75}$$

for some $R \in [0,\infty)$, $(\widetilde{\mu}_{0,R}(t;\omega), \sigma_0(t;\omega), h(t,q\omega)$ are integrable function and ω is some background random variable.

The **Lévy-driven geometric Lévy process** [223, Example 1.15, p. 7] is similarly defined,

$$\begin{aligned} dX(t) = \; & X(t)\Big(\widetilde{\mu}_{0,R}(t;\omega)dt + \sigma_0(t;\omega)dW(t) \\ & + \textstyle\int_{|q|<R} h(t,q;\omega)\widetilde{\mathcal{P}}(dt,dq) + \int_{|q|\geq R} h(t,q;\omega)\mathcal{P}(dt,dq)\Big), \end{aligned} \tag{12.76}$$

where, in addition, the jump-amplitude $h(t,q;\omega) \geq -1$ to preserve positivity assuming $X(0) > 0$, with more potential uses in financial applications.

In general, these processes are special cases of what Øksendal and Sulem [223, Th. 1.19,p. 10] call **Lévy diffusions** governed by **Lévy stochastic differential equations**,

$$\begin{aligned} d\mathbf{X}(t) = \; & \widetilde{\mu}(t,\mathbf{X}(t))dt + \sigma(t,\mathbf{X}(t))d\mathbf{W}(t) \\ & + \textstyle\int_{|q|<R} h(t,q;\omega)\widetilde{\mathcal{P}}(dt,dq) + \int_{|q|\geq R} h(t,\mathbf{X}(t),q)\mathcal{P}(dt,dq), \end{aligned} \tag{12.77}$$

where $0 \leq t \leq T$, $\mathbf{X} \in \mathbb{R}^{n_x}$, $\widetilde{\mu} \in \mathbb{R}^{n_x}$, $\mathbf{W} \in \mathbb{R}^{n_w}$, $\sigma \in \mathbb{R}^{n_x \times n_w}$, $\mathcal{P} \in \mathbb{R}^{n_p}$, $\mathbf{Q} \in \mathbb{R}^{n_x}$ and $h \in \mathbb{R}^{n_x \times n_p}$, subject to the usual linear growth and Lipschitz continuity conditions.

For many other Lévy process models, including models which push the limits of the assumptions here, see Applebaum [12, Subsect. 5.4.7, p. 286ff]

Concluding this subsection like the last, the size of the Lévy processes is compared to that of jump-diffusions. According to the strict Lévy process definition leading to a restriction to constant coefficients,

$$\left\{ \begin{array}{c} \textbf{\textit{constant coefficient}} \\ \textbf{\textit{jump-diffusion processes}} \end{array} \right\} \subset \textbf{\textit{Lévy processes}}, \tag{12.78}$$

since ordinarily jump-diffusions based upon Poisson processes do not allow for infinite jump-rates on $[0, t]$. However, if the infinite jump activity is controlled for, then

$$\left\{ \begin{array}{c} \textbf{\textit{finite jump-rate}} \\ \textbf{\textit{Lévy processes}} \end{array} \right\} \subset \textbf{\textit{jump-diffusion processes}}, \tag{12.79}$$

since jump-diffusions in general include variable coefficients and nonlinear terms.

If the comparison is made to the Lévy-driven processes discussed by Øksendal and Sulem [223] and summarized here, then

$$\{\textbf{\textit{jump-diffusion processes}}\} \subset \textbf{\textit{Lévy-driven processes}}, \tag{12.80}$$

due to the inclusion of infinite jump-rates with nonlinear and time-dependent coefficients in Lévy-driven processes.

12.4 Exercises

1. Similar to the Examples 12.10, find the martingale properties as a function of the additional drift for the **geometric jump diffusion** problem with constant coefficients,

$$dX(t) = X(t) \left(\mu_0 dt + \sigma_0 dW(t) + \nu_0 \sum_{i=1}^{dP(t)} \left(e^{Q_i} - 1 \right) \right),$$

where again the marks are IID with mean μ_Q and variance σ_Q^2.

Suggested References for Further Reading

- Applebaum, 2004 [12].
- Bain, 2006 [17].
- Baxter and Rennie, 1996 [22].
- Billingsley, 1986 Billingsley86.
- Bingham and Kiesel, 2004 [33]
- Bossaerts, 2002 [41].
- Brémaud, 1981 [43].
- Cont and Tankov, 2004 [59].
- Cyganowski, Kloeden and Ombach, 2002 [66].
- Doob, 1953 [69].
- Duffie, 1992 [74].

- Gihman and Skorohod, 1972 [94].
- Glasserman, 2003 [96].
- Harrison and Pliska, 1981, [132].
- Harrison and Pliska, 1983, [133].
- Heath and Schweizer, 2000 [135].
- Hull, 2000 [147].
- Itô, 1951 [149].
- Karatzas and Shreve, 1998 [160].
- Karlin and Taylor, 1981 [162].
- Klebaner, 1998 [164].
- Mikosch, 1998 [209].
- Neftci, 2000 [217].
- Øksendal, 1998 [222].
- Øksendal and Sulem, 2005 [223].
- Pliska, 1997 [225].
- Protter, 2004 [232].
- Runggaldier, 2003 [239].
- Rogers and Williams, 2000 [236].
- Shreve, 2004 [248].
- Steele, 2001 [256].
- Yong and Zhou, 1999 [288].

Bibliography

[1] M. L. Abell and J. P. Braselton, *The Maple V Handbook*, Academic Press, New York, NY, 1994.

[2] M. Abramowiitz and I. A. Stegun, (eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Applied Mathematics Series 55, National Bureau of Standards, Washington, DC, 1964.

[3] R. M. Abu-Saris and F. B. Hanson, *Computational Suboptimal Filter for a Class of Wiener-Poisson Driven Stochastic Process*, Dynamics and Control, vol. 7, no. 3, 1997, pp. 279-292.

[4] N. Ahmed and K. Teo, *Optimal Control of Distributed Parameter Systems*, North Holland, New York, NY, 1981.

[5] Y. Aït-Sahalia, *Disentangling Diffusion from Jumps*, J. Financial Economics, vol. 74, 2004, pp. 487–528.

[6] T. G. Andersen, L. Benzoni and J. Lund, *An Empirical Investigation of Continuous-Time Equity Return Models*, J. Finance, vol. 57, no. 3, 2002, pp. 1239-1284.

[7] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

[8] B. D. O. Anderson and J. B. Moore, *Optimal Control: Linear Quadratic Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

[9] P. Andersen and J. G. Sutinen, *Stochastic Bioeconomics: A Review of Basic Methods and Results*, Marine Res. Econ., vol. 1, 1982, pp. 1-10.

[10] H. L. Anderson, *Metropolis, Monte Carlo Method, and the MANIAC*, Los Alamos Science, Fall 1986, pp. 96-107.

[11] C. A. Aourir, D. Okuyama, C. Lott and C. Eglinton, *Exchanges - Circuit Breakers, Curbs, and Other Trading Restrictions*, `http://invest-faq.com/articles/exch-circuit-brkr.html` .

[12] D. Applebaum, *Lévy Processes and Stochastic Calculus*, Cambridge University Press, Cambridge, UK, 2004.

[13] L. Arnold, *Stochastic Equations: Theory and Applications*, John Wiley, New York, NY, 1974.

[14] Y. Ashkenazy, J. M. Hausdorff, P. C. Ivanov and H. E. Stanley, *A Stochastic Model of Human Gait Dynamics*, Physica A, vol. 316, 2002, pp. 662-670.

[15] M. Athans and P. L. Falb, *Optimal Control: An Introduction to the Theory and Its Applications*, McGraw-Hill Book Co., New York, NY, 1966.

[16] L. Bachelier, *Théorie de la Spéculation*, Annales de l'Ecole Normale Supérieure, vol. 17, 1900, pp. 21-86. English translation by A. J. Boness in *The Random Character of Stock Market Prices*, P. H. Cootner (Editor), MIT Press, Cambridge, MA, 1967, pp. 17-78.

[17] A. Bain, *Stochastic Calculus*, `http://www.chiark.greenend.org.uk/~alanb/stoc-calc.pdf` .

[18] C. A. Ball and W. N. Torous, *On Jumps in Common Stock Prices and Their Impact on Call Option Prices*, J. Finance, vol. 40, 1985, pp. 155-173.

[19] M. S. Bartlett, *Stochastic Equations: Theory and Applications*, 3rd ed., Cambridge University Press, Cambridge, UK, 1978.

[20] O. E.. Barndorff-Nielsen and N. Shepherd, *Non-Gaussian Ornstein-Uhlenb6eck-Based Models and Some of their Uses in Financial Economics*, ser. B, vol. 3, part. 2, 2001, pp. 167-241.

[21] T. Basar, *Twenty-five Seminal Papers in Control*, IEEE Control Systems Magazine, vol. 20, no. 1, February 2000, pp. 69-70.

[22] M. Baxter and A. Rennie, *Financial Calculus: An Introduction to Derivative Pricing*, Cambridge University Press, Cambridge, UK, 1996.

[23] I. Beichl and F. Sullivan, *The Metropolis Algorithm*, Computng in Sci. & Engineering, vol. 2, no. 1, 2000, pp. 65-69.

[24] D. J. Bell and D. H. Jacobson, *Singular Optimal Control Problems*, Academic Press, New York, NY, 1994.

[25] R. E. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ, 1957.

[26] R. E. Bellman, *Adaptive Control Processes: A Guided Tour*, Princeton University Press, Princeton, NJ, 1961.

[27] R. Bellman and R. Kalaba, *Selected Papers on Mathematical Trends in Control Theory*, Dover Publications, New York, NY, 1964.

[28] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill Book Co., New York, NY, 1978.

[29] D. S. Bernstein, *Feedback Control: An Invisible Thread in the History of Technology*, IEEE Control Systems Magazine, vol. 22, no. 2, April 2002, pp. 53-68.

[30] J. T. Betts, *Practical Methods for Optimal Control Using Nonlinear Programming*, SIAM, Philadelphia, PA, 2001.

[31] A. T. Bharucha-Reid, *Elements of the Theory of Markov Processes and their Applications*, McGraw-Hill Book Co., New York, NY, 1960.

[32] P. Billingsley, *Probability and Measures*, 2nd Edition, John Wiley, New York, NY, 1986.

[33] N. H. Bingham and R. Kiesel, *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*, Springer-Verlag, New York, NY, 2004.

[34] F. Black and M. Scholes, *The Pricing of Options and Corporate Liabilities*, J. Political Economy, vol. 81, 1973, pp. 637-659.

[35] F. Black, *How We Came Up with the Option Formula*, J. Portfolio Mgmt., vol. 15, Winter 1989, pp. 4-8.

[36] A. Boker, C. J. Haberman, L. Girling, R. P.Guzman, G. Louridas, J. R. Tanner, M. Cheang. M. Math, B. W. Maycher, D. D. Bell and G. J. Doak, *Variable Ventilation Improves Perioperative Lung Function in Patients Undergoing Abdominal Aortic Aneurysmectomy*, Anesthesiology, vol. 100, no. 3, 2004, pp. 608-616.

[37] G. E. P. Box and M. E. Muller, *A Note on the Generation of Random Normal Deviates*, Ann. Math. Stat., vol. 29, 1958, pp. 610-611.

[38] P. Boyle, *Options: A Monte Carlo Approach*, J. Financial Economics, vol. 4, 1977, pp. 323-338.

[39] P. Boyle, M. Broadie and P. Glasserman, *Monte Carlo Methods for Security Pricing*, J. Economic Dynamics and Control, vol. 21, 1997, pp. 1267-1321.

[40] G. A. Bliss, *Lectures on the Calculus of Variations*, University Chicago Press, Chicago, IL, 1946.

[41] P. Bossaerts, *The Paradox of Asset Pricing*, Princeton University Press, Princeton, NJ, 2002.

[42] P. W. Bridgeman, *Dimensional Analysis*, Yale University Press, New Haven, CT, 1963.

[43] P. Brémaud, *Point Processes and Queues: Martingale Dynamics*, Springer-Verlag, New York, NY, 1981.

[44] A. E. Bryson and Y. C. Ho, *Applied Optimal Control*, John Wiley, New York, NY, 1975.

[45] T. G. Buchman, *Nonlinear Dynamics, Complex Systems, and the Pathobiology of Critical Illness*, Curr. Opin. Crit. Care, vol. 10, no. 5, 2004, pp. 378-382.

[46] P. Carr, H. Geman, D. B. Madan and M. Yor, *Stochastic Volatility for Lévy Processes*, Math. Fin., vol. 13, no. 3, 3003, pp. 345-382.

[47] P. Carr and D. B. Madan, *Option Valuation Using the Fast Fourier Transform*, J. Comp. Fin., vol. 2, 1999, pp. 61-73.

[48] S. P. Chakrabarty and F. B. Hanson, *Optimal Control of Drug Delivery to Brain Tumors for a Distributed Parameters Model*, in Proc. 2005 American Control Conference, 2005, pp. 973-978.

[49] S. P. Chakrabarty and F. B. Hanson, *Optimal Control of Drug Delivery to Brain Tumors for a PDE Driven Model Using the Galerkin Finite Element Method*, in Proc. 44nd IEEE Conference on Decision and Control and European Control Conference 2005, 2005, pp. 1613-1618.

[50] S. P. Chakrabarty and F. B. Hanson, *Cancer Drug Delivery in Three Dimensions For a Distributed Parameter Control Model Using Finite Elements*, in Proc. 45th IEEE Conference on Decision, 23 February 2006, pp. 1-8, submitted for review; `URL:http://www.math.uic.edu/~hanson/pub/CDC2006/cdc06spcfhweb.pdf` .

[51] G. Chichilnisky, *Fischer Black: The Mathematics of Uncertainty*, Notices of the AMS, vol. 43 (3), 1996, pp. 319-322.

[52] S.-L. Chung and F. B. Hanson, *Optimization Techniques for Stochastic Dynamic Programming*, in Proc. 29th IEEE Conference on Decision and Control, vol. 4, 1990, pp. 2450-2455.

[53] S.-L. Chung and F. B. Hanson, *Parallel Optimizations for Computational Stochastic Dynamic Programming*, in Proc. 1990 International Conference on Parallel Processing, vol. 3: Algorithms and Applications, P.-C. Yew, ed., 1990, pp. 254-260.

[54] S.-L. Chung, F. B. Hanson and H. H. Xu, *Parallel Stochastic Dynamic Programming: Finite Element Methods*, Lin. Alg. Applic., vol. 172, 1992, pp. 197-218.

[55] E. Çinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[56] C. W. Clark, *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*, 1st and 2nd Editions, John Wiley, New York, NY, 1976 and 1990.

[57] C. W. Clark and R. Lamberson, *An Economic History and Analysis of Pelagic Whaling*, Marine Policy, vol. 6, 1982, pp. 103-120.

[58] J. E. Cohen, *Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better*, PLoS Biology, vol. 2, issue 12, 2004, p. 2017-2023.

[59] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall/CRC, Boca Raton, FL, 2004.

[60] E. T. Copson, *Asymptotic Expansions*, Cambridge University Press, Cambridge, UK, 1965.

[61] J. M. Courtault, Y. Kabanov, B. Bru, P. Crépel, I. Lebon and A. L. Marchand, *Louis Bachelier on the centenary of Théorie De La Spéculation*, Mathematical Finance, vol. 10, No. 3, 2000, pp. 341-353.

[62] D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes*, Chapman and Hall, London, UK, 1965.

[63] J. C. Cox and M. Rubinstein, *Options Markets*, Prentice-Hall, Englewood Cliffs, NJ, 1985.

[64] S. Cyganowski, L. Grüne and P. Kloeden, *Maple for Jump-Diffusion Stochastic Differential Equations in Finance*, Programming Languages and Systems in Computational Economics and Finance, S. S. Nielsen (Editor), Kluwer Academic Publishers, Amsterdam, 2002, pp. 233-269; `URL:http://www.uni-bayreuth.de/departments/math/~lgruene/papers/jumpfin.html` .

[65] S. Cyganowski and P. Kloeden, *Maple Schemes for Jump-Diffusion Stochastic Differential Equations*, Proc. 16th IMACS World Congress, Lausanne 2000, M. Deville, R. Owens (Editors), Dept. Computer Science, Rutgers University, 2000, CD-ROM Paper 216-9, pp. 1-16; `http://www.math.uni-frankfurt.de/~numerik/maplestoch/jumpdiff.pdf` .

[66] S. Cyganowski, P. Kloeden and J. Ombach, *Elementary Probability to Stochastic Differential Equations with Maple*, Springer-Verlag, New York, NY, 2002.

[67] B. N. Datta, *Numerical Linear Algebra and Applications*, Brooks/Cole, New York, NY, 1995.

[68] J. Dongarra and F. Sullivan, *Guest Editor's Introduction: The Top Ten Algorithms*, Computng in Sci. & Engineering, vol. 2, no. 1, 2000, pp. 22-23.

[69] J. L. Doob, *Stochastic Processes*, John Wiley, New York, NY, 1953.

[70] G. D. Doolen and John Hendricks, *Monte Carlo at Work*, Los Alamos Science, Special Issue Dedicated to S. Ulam, 1987, pp. 142-143; composite `http://www.fas.org/sgp/othergov/doe/lanl/pubs/00326867.pdf` .

[71] P. Dorato, C. Abdallah and V. Cerone, *Linear-Quadratic Control: An Introduction*, Prentice-Hall, Englewood Cliffs, NJ, 1995.

[72] J. Douglas Jr. and T. Dupont, *Galerkin Methods of Parabolic Equations*, SIAM J. Numerical Analysis, vol. 7, 1970, pp. 575-626.

[73] J. Douglas Jr., *Time Step Procedures of Nonlinear Parabolic PDEs* in Mathematics of Finite Elements and Applications, MAFELAP, J. Whiteman (Editor), Academic Press, London, 1979, pp. 289-304.

[74] D. Duffie, *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, NJ, 2001.

[75] D. Düvelmeyer *Untersuchungen zu Chancen und Risiken von Anti-Trend-Strategien am Beispiel des DAX-Futures*, Thesis, Facultät für Mathematik, Technische Universität Chemnitz, Chemnitz, 2001; `URL:http://www-usercgi.tu-chemnitz.de/{$\sim$}dana\diplom\_pdf\_dd.zip` .

[76] P. Dyer and S. R. McReynolds, *The Computation and Theory of Optimal Control*, Academic Press, New York, NY, 1970.

[77] E. B. Dykin, *Markov Processes I and II*, Academic Press, New York, NY, 1965.

[78] R. Eckhardt, *Ulam, John von Neumann, and the Monte Carlo Method*, Los Alamos Science, Special Issue Dedicated to S. Ulam, 1987, pp. 131-137. `http://www.fas.org/sgp/othergov/doe/lanl/pubs/00326867.pdf` .

[79] P. Embrechts , *Multivariate Extremes and Market Risk Scenarios*, Bachelior Finance Society Third World Congress, Chicago, July 2004.

[80] H. H. Engelhard, *Brain Tumors and the Blood-Brain Barrier*, Neuro-Oncology: The Essentials, Thieme Medical Publishers, Inc., New York 2000, pp. 49-53.

[81] J. D. Esary, F. Proschan and D. W. Walkup, *Association of Random Variables, with Applications*, Annals of Mathematical Statistics, vol. 38, 1967, pp. 1466-1474.

[82] M. Evans, N. Hastings and B. Peacock, *Statistical Distributions*, 3rd ed., John Wiley, New York, NY, 2000.

[83] W. Feller, *An Introduction to Probability Theory and Its Application*, vol. 1, 3rd ed., John Wiley, New York, NY, 1968.

[84] W. Feller, *An Introduction to Probability Theory and Its Application*, vol. 2, 2nd ed., John Wiley, New York, NY, 1971.

[85] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer-Verlag, New York, NY, 1975.

[86] J. J. Florentin, *Optimal Control of Systems with Generalized Poisson Inputs*, ASME Trans., vol. 85D (also J. Basic Engr., vol. 2), 1963, pp. 217-221.

[87] G. E. Forsythe, M. A. Malcolm and C. Moler, *Computer Methods for Mathematical Computations*, Prentice-Hall, Englewood Cliffs, NJ, 1977.

[88] B. Friedman, *Principles and Techniques of Applied Mathematics*, John Wiley, New York, NY, 1956.

[89] Future Directions in Control Theory Panel, *Report of the Panel on Future Directions in Control Theory: A Mathematical Perspective*, W. H. Fleming (Chairman), Society for Industrial and Applied Mathematics, Philadelphia, PA, 1988.

[90] Future Directions in Control, Dynamics, and Systems Panel, *Future Directions in Control in an Information Rich World: Report of the Panel on Future Directions in Control, Dynamics, and Systems*, R. M. Murray (Editor), Society for Industrial and Applied Mathematics, Philadelphia, PA, July 2003.

[91] T. C. Gard, *Introduction to Stochastic Differential Equations*, Marcel-Dekker, New York, NY, 1988.

[92] R. A. Gatenby and E. T. Gawlinski, *A Reaction-Diffusion Model of Cancer Invasion*, Cancer Research, vol. 56, 1996, pp. 5745-5753.

[93] S. B. Gershwin, *Manufacturing Systems Engineering*, Prentice-Hall, Englewood Cliffs, NJ, 1994.

[94] I. I. Gihman and A. V. Skorohod, *Stochastic Differential Equations*, Springer-Verlag, New York, NY, 1972.

[95] I. I. Gihman and A. V. Skorohod, *Controlled Stochastic Processes*, Springer-Verlag, New York, NY, 1979.

[96] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer-NY, New York, NY, 2003.

[97] P. W. Glynn and W. Whitt, *The Asymptotic Efficiency of Simulation Estimators*, Operations Research, vol. 40, no. 3, 1992, pp. 505-520.

[98] N. S. Goel and N. Richter-Dyn, *Stochastic Models in Biology*, Springer-NY, New York, NY, 2003.

[99] J. H. Goldie and A. J. Coldman, *Drug Resistance in Cancer: Mechanisms and Models*, Cambridge University Press, Cambridge, UK, 1998.

[100] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[101] M. D. Gunzburger, *Perspectives in Flow Control and Optimization*, SIAM, Philadelphia, PA, 2003.

[102] R. Haberman, *Elementary Applied Partial Differential Equations with Fourier Series and Boundary Value Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[103] W. Hackbusch, *A Numerical Method for Solving Parabolic Equations with Opposite Orientations*, Computing, vol. 20, 1978, pp. 229-240.

[104] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, Methuen, London, UK, 1964.

[105] F. B. Hanson, *Bioeconomic Model of the Lake Michigan Alewife Fishery*, Canadian Journal of Fisheries and Aquatic Sciences, vol. 44, suppl. 2, 1987, pp. 298-305.

[106] F. B. Hanson, *Stochastic Dynamic Programming: Advanced Computing Constructs*, in Proc. 28th IEEE Conference on Decision and Control, vol. 1, 1989, pp. 901-903.

[107] F. B. Hanson, *Computational Dynamic Programming on a Vector Multiprocessor*, IEEE Transactions on Automatic Control, vol. 36, no. 4, 1991, pp. 507-511.

[108] F. B. Hanson, *Computational Stochastic Dynamic Programming* in Stochastic Digital Control System Techniques, within series Control and Dynamic Systems: Advances in Theory and Applications, vol. 76, C. T. Leondes, ed., Academic Press, New York, NY, 1996, pp. 103-162.

[109] F. B. Hanson, *Local Supercomputing Training in the Computational Sciences Using Remote National Centers*, Future Generation Computer Systems: Special Issue on Education in the Computational Science, vol. 19, 2003, pp. 1335-1347.

[110] F. B. Hanson, *Computational Stochastic Control: Basic Foundations, Complexity and Techniques*, in Proc. 42nd IEEE Conference on Decision and Control, 2003, pp. 3024-3029.

[111] F. B. Hanson and K. Naimipour, *Convergence of Numerical Method for Multistate Stochastic Dynamic Programming*, in Proc. Int. Federation of Automatic Control 12th World Congress, vol. 9, 1993, pp. 501-504.

[112] F. B. Hanson, C. J. Pratico, M. S. Vetter, and H. H. Xu, *Multidimensional Visualization Applied to Renewable Resource Management*, in Proc. Sixth SIAM Conference on Parallel Processing for Scientific Computing, vol. 2, 1993, pp. 1033-1036.

[113] F. B. Hanson and D. Ryan, *Optimal Harvesting with Density Dependent Random Effects*, Natural Resource Modeling, vol. 2, 1988, pp. 439-455.

[114] F. B. Hanson and D. Ryan, *Mean and Quasideterministic Equivalence for Linear Stochastic Dynamics*, Mathematical Biosciences, vol. 93, 1989, pp. 1-14.

[115] F. B. Hanson and D. Ryan, *Optimal Harvesting with Both Population and Price Dynamics*, Mathematical Biosciences, vol. 148, 1998, pp. 129-146.

[116] F. B. Hanson and C. Tier, *An Asymptotic Solution to the First Passage Time Problem for Singular Diffusion Arising in Population Biology*, SIAM J. Appl. Math., vol. 40, 1981, pp. 113-132.

[117] F. B. Hanson and C. Tier, *A Stochastic Model of Tumor Growth*, Mathematical Biosciences, vol. 61, 1982, pp. 73-100.

[118] F. B. Hanson and H. C. Tuckwell, *Persistence Times of Populations with Large Random Fluctuations*, Theor. Population Biol., vol. 14, 1978, pp. 46-61.

[119] F. B. Hanson and H. C. Tuckwell, *Logistic Growth with Random Density Independent Disasters*, Theor. Population Biol., vol. 19, 1981, pp. 1-18.

[120] F. B. Hanson and H. C. Tuckwell, *Diffusion Approximations for Neuronal Activity including Reversal Potentials*, J. Theoretical Neurobiology, vol. 2, 1983, pp. 127-153.

[121] F. B. Hanson and H. C. Tuckwell, *Population Growth with Randomly Distributed Jumps*, J. Mathematical Biology, vol. 36, no. 2, 1997, pp. 169-187.

[122] F. B. Hanson and J. J. Westman, *Optimal Consumption and Portfolio Policies for Important Jump Events: Modeling and Computational Considerations*, in Proc. American Control Conference, 2001, pp. 4456-4661.

[123] F. B. Hanson and J. J. Westman, *Optimal Consumption and Portfolio Control for Jump-Diffusion Stock Process with Log-Normal Jumps*, in Proc. American Control Conference, July 2002, pp. 4256-4261; for corrected version see `ftp://www.math.uic.edu/pub/Hanson/ACC02/acc02webcor.pdf` .

[124] F. B. Hanson and J. J. Westman, *Stochastic Analysis of Jump-Diffusions for Financial Log-Return Processes*, in Stochastic Theory and Control, Proceedings of a Workshop held in Lawrence, Kansas, October 18-20, 2001, Lecture Notes in Control and Information Sciences, vol. 280, B. Pasik-Duncan (Editor), Springer- Verlag, New York, pp. 169-184, 2002; for corrected version see `ftp://www.math.uic.edu/pub/Hanson/KU01/ku02hwfmcor.pdf` .

[125] F. B. Hanson and J. J. Westman, *Jump-Diffusion Stock Return Models in Finance: Stochastic Process Density with Uniform-Jump Amplitude*, in Proc. 15th Int. Sympos. Mathematical Theory of Networks and Systems, August 2002, pp. 1-7.

[126] F. B. Hanson and J. J. Westman, *Portfolio Optimization with Jump–Diffusions: Estimation of Time-Dependent Parameters and Application*, in Proc. 41st Conference on Decision and Control, December 2002, pp. 377–382.

[127] F. B. Hanson and J. J. Westman, *Jump-Diffusion Stock-Return Model with Weighted Fitting of Time-Dependent Parameters*, in Proc. American Control Conference, 2003, pp. 4869-4874.

[128] F. B. Hanson, J. J. Westman and Z. Zhu, *Maximum Multinomial Likelihood Estimation of Market Parameters for Stock Jump-Diffusion Models*, in Mathematics of Finance: Proc. 2003 AMS-IMS-SIAM Joint Summer Research Conference on Mathematics of Finance, AMS Contemporary Mathematics, vol. 351, G. Yin and Q. Zhang (Editors), 2004, pp. 155-169.

[129] F. B. Hanson and J. J. Westman, *Optimal Portfolio and Consumption Policies Subject to Rishel's Important Jump Events Model: Computational Methods*, IEEE Transactions on Automatic Control, vol. 48, no. 3, Special Issue on Stochastic Control Methods in Financial Engineering, 2004, pp. 326-337.

[130] F,B. Hanson and G. Yan, *American Put Option Pricing for Stochastic-Volatility, Jump-Diffusion Models*, Proc. 2007 American Control Conference, 2007, pp. 1-6, submitted invited paper; `http://www.math.uic.edu/~hanson/pub/GYan/ACC07fhgywebpub.pdf` .

[131] F. B. Hanson and Z. Zhu, *Comparison of Market Parameters for Jump-Diffusion Distributions Using Multinomial Maximum Likelihood Estimation*, in Proc. 42nd IEEE Conference on Decision and Control, December 2004, pp. 3919-3924.

[132] J. M. Harrison and S. R. Pliska, *Martingales and Stochastic Integrals in the Theory of Continuous Trading*, Stochastic Processes & Appl., vol. 11, pp. 215-260, 1981.

[133] J. M. Harrison and S. R. Pliska, *A Stochastic Calculus Model of Continuous Trading: Complete Markets*, Stochastic Processes & Appl., vol. 15, pp. 313-316, 1983.

[134] M. B. Haugh and A. W. Lo, *Computational Challenges in Portfolio Management*, Computing in Sci. &. Engr., pp. 54l-59, May/June 2000.

[135] D. Heath and M. Schweizer, *Martingales versus PDEs in Finance: An Equivalence Result with Examples*, J. Appl. Prob., vol. 37, pp. 947-957, 2000.

[136] J. W. Helton and O. Merino, *Classical Control Using H^∞ Methods: An Introduction to Design*, SIAM, Philadelphia, PA, 1998.

[137] J. W. Helton and O. Merino, *Classical Control Using H^∞ Methods: Theory, Optimization, and Design*, SIAM, Philadelphia, PA, 1998.

[138] R. C. Hennemuth, J. E. Palmer and B. E. Brown, *A Statistical Description of Recruitment in Eighteen Selected Fish Stocks*, J. Northw. Alt. Fish Sci., vol. 1, pp. 101-111, 1980.

[139] D. J. Higham, *An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations*, SIAM Review, vol. 43, no. 3, 2001, pp. 525-546.

[140] D. J. Higham, *An Introduction to Financial Option Valuation: Mathematics, Stochastics and Computation*, Cambridge University Press, Cambridge, UK, 2004.

[141] D. J. Higham, *Black-Scholes for Scientific Computing Students*, Computing in Science & Engineering, vol. 6, no. 6, pp. 72-79, November/December 2004.

[142] D. J. Higham and N. J. Higham, *MATLAB Guide*, SIAM, Philadelphia, PA, 2000.

[143] D. J. Higham and P. E. Kloeden, *Maple and MATLABfor Stochastic Differential Equations in Finance*, in Programming Languages and Systems in Computational Economics and Finance, S. S. Neilsen (Editor), Kluwer, 2002, pp. 233-270. `http://www.maths.strath.ac.uk/~aas96106/algfiles.html` .

[144] D. J. Higham and P. E. Kloeden, *Numerical Methods for Nonlinear stochastic differential equations with jumps*, Numerische Mathematik, vol. 101, 2005, pp. 101-119.

[145] D. J. Higham and P. E. Kloeden, *Convergence and Stability of Implicit Methods for Jump-Diffusion Systems*, Numerische Mathematik, vol. 101, 2005, pp. 101-119.

[146] D. J. Higham, X. Mao and A. M. Stuart, *Strong Convergence of Euler-Type Methods for Nonlinear Stochastic Differential Equations*, SIAM J. Numer. Anal., vol. 40, no. 3, 2002, pp. 1041-1063.

[147] J. C. Hull, *Options, Futures, & Other Derivatives*, 4th Edition, Prentice-Hall, Englewood Cliffs, NJ, 2000.

[148] International Pacific Halibut Commission, *Annual Reports*, Seattle, WA , 1984-1985.

[149] K. Itô, *On Stochastic Differential Equations*, Mem. Amer. Math. Soc., no. 4, 1951, pp. 1-51.

[150] P. Jäckel, *Monte Carlo Methods in Finance*, John Wiley, New York, NY, 2002.

[151] D. H. Jacobson and D. Q. Mayne, *Differential Dynamic Programming*, American Elsevier, New York, NY, 1970.

[152] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin, 1987.

[153] R. A. Jarrow and E. R. Rosenfeld, *Jump Risks and the Intertemporal Capital Asset Pricing Model*, J. Business, vol. 57 (3), 1984, pp. 337-351.

[154] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, NY, 1970.

[155] P. Jorion, *On Jump Processes in the Foreign Exchange and Stock Markets*, Rev. Fin. Studies, vol. 88, no. 4, 1989, pp. 427-445.

[156] R. E. Kalman, *Contributions to the Theory of Optimal Control*, Bol. Soc. Mat. Mex., vol. 5, 1960, pp. 102-119.

[157] M. H. Kalos and P. A. Whitlock, Monte Carlo Methods, Volume I: Basics, John Wiley and Sons, NY, 1986.

[158] M. I. Kamien and N. L. Schwartz, *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, North-Holland, New York, NY, 1981.

[159] I. Karatzas, J. P. Lehoczky, S. P. Sethi and S. E. Shreve, *Explicit Solution of a General Consumption/Investment Problem*, Math. Oper. Res., vol. 11, 1986, pp. 261–294. (Reprinted in Sethi [245, Chapter 2].)

[160] I. Karatzas and S. E. Shreve, *Methods of Mathematical Finance*, Springer-Verlag, New York, NY, 1998.

[161] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd ed., Academic Press, New York, NY, 1975.

[162] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes*, Academic Press, New York, NY, 1981.

[163] D. E. Kirk, *Optimal Control Theory: An Introduction*, Prentice-Hall, Englewood Cliffs, NJ, 1970. (Reprinted by Dover Publications, Mineola, NY, 2004.)

[164] F. C. Klebaner, *Introduction to Stochastic Calculus with Applications*, Imperial College Press, London, UK, 1998.

[165] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, New York, NY, 1992.

[166] P. E. Kloeden, E. Platen and H. Schurz, *Numerical Solution of SDE Though Computer Experiments*, Springer-Verlag, New York, NY, 1994.

[167] P. Kokotović, H. K. Khalil and J. O'Reilly, *Singular Perturbation Methods in Control: Analysis and Design*, Academic Press, New York, NY, 1986.

[168] A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis*, R. A. Silverman (Translator), Dover Publications, New York, NY, 1970.

[169] S. G. Kou, *A Jump Diffusion Model for Option Pricing*, Management Science, vol. 48, 2002, pp. 1086–1101.

[170] S. G. Kou and H. Wang, *Option Pricing Under a Double Exponential Jump Diffusion Model*, Management Science, vol. 50 (9), 2004), pp. 1178–1192.

[171] K. Koya, H. Kimura and M. Kawato, *Neural Mechanisms of Learning and Control* IEEE Control Systems Magazine, vol. 21, no. 4, August 2001, pp. 42-54.

[172] H. Kunita and S. Watanabe, *On Square-Integrable Martingales*, Nagoya Math. J., vol. 30, 1967, pp. 209-245.

[173] H. J. Kushner, *Stochastic Stability and Control*, Academic Press, New York, NY, 1967.

[174] H. J. Kushner, *A Survey of Some Applications of Probability and Stochastic Control Theory to Finite Difference Methods for Degenerate Elliptic and Parabolic Equations*, SIAM Review, vol. 18, 1976, pp. 545-577.

[175] H. J. Kushner, *Numerical Methods for Stochastic Control in Continuous Time*, SIAM J. Control and Optimizations, vol. 28, 1990, pp. 999-1048.

[176] H. J. Kushner, *Numerical Methods for Stochastic Control in Continuous Time*, SIAM J. Control and Optimizations, vol. 28, 1990, pp. 999-1048.

[177] H. J. Kushner, *Jump-Diffusions with Controlled Jumps: Existence and Numerical Methods*, J. Math. Anal. Applic., vol. 249, no. 1, 2000, pp. 179-198.

[178] H. J. Kushner amd G. DiMasi, *Approximation for Functionals and Optimal Control Problems on Jump Diffusions Processes*, J. Math. Anal. Applic., vol. 63, 1978, pp. 772-800.

[179] H. J. Kushner and P. G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2nd Edition, Springer-Verlag, New York, NY, 2001, .

[180] H. J. Kushner and D. J. Jarvis, *Large-Scale Computations for High Dimension Control Systems*, in Proc. 33rd IEEE Conference on Decision and Control, vol. 1, 1994, pp. 461-465.

[181] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms Recursive Algorithms and Applications*, Springer-Verlag, New York, NY, 2003.

[182] R. E. Larson, *A Survey of Dynamic Programming Computational Procedures*, IEEE Transactions on Automatic Control, vol. AC-16, 1967, pp. 767-774.

[183] G. P. Lepage, *A New Algorithm for Adaptive Multidimensional Integration* , Journal of Computational Physics, vol. 27, no. 2, 1978, pp. 192-203.

[184] F. L. Lewis, *Optimal Estimation with an Introduction to Stochastic Control Theory*, John Wiley, New York, NY, 1986.

[185] M. J. Lighthill, *Introduction to Fourier Analysis and Generalised Functions*, Cambridge University Press, Cambridge, UK, 1964.

[186] A. Lipton, *Mathematical Methods for Foreign Exchange: A Financial Engineer's Approach* , World Scientific, Singapore, 2001.

[187] D. Ludwig, *Stochastic Population Theories*, Springer-Verlag, New York, NY, 1974.

[188] D. Ludwig, *Persistence of Dynamical Systems Under Random Perturbations*, SIAM Review, vol. 17, 1979, pp. 605-640.

[189] D. Ludwig, *Optimal Harvesting of a Randomly Fluctuating Resource. I: Application of Perturbation Methods*, SIAM J. Applied Mathematics, vol. 37, 1979, pp. 166-184.

[190] D. Ludwig and J. M. Varah, *Optimal Harvesting of a Randomly Fluctuating Resource. II: Numerical Methods and Results*, SIAM J. Applied Mathematics, vol. 37, 1979, pp. 185-205.

[191] Y. Maghsoodi, *Mean Square Efficient Numerical Solution of Jump-Diffusion Stochastic Differential Equations*, Sankhyā: Indian J. Stat., vol. 58, ser. A, pt. 1, 1996, pp. 25-47.

[192] Y. Maghsoodi and C. J. Harris, *In-Probability Approximation and Simulation of Nonlinear Jump-Diffusion Stochastic Differential Equations*, IMA J. Math Control & Info., vol. 4, 1996, pp. 65-492.

[193] M. Mangel, *Decision and Control in Uncertain Resource Systems*, Academic Press, New York, NY, 1985.

[194] M. Mariton, *Jump Linear Systems in Automatic Control*, M. Dekker, New York, NY, 1990.

[195] G. Marsaglia and T. A. Bray, *A Convenient Method for Generating Normal Variables*, SIAM Rev., vol. 6, 1964, pp. 260-264.

[196] D. Q. Mayne, *A Second–Order Gradient Method for Determining Optimal Control of Non–Linear Discrete Time Systems*, International Journal of Control. vol. 3, 1966, pp. 85-95.

[197] D. Q. Mayne, *Differential Dynamic Programming — A Unified Approach to the Optimization of Dynamical Systems*, Control and Dynamical Systems: Advances in Theory and Applications, vol. 10, C. T. Leondes, ed., Academic Press, New York, NY, 1973, pp. 179-254.

[198] R. C. Merton, *Lifetime Portfolio Selection Under Uncertainty: The Continuous-Time Case*, Rev. Econ. and Stat., vol. 51, 1969, pp. 247-257. (Reprinted in Merton [203, Chapter 4].)

[199] R. C. Merton, *Optimum Consumption and Portfolio Rules in a Continuous-Time Model*, J. Econ. Theory, vol. 3 (4), 1971, pp. 373-413. (Reprinted in Merton [203, Chapter 5].)

[200] R. C. Merton, *Eratum*, J. Econ. Theory, vol. 6, no. 2, 1973, pp. 213-214.

[201] R. C. Merton, *Theory of Rational Option Pricing*, Bell J. Econ. Mgmt. Sci., vol. 4, 1973 (Spring), pp. 141-183. (Reprinted in Merton [203, Chapter 8].)

[202] R. C. Merton, *Option Pricing When Underlying Stock Returns are Discontinuous*, J. Financial Economics, vol. 3, 1976, pp. 125-144. (Reprinted in Merton [203, Chapter 9].)

[203] R. C. Merton, *Continuous Time Finance*, Blackwell Publishers, Cambridge, MA, 1992.

[204] R. C. Merton and M. S. Scholes, *Fischer Black*, J. Finance, vol. 50, no. 5, 1996, pp. 1359-1369.

[205] W. C. Messner and D. M. Tilbury, *Control Tutorials for MATLAB And Simulink: User's Guide*, Addison-Wesley Publ. Co., 2002; see also `http://www.engin.umich.edu/group/ctm/` .

[206] N. Metropolis, *The Beginning of the Monte Carlo Method*, Los Alamos Science, Special Issue Dedicated to S. Ulam, 1987, pp. 125-130. `http://www.fas.org/sgp/othergov/doe/lanl/pubs/00326866.pdf` .

[207] N. Metropolis, A. N. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller' *Equation of State Calculation by Fast Computing Machines*, J. Chem. Phys., vol. 21, no. 6, 1953, 10871092.

[208] N. Metropolis and S. Ulam, *The Monte Carlo Method*, J. Amer. Stat. Assoc., vol. 44, no. 247, 1949, pp. 335-341.

[209] T. Mikosch, *Elementary Stochastic Calculus: with Finance in View*, World Scientific, Singapore, 1998.

[210] C. Moler, et al., *Using MATLAB*, vers 6., Mathworks, Natick, MA, 2000.

[211] F. Moss, L. M. Ward and W. G. Sannita, *Stochastic Resonance and Sensory Information Processing*, Clinical Neurophysiology, vol. 115, 2004, pp. 267-281.

[212] G. I. Murphy., *Clupeoids*, in Fish Population Dynamics, J. A. Gulland (Editor), John Wiley, New York, NY, 1977, pp. 283-308.

[213] J. D. Murray., *Mathematical Biology, I: An Introduction*, Springer-NY, New York, NY, 2002.

[214] J. D. Murray., *Mathematical Biology, II: Spatial Models and Biomedical Applications*, Springer-NY, New York, NY, 2003.

[215] R. M. Murray, K. J. Aström, R. W. Brockett and G. Stein, *Future Directions in Control in an Information Rich World: A Summary of the Report of the Panel on Future Directions in Control, Dynamics, and Systems*, IEEE Control Systems Magazine, vol. 23, no. 2, April 2003, pp. 20-33; see also URL:http://www.cds.caltech.edu/~murray/cdspanel/, California Institute of Technology, April 2002.

[216] K. Naimipour and F. B. Hanson, *Convergence of a Numerical Method for the Bellman Equation of Stochastic Optimal Control with Quadratic Costs and Constrained Control*, Dynamics and Control, vol. 3, no. 3, 1993, pp. 237-259.

[217] S. N. Neftci, *Introduction to the Mathematics of Financial Derivative*, 2nd ed., Academic Press, New York, NY, 2000.

[218] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, PA, 1989.

[219] R. M. Nisbet and W. S. C. Gurney, *Modelling Fluctuating Populations*, John Wiley, New York, NY, 1982.

[220] N. S. Nise, *Control Systems Engineering*, John Wiley, New York, NY, 2000.

[221] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer-Verlag, New York, NY, 1999.

[222] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Fifth ed., Springer-Verlag, New York, NY, 1998.

[223] B. Øksendal and A. Sulem, *Applied Stochastic Control of Jump Diffusions*, Springer-Verlag, Berlin, 2005.

[224] E. Parzen, *Stochastic Processes*, Holden-Day, San Francisco, CA, 1962.

[225] S. R. Pliska, *Introduction to Mathematical Finance: Discrete Time Models*, Blackwell Publishers, Cambridge, MA, 1997.

[226] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, Wiley-Interscience Publishers, New York, NY, 1962.

[227] E. Polak, *An Historical Survey of Computational Methods in Optimal Control*, SIAM Review, vol. 15, 1973, pp. 553-584.

[228] C. J. Pratico, F. B. Hanson, H. H. Xu, D. J. Jarvis and M. S. Vetter, *Visualization for the Management of Renewable Resources in an Uncertain Environment*, in Proc. Supercomputing '92, 1992, pp. 258-266, color plates p. 843.

[229] W. H. Press and G. R. Farrar, *Recursive Stratified Sampling for Multidimensional Monte Carlo Integration*, Computers in Physics, vol. 4, no. 2, 1990, pp. 190-195.

[230] W. H Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing*, 2nd ed., Cambridge University Press, Cambridge, UK, 2002.

[231] A. Priplata, J. Niemi, M. Salen, J. Harry L. A. Lipsitz and J. J. Collins, *Noise-Enhanced Human Balance Control*, Phys. Rev. Let., vol. 89, no. 23, 2002, pp. 2381011-2381014.

[232] P. Protter, *Stochastic Integration and Differential Equations: A New Approach*, Springer-Verlag, Berlin, 2004.

[233] D. W. Repperger, *Celebrating the 100th Anniversary of Controlled, Sustained, and Powered Air Flight,* in IEEE Control Systems Magazine, vol. 23, no. 6, 2003, pp. 12-16.

[234] *Research Directions in Distributed Parameter Systems*, R. C. Smith and M. A. Demetriou (Editors), SIAM, Philadelphia, PA, 2003.

[235] R. Rishel, *Modeling and Portfolio Optimization for Stock Prices Dependent on External Events,* in Proc. 38th IEEE Conference on Decision and Control, 1999, pp. 2788-2793.

[236] L. C. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales*, Cambridge University Press, Cambridge, UK, 2000.

[237] S. M. Ross, *Stochastic Processes*, John Wiley, New York, NY, 1983.

[238] S. M. Ross, *Introduction to Probability Models*, 7th ed., Academic Press, 2000.

[239] W. J. Runggaldier, *Jump-Diffusion Models*, in Handbook of Heavy Tailed Distributions in Finance, S. T. Rachev (Editor), Handbooks in Finance, Elsevier/North-Holland, New York, NY, 2003, pp. 169-209.

[240] D. Rust, *Do People Behave According to Bellman's Principle Optimality?*, Hoover Institution Working Paper E-92-10, 65 pages, 1994.

[241] D. Ryan and F. B. Hanson, *Optimal Harvesting with Exponential Growth in an Environment with Random Disasters and Bonanzas*, Math. Biosci., vol. 74, 1985, pp. 37-57.

[242] D. Ryan and F. B. Hanson, *Optimal Harvesting of a Logistic Population in an Environment with Stochastic Jumps*, J. Math. Biol., vol. 24, 1986, pp. 259-277.

[243] T. H. Rydberg, *The Normal Inverse Gaussian Lévy Process: Simulation and Approximation*, Comm. Stat. Stoch. Models, vol. 13, 1997, pp. 887-910.

[244] Z. Schuss, *Theory and Applications of Stochastic Differential Equations*, John Wiley, New York, NY, 1980.

[245] S. P. Sethi, *Optimal Consumption and Investment with Bankruptcy* Kluwer Academic Publishers, Boston, MA, 1997.

[246] S. P. Sethi and M. Taksar, *A Note on Merton's "Optimum Consumption and Portfolio Rules in a Continuos-Time Model"* J. Econ. Theory, vol. 46 (2), 1988, pp. 395-401. (Reprinted in Sethi [245, Chapter 3].)

[247] S. P. Sethi and Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems* Birkhäuser, Boston, MA, 1994.

[248] S. E. Shreve, *Stochastic Calculus Models for Finance II: Continuous-Time Models*, Springer-Verlag, New York, NY, 2004.

[249] L. Simpson-Herren and H. H. Lloyd, *Kinetics Parameters and Growth Curves for Experimental Tumor Systems*, Cancer Chemo. Reps., vol. 54, 1970, pp. 143-174.

[250] C. Sindermann, *Principal Diseases of Marine Fish and Shellfish*, Academic Press, New York, NY, 1970.

[251] I. N. Sneddon, *Elements of Partial Differential Equations*, McGraw-Hill Book Co., New York, NY, 1957.

[252] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*, 2nd ed., Springer-Verlag, New York, NY, 1991.

[253] I. M. Sobol', *On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals* Comp. Math. and Math. Physics (Engl. Transl.), vol. 7, 1967, pp. 784–802.

[254] H. W. Sorensen (ed.), *Kalman Filtering: Theory and Application*, IEEE Press, New York, NY, 1985.

[255] G. G. Steel, *Growth Kinetics of Tumors*, Clarendon Press, Oxford, UK, 1977.

[256] J. M. Steele, *Stochastic Calculus and Financial Applications*, Springer-NY, New York, NY, 2001.

[257] R. B. Stein, *Nerve and Muscle: Membranes, Cells, and Systems*, Plenum Press, New York, NY, 1980.

[258] R. Stengel, *Stochastic Optimal Control: Theory and Application*, Dover Publications, New York, NY, 1994.

[259] S. Stojanovic, *Computational Financial Mathematics using Mathematica: Optimal Trading in Stocks and Options*, Birkhäuser, Boston, MA, 2002.

[260] R. L. Stratonovich, *A New Representation for Stochastic Integrals and Equations*, SIAM J. Control, vol. 4, 1966, pp. 362-371.

[261] G. W. Swan, *Applications of Optimal Control in Biomedicine*, Marcel Dekker, New York, NY, 1984.

[262] K. R. Swanson, *Mathematical Modeling of the Growth and Control of Tumors*, Ph.D. Thesis, University of Washington, Seattle, 1999.

[263] A. E. Taylor and W. R. Mann, *Advanced Calculus*, 2nd ed., Xerox College Publishing, Lexington, MA, 1972.

[264] D. Tavella and C. Randall, *Pricing Financial Instruments: The Finite Difference Method*, John Wiley, New York, NY, 2000.

[265] H. M. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, 3rd ed., Academic Press, New York, NY, 1998.

[266] C. Tier and F. B. Hanson, *Persistence in Density Dependent Stochastic Populations*, Mathematical Biosciences, vol. 53, 1981, pp. 89-117.

[267] H .C. Tijms, *Stochastic Modelling and Analysis: A Computational Approach*, John Wiley, New York, NY, 1986.

[268] D. M. Tilbury and W. C. Messner, *Control Tutorials for Software Instruction over the World Wide Web*, IEEE Trans. Automatic Control, vol. 42, no. 4, 1999, pp. 237-246.

[269] H .C. Tuckwell, *Stochastic Processes in the Neurosciences*, SIAM, Philadelphia, PA, 1989.

[270] H .C. Tuckwell, *Elementary Applications of Probability Theory*, Chapman and Hall, London, UK, 1995.

[271] M. Turelli, *Random Environments and Stochastic Calculus*, Theor. Population Biol., vol. 12, 1977, pp. 140-178.

[272] W. T. Vetterling, S. A. Teukolsky, W. H Press and B. P. Flannery, *Numerical Recipes: Example Book [C++]*, 2nd ed., Cambridge University Press, Cambridge, UK, 2002.

[273] J. von Neumann, *Various Techniques Used in Connection with Random Digits*, in Applied Mathematics Series, vol. 12, U. S. National Bureau of Standards, 1951, pp. 36-38.

[274] J. J. Westman and F. B. Hanson, *The LQGP Problem: A Manufacturing Application*, in Proc. American Control Conference, 1997, pp. 566-570.

[275] J. J. Westman and F. B. Hanson, *Nonlinear State Dynamics: Computational Methods and Manufacturing Example*, International Journal of Control, 2000, vol. 73, pp. 464-480.

[276] J. J. Westman and F. B. Hanson, *State Dependent Jump Models in Optimal Control*, in Proc. 38th IEEE Conference on Decision and Control, 1999, pp. 2378-2383.

[277] J. J. Westman and F. B. Hanson, *Nonlinear State Dynamics: Computational Methods and Manufacturing Example*, International Journal of Control, vol. 73, 2000, pp. 464-480.

[278] J. J. Westman and F. B. Hanson, *MMS Production Scheduling Subject to Strikes in Random Environments*, in Proc. American Control Conference, 2000, pp. 2194-2198.

[279] J. J. Westman, F. B. Hanson and E. K. Boukas, *Optimal Production Scheduling for Manufacturing Systems with Preventive Maintenance in an Uncertain Environment*, in Proc. American Control Conference, 2001, pp. 1375-1380.

[280] N. Wiener, *Differential Space*, Journal of Math. and Physics, vol. 2, 1923, pp. 132-174.

[281] N. Wiener, *Generalized Harmonic Analysis*, Acta Mathematica, vol. 55, 1930, pp. 117-258.

[282] P. Wilmott, S. Howison and J. Dewynne, *The Mathematics of Financial Derivatives: A Student Introduction*, Cambridge University Press, Cambridge, UK, 1996.

[283] P. Wilmott, *Paul Wilmott on Quantitative Finance*, vols. 1 & 2, John Wiley, New York, NY, 2000.

[284] S. Wolfram, *Mathematica: A System for Doing Mathematics by Computer*, Addison-Wesley, Reading, MA, 1988.

[285] W. M. Wonham, *Random Differential Equations in Control Theory*, in Probabilistic Methods in Applied Mathematics, vol. 2, Academic Press, New York, NY, 1970, pp. 131-212.

[286] Yahoo! Finance, *Historical Quotes, S & P 500 Symbol ^SPC*, `URL:http://chart.yahoo.com/` , February 2001.

[287] G. Yan and F. B. Hanson, *Option Pricing for a Stochastic-Volatility Jump-Diffusion Model with Log-Uniform Jump-Amplitudes*, Proc. 2006 American Control Conference, 2006, pp. 2989-2994.

[288] J. Yong and X. Y. Zhou, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, Springer-Verlag, New York, NY, 1999.

[289] Z. Zhu, *Option Pricing and Jump-Diffusion Models*, Ph.D. Thesis in Mathematics, Univ. Illinois at Chicago, Chicago, IL, 17 October 2005, 165 pages.

[290] Z. Zhu and F. B. Hanson, *A Monte-Carlo Option-Pricing Algorithm for Log-Uniform Jump-Diffusion Model* in Proc. 44nd IEEE Conference on Decision and Control and European Control Conference, 2005, pp. 5221-5226.

[291] Z. Zhu and F. B. Hanson, *Optimal Portfolio Application with Double-Uniform Jump Model* in Control Theory Applications in Financial Engineering and Manufacturing, International Series in Operations Research and Management Sciences, Springer/Kluwer, New York, NY, 2006, 28 pages, invited chapter.

Index

Appendix A

Deterministic Optimal Control

As far as the laws of mathematics refer to reality,
they are not certain;
and as far as they are certain,
they do not refer to reality.
—Albert Einstein (1879-1955) [quoted by J.R. Newman in *The World of Mathematics*].

$m = L/c^2$,
—Albert Einstein (1879-1955), the original form of his famous energy-mass relation $E = mc^2$, where L is the Lagrangian, sometimes a form of energy and the cost part of the Hamiltonian in deterministic control theory.

It probably comes as a surprise to many Americans that the Wright brothers, Orville and Wilbur, did not invent flying, but they developed the *first free, controlled and sustained powered flight by man* as reviewed in Repperger's historical perspective on their technical challenges [233]. Indeed, control is embedded in many modern appliances working silently in computers, motor vehicles and other useful appliances. Beyond engineering design there are natural control systems like the remarkable human brain together with other components of the central nervous system [171]. Basar [21] lists twenty-five seminal papers in control and Bernstein [29] reviews control history through feedback control. The state and future directions of control of dynamical systems were summarized in the 1988 Fleming panel report [89] and more recently in the 2003 Murray panel report [90].

This chapter provides summary background as a review to provide a basis for examining the difference between deterministic optimal control and stochastic optimal control, treated in Chapter 6. Summarized with commentary are Hamil-

ton's equations, the maximum principle and dynamic programming formulation. A special and useful canonical model, the linear quadratic (LQ) model, is presented.

A.1 Hamilton's Equations: Hamiltonian and Lagrange Multiplier Formulation of Deterministic Optimal Control

For deterministic control problems [163, 44], many can be cast as systems of ordinary differential equations so there are many standard numerical methods that can be used for the solution. For example, if $\mathbf{X}(t)$ is the state n_x-vector on the state space $\mathcal{X}$ in continuous time t and $\mathbf{U}(t)$ is the control n_u-vector on the control space $\mathcal{U}$, then the differential equation for the deterministic system dynamics is

$$\frac{d\mathbf{X}}{dt}(t) = \mathbf{f}(\mathbf{X}(t), \mathbf{U}(t), t), \quad \mathbf{X}(t_0) = \mathbf{x}_0 \, . \tag{A.1}$$

Here, $\mathbf{f}(\mathbf{x}, \mathbf{u}, t)$ is called the **plant function** and may be nonlinear. The **cost objective** functional or **performance index** is to achieve the minimal cumulative **running** or **instantaneous costs** $C(\mathbf{x}, \mathbf{u}, t)$ on (t_0, f_f) plus **terminal cost** function $S(\mathbf{x}, t)$, that is,

$$V[\mathbf{X}, \mathbf{U}, t_f](\mathbf{x}_0, t_0) = \int_{t_0}^{t_f} C\left(\mathbf{X}(t), \mathbf{U}(t), t\right) dt + S\left(\mathbf{X}(t_f), t_f\right) \, . \tag{A.2}$$

Often in deterministic control theory and the calculus of variations, the cost function is also called the **Lagrangian**, i.e., $L(\mathbf{x}, \mathbf{u}, t) = C(\mathbf{x}, \mathbf{u}, t)$, from analogy with classical mechanics. The notation $V[\mathbf{X}, \mathbf{U}, t_f](\mathbf{x}_0, t_0)$ means that the cost is a functional of the state and control trajectory functions $V[\mathbf{X}, \mathbf{U}, t_f]$, i.e., a function of functions, but also is a function of the values of the initial data $(\mathbf{x}_0, t_0)$, i.e., a function dependence in the ordinary sense. This fairly general functional form with running and terminal costs is called the **Bolza form** of the objective functional. However, the notation $C(\mathbf{x}, \mathbf{u}, t)$ will be used for the instantaneous component of the objective even when it is not a cost and the overall objective is maximization rather than minimization, e.g., the maximization of profit.

Here, the value of the minimum total costs with respect to the control space $\mathcal{U}$ will be considered,

$$v^*(\mathbf{x}_0, t_0) = \min_{\mathbf{U} \in \mathcal{U}} \left[V[\mathbf{X}, \mathbf{U}, t_f](x_0, t_0) \right] \, , \tag{A.3}$$

unless otherwise specified, subject to the initial value problem for the controlled dynamics in (A.1). The is very little difference between the global minimum and the global maximum problem, the smallest value is found in the former and the largest value in the latter. The search in both cases is over all **critical points**, which consist of the set of all **regular points** or local optima which here are points where the control derivative or gradient is zero, **boundary points** of the control domain and **singular points** or other irregular points. If the control space $\mathcal{U}$ is the

whole space $\mathbb{R}^{n_u}$ the control problem is said to be **unconstrained** or in absence of constraints and then the problem is mainly searching for regular points, assuming there are no singular points, so

$$v^*(\mathbf{x}_0, t_0) = v^{(\text{reg})}(\mathbf{x}_0, t_0) = \min_{\mathbf{U}\in\mathbb{R}^{n_u}} \left[V[\mathbf{X}, \mathbf{U}, t_f](x_0, t_0)\right] . \tag{A.4}$$

In the Hamiltonian formulation [163], the Bolza form of optimization objective is replaced by a running cost optimal objective extended to include the state dynamics and the new optimization objective function is called the **Hamiltonian**:

$$\mathcal{H}(\mathbf{X}(t), \mathbf{U}(t), \boldsymbol{\lambda}(t), t) \equiv C(\mathbf{X}(t), \mathbf{U}(t), t) + \boldsymbol{\lambda}^T(t)\mathbf{f}(\mathbf{X}(t), \mathbf{U}(t), t) , \tag{A.5}$$

where $\boldsymbol{\lambda}(t)$ is the n_x-vector **Lagrange multiplier**, also called the **adjoint state** or **co-state** or auxiliary vector. The Lagrange multiplier provides the objective extension for including the state dynamics. The symbol $\boldsymbol{\lambda}$ should not be confused with the Poisson rate use in stochastic jump modeling, since the jump rate does not appear in deterministic problems, but both deterministic and stochastic uses are standard notations in the appropriate context.

Theorem A.1. ***Gradient necessary conditions for a regular control optimum – Interior point optimum principle:***
Let the Hamiltonian $\mathcal{H}$ have continuous first order derivatives in the state, co-state and control vectors, $\{\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}\}$. Then the ***necessary conditions*** *for an* ***interior point optimum*** *(maximum or minimum) of the Hamiltonian $\mathcal{H}$ at the optimal set of three vectors, $\{\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t)\}$, marked with an asterisk $(*)$, are called* ***Hamilton's equations:***

$$\frac{d\mathbf{X}^*}{dt}(t) = \left(\frac{\partial\mathcal{H}}{\partial\boldsymbol{\lambda}}\right)^* \equiv \left(\frac{\partial\mathcal{H}}{\partial\boldsymbol{\lambda}}\right)(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t), t) = \mathbf{f}(\mathbf{X}^*(t), \mathbf{U}^*(t), t) , \tag{A.6}$$

$$-\frac{d\boldsymbol{\lambda}^*}{dt}(t) = \left(\frac{\partial\mathcal{H}}{\partial\mathbf{x}}\right)^* \equiv \left(\frac{\partial\mathcal{H}}{\partial\mathbf{x}}\right)(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t), t) = \left(\frac{\partial C}{\partial\mathbf{x}} + \frac{\partial\mathbf{f}^T}{\partial\mathbf{x}}\boldsymbol{\lambda}\right)^* , \tag{A.7}$$

$$\mathbf{0} = \left(\frac{\partial\mathcal{H}}{\partial\mathbf{u}}\right)^* \equiv \left(\frac{\partial\mathcal{H}}{\partial\mathbf{u}}\right)(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t), t) = \left(\frac{\partial C}{\partial\mathbf{u}} + \frac{\partial\mathbf{f}^T}{\partial\mathbf{u}}\boldsymbol{\lambda}\right)^* , \tag{A.8}$$

where from the critical condition (A.8) , the optimal control is the regular control, i.e.,

$$\mathbf{U}^*(t) = \mathbf{U}^{(\text{reg})}(t),$$

at a ***regular or interior point*** *and $\mathbf{U}^{(\text{reg})}(t)$ is called a* ***regular control****, so critical condition (A.8) does not necessarily apply to boundary points or singular points of the control, but certainly does apply to the case of unconstrained control. The associated final conditions are listed in Table A.1 below.*

Proof. The proof is a standard optimization proof in the calculus of variations [40, 15, 163, 44] and is a significant generalization of the usual *first derivative optima*

test. Our formal justification is a brief formulation after Kirk's description [163], but in our notation.

Note that the gradient,

$$\left(\frac{\partial \mathcal{H}}{\partial \mathbf{x}}\right)^* \equiv \nabla_{\mathbf{x}}[\mathcal{H}](\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t), t) = \left[\frac{\partial \mathcal{H}}{\partial x_i}(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t), t)\right]_{n_x \times 1},$$

so is the $\mathbf{x}$-gradient and is a column n_x-vector like $\mathbf{X}$ itself here (elsewhere row vector gradients may be used, e.g., [44]), and so forth, including the gradients of C and $\mathbf{f}$. The triple set (A.6, A.7, A.8) of equations form a set of three vector ordinary differential equations for the optimal trajectory under the optimal control $\mathbf{U}^*(t)$. The first equation (A.6) merely reaffirms the specified state dynamical system (A.1) and that the inclusion with the Lagrange multiplier $\boldsymbol{\lambda}^*(t)$ is proper. The prefix minus on the time-derivative of the Lagrange multiplier in (A.7) indicates that it is a backward-time ODE, in contrast the the forward-time state ODE (HamEqX).

For the calculus of variations, the objective (A.2) is extended in two ways. First, the terminal cost is absorbed in the integral of running costs using the fundamental theorem of calculus,

$$\begin{aligned} S(\mathbf{X}(t_f), t_f) &= S(\mathbf{x}_0, t_0) + \int_{t_0}^{t_f} \frac{dS}{dt}(\mathbf{X}(t), t)dt \\ &= S(\mathbf{x}_0, t_0) + \int_{t_0}^{t_f} \left(\frac{\partial S}{\partial t}(\mathbf{X}(t), t) + \dot{\mathbf{X}}^\top(t)\frac{\partial S}{\partial \mathbf{x}}(\mathbf{X}(t), t)\right) dt, \end{aligned}$$

noting that the initial condition $S(\mathbf{x}_0, t_0)$ is fixed, so can be ignored in the optimization, but the final time t_f will be allowed to be free rather than fixed.

Second, the negative of the state derivative, $-\dot{\mathbf{X}}(t)$, is included in the Lagrange coefficient of the Hamiltonian. Thus, the extended or augmented objective is

$$V^+[\mathbf{Z}, \dot{\mathbf{X}}, t_f] \equiv \int_{t_0}^{t_f} C^+(\mathbf{Z}(t), \dot{\mathbf{X}}(t), t)dt, \tag{A.9}$$

where for brevity an extended state vector is defined as

$$\mathbf{Z}(t) \equiv \begin{bmatrix} \mathbf{X}(t) \\ \mathbf{U}(t) \\ \boldsymbol{\lambda}(t) \end{bmatrix} \tag{A.10}$$

and the extended cost function is

$$C^+(\mathbf{Z}(t), \dot{\mathbf{X}}(t), t) \equiv \mathcal{H}(\mathbf{Z}(t), t) + \frac{\partial S}{\partial t}(\mathbf{X}(t), t) + \dot{\mathbf{X}}^\top(t)\left(\frac{\partial S}{\partial \mathbf{x}}(\mathbf{X}(t), t) - \boldsymbol{\lambda}(t)\right). \tag{A.11}$$

The objective extension also enables the optimal treatment of the final or stopping time t_f when t_f is a free variable.

Next, the variations of the independent variables about potential optima, e.g., $\mathbf{Z}^*(t)$, are introduced,

$$\begin{aligned} \mathbf{Z}(t) &\equiv \mathbf{Z}^*(t) + \delta\mathbf{Z}(t); \\ \dot{\mathbf{X}}(t) &\equiv \dot{\mathbf{X}}^*(t) + \delta\dot{\mathbf{X}}(t); \\ t_f &\equiv t_f^* + \delta t_f, \end{aligned}$$

the latter permitting optimal stopping times t_f^* in addition to free final states for generality. Assuming all variations are small and neglecting higher order variations, i.e., $\mathrm{O}(|\delta\mathbf{Z}(t)|^2)$, a preliminary form of the first variation of the extended objective

$$V^+[\mathbf{Z}, \dot{\mathbf{X}}, t_f] \simeq V^+[\mathbf{Z}^*, \dot{\mathbf{X}}^*, t_f^*] + \delta V^+[\mathbf{Z}, \dot{\mathbf{X}}, t_f]$$

is

$$\delta V^+[\mathbf{Z}, \dot{\mathbf{X}}, t_f] \simeq \int_{t_0}^{t_f^*} \left(\delta\mathbf{Z}^\top \left(\frac{\partial C^+}{\partial \mathbf{z}} \right)^* + \delta\dot{\mathbf{X}}^\top \left(\frac{\partial C^+}{\partial \dot{\mathbf{x}}} \right)^* \right) dt + \delta t_f \; (C^+)^* \Big|_{t=t_f^*},$$

where the latter term derives from a forward approximation of the final integral fragment on $[t_f^*, t_f^* + \delta t_f]$ for small first variation δt_f, ignoring second variations. Also, the shorthand notation such as $(\partial C^+/\partial \mathbf{z})^* = (\partial C^+/\partial \mathbf{z})(\mathbf{Z}^*(t), \dot{\mathbf{X}}^*(t), t)$ has been used.

Since

$$\delta\mathbf{X}(t) = \delta\mathbf{X}(t_0) + \int_{t_0}^{t} \delta\dot{\mathbf{X}}(s) ds,$$

the variation $\delta\dot{\mathbf{X}}(t)$ is not independent of its integral $\delta\mathbf{X}(t)$, but this dependence can be removed by a primary applied mathematics technique of *integration by parts*. So, replacing the objective variation δV^+ by $\delta V^\dagger$ without $\delta\dot{\mathbf{X}}(t)$,

$$\begin{aligned} \delta V^\dagger[\mathbf{Z}, t_f] \simeq \; & \int_{t_0}^{t_f^*} \left(\delta\mathbf{Z}^\top \left(\frac{\partial C^+}{\partial \mathbf{z}} \right)^* - \delta\mathbf{X}^\top \frac{d}{dt} \left(\frac{\partial C^+}{\partial \dot{\mathbf{x}}} \right)^* \right) dt \\ & + \left(\delta t_f \; (C^+)^* + \delta\mathbf{X}^\top \left(\frac{\partial C^+}{\partial \dot{\mathbf{x}}} \right)^* \right) \Big|_{t=t_f^*}. \end{aligned}$$

However, the variation

$$\delta\mathbf{X}(t_f^*) \equiv \mathbf{X}(t_f^*) - \mathbf{X}^*(t_f^*)$$

is only the variation at $t = t_f^*$ and not the total final variation required, which is

$$\delta\widehat{\mathbf{X}}(t_f^*) \equiv \mathbf{X}(t_f^* + \delta t_f) - \mathbf{X}^*(t_f^*),$$

the difference between a final trial value at $t_f = t_f^* + \delta t_f$ and an final optimal state value at the optimal stopping time $t = t_f^*$. By using a tangent line approximation, the former can be converted to the other with sufficient first variation accuracy,

$$\delta\widehat{\mathbf{X}}(t_f^*) \simeq \mathbf{X}(t_f^*) + \dot{\mathbf{X}}(t_f^*)\delta t_f - \mathbf{X}^*(t_f^*) \simeq \delta\widehat{\mathbf{X}}(t_f^*) + \dot{\mathbf{X}}^*(t_f^*)\delta t_f,$$

where $\dot{\mathbf{X}}(t_f^*)\delta t_f \simeq \dot{\mathbf{X}}^*(t_f^*)\delta t_f$ within first variation accuracy. Hence, the proper final first variation $\delta\widehat{\mathbf{X}}(t_f^*)$ with tangent correction can be substituted for $\delta\widehat{\mathbf{X}}(t_f^*)$, yielding

$$\begin{aligned} \delta V^\dagger[\mathbf{Z}, t_f] \simeq \; & \int_{t_0}^{t_f^*} \left(\delta\mathbf{Z}^\top \left(\frac{\partial C^+}{\partial \mathbf{z}} \right)^* - \delta\mathbf{X}^\top \frac{d}{dt} \left(\frac{\partial C^+}{\partial \dot{\mathbf{x}}} \right)^* \right) dt \\ & + \left(\delta t_f \left(C^+ - \left(\dot{\mathbf{X}} \right)^\top \left(\frac{\partial C^+}{\partial \dot{\mathbf{x}}} \right)^* \right) + \delta\widehat{\mathbf{X}}^\top \left(\frac{\partial C^+}{\partial \dot{\mathbf{x}}} \right)^* \right) \Big|_{t=t_f^*}. \end{aligned} \tag{A.12}$$

The *Fundamental Theorem of the Calculus of Variations* [163] states that the first variation, here $\delta V^\dagger[\mathbf{Z}, t_f]$, must vanish for all admissible variations, here assuming $\delta\mathbf{Z}(t)$ is continuous, on an optimal trajectory, here $\mathbf{Z}^*(t)$. Thus,

$$\delta V^\dagger[\mathbf{Z}, t_f] = 0.$$

Further, the *Fundamental Lemma of the Calculus of Variations* [163] states that given a continuous function $\mathcal{F}_i(t)$ and

$$\int_{t_0}^{t_f} \delta X_i(t)\mathcal{F}_i(t)dt = 0$$

for every continuous trajectory $\delta X_i(t)$ on $[t_0, t_f]$, then

$$\mathcal{F}_i(t) = 0$$

on $[t_0, t_f]$. For multidimensional trajectories and independent component variations $\delta X_i(t)$ for $i = 1{:}n_x$, then the result holds for all components.

Using the definition of the extended cost C^+ in (A.11), extended state $\mathbf{Z}$ in (A.10) and the Hamiltonian (A.5) with the first variation $\delta V^\dagger[\mathbf{Z}, t_f]$ in (A.12), we have

- Coefficient of $\delta\boldsymbol{\lambda}^\top(t) \implies$

$$\left(\frac{\partial C^+}{\partial \boldsymbol{\lambda}}\right)^* = \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}}\right)^* - \dot{\mathbf{X}}^*(t) = \mathbf{0} \implies$$
$$\dot{\mathbf{X}}^*(t) = \left(\frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}}\right)^* = \mathbf{f}(\mathbf{X}^*(t), \mathbf{U}^*(t), t) \text{ on } t_0 < t \le t_f.$$

- Coefficient of $\delta\mathbf{X}^\top(t) \implies$

$$\left(\frac{\partial C^+}{\partial \mathbf{x}}\right)^* - \frac{d}{dt}\left(\frac{\partial C^+}{\partial \dot{\mathbf{x}}}\right)^* = \left(\frac{\partial \mathcal{H}}{\partial \mathbf{x}}\right)^* + \dot{\boldsymbol{\lambda}}^*(t) = \mathbf{0} \implies$$
$$\dot{\boldsymbol{\lambda}}^*(t) = -\left(\frac{\partial \mathcal{H}}{\partial \mathbf{x}}\right)^* = -\left(\frac{\partial C}{\partial \mathbf{x}} + \frac{\partial \mathbf{f}^T}{\partial \mathbf{x}}\boldsymbol{\lambda}\right)^*, \text{ on } t_0 \le t < t_f.$$

- Coefficient of $\delta\mathbf{U}^\top(t) \implies$

$$\left(\frac{\partial C^+}{\partial \mathbf{u}}\right)^* = \left(\frac{\partial \mathcal{H}}{\partial \mathbf{u}}\right)^* = \left(\frac{\partial C}{\partial \mathbf{u}} + \frac{\partial \mathbf{f}^T}{\partial \mathbf{u}}\boldsymbol{\lambda}\right)^* = \mathbf{0}, \text{ on } t_0 \le t < t_f.$$

Cautionary Remark: *This critical point result is only valid for isolated, interior critical optima, so it would not be valid for the case that $\mathcal{H}$ is linear in $\mathbf{U}$ or a singular case. However, the equations for $\dot{\mathbf{X}}^*(t)$ and $\dot{\boldsymbol{\lambda}}^*(t)$ remain valid.*

- Coefficient of $\delta t_f \implies$
  If t_f is fixed, then $\delta t_f \equiv 0$ and no information can be implied about the coefficient, else if t_f is free and if $\delta t_f \neq 0$ is otherwise arbitrary then

$$\left((C^+)^* - \left(\dot{\mathbf{X}}^*\right)^\top \left(\frac{\partial C^+}{\partial \dot{\mathbf{x}}}\right)^*\right)\Bigg|_{t=t_f^*} = \left(\mathcal{H}^* + \left(\frac{\partial S}{\partial t}\right)^*\right)\Bigg|_{t=t_f^*} = 0.$$

- Coefficient of $\delta\widehat{\mathbf{X}}^\top(t_f^*) \implies$
  If $\mathbf{X}(t_f)$ is fixed and t_f fixed, then $\delta\widehat{\mathbf{X}}^\top(t_f^*) \equiv 0$ and no information can be implied about the coefficient, else if $\mathbf{X}(t_f)$ is free and t_f is fixed, then $\delta\widehat{\mathbf{X}}^\top(t_f^*) \neq 0$ and

$$\left(\frac{\partial C^+}{\partial \dot{\mathbf{x}}}\right)^*\Bigg|_{t=t_f^*} = \left(\left(\frac{\partial S}{\partial \mathbf{x}}\right)^* - \boldsymbol{\lambda}^*\right)\Bigg|_{t=t_f} = \mathbf{0} \implies$$
$$\boldsymbol{\lambda}^*(t_f) = \frac{\partial S}{\partial \mathbf{x}}(\mathbf{X}^*(t_f), t_f),$$

  or else if both $\mathbf{X}(t_f)$ and t_f are free, then the combined conditions are

$$\boldsymbol{\lambda}_f^{**} \equiv \boldsymbol{\lambda}^*(t_f^*) = \frac{\partial S}{\partial \mathbf{x}}(\mathbf{X}^*(t_f^*), t_f^*),$$
$$\left(\mathcal{H} + \frac{\partial S}{\partial t}\right)^{**} \equiv \left(\mathcal{H}^* + \left(\frac{\partial S}{\partial t}\right)^*\right)\Bigg|_{t=t_f^*} = 0,$$

  the double asterisk notion denoting the optimal stopping time on the optimal path.

The first three items complete the proof of the theorem, while the last two items complete the justifications of the final conditions listed in Table A.1 below. □

The state vector $\mathbf{X}^*(t)$ satisfies specified initial conditions $\mathbf{X}^*(t_0) = \mathbf{x}_0^*$ at t_0. The **final conditions** for the state $\mathbf{X}^*(t_f)$ and **co-state** or **adjoint state** $\boldsymbol{\lambda}^*(t_f)$, if any, depend on the application and a fairly complete set is tabulated in Kirk [163], Bryson-Ho [44] and Athans-Falb [15]. The final conditions depend on whether the final time t_f is fixed (specified) or free (unspecified) and whether the corresponding final state vector $\mathbf{x}_f = \mathbf{X}(t_f)$ is fixed or free. A partial list of some of the conditions is given in Table A.1: See the classical sources of Athans-Falb [15], Kirk [163] and Bryson-Ho [44] for additional final conditions such as moving boundaries $\boldsymbol{\Gamma}(\mathbf{X}(t)) = \mathbf{0}$ or $\boldsymbol{\Gamma}(\mathbf{X}(t), t) = \mathbf{0}$ and other variants that enter into the final conditions. For other variants with more economic interpretations, the bio-economics book of Clark [56] is very readable and useful. Other condition variants include a multitude of mixed and hybrid cases that are vector component combinations of the purely fixed and free vector cases presented in the Table A.1. Some of these final conditions arise as natural boundary conditions because they

Table A.1. *Some final conditions for deterministic optimal control.*

	$\mathbf{X}(t_f) = \mathbf{x}_f$ Fixed	$\mathbf{X}(t_f)$ Free & t_f-Independent
t_f Fixed	$\mathbf{x}_f^* = \mathbf{X}^*(t_f) = \mathbf{x}_f$ at $t = t_f$	$\boldsymbol{\lambda}_f^* = \boldsymbol{\lambda}^*(t_f) = \nabla_{\mathbf{x}}[S](\mathbf{x}_f^*, t_f)$ at $t = t_f$
t_f Free	$\mathbf{x}_f^{**} = \mathbf{X}^*(t_f^*) = \mathbf{x}_f$ $(\mathcal{H} + S_t)_f^{**} = 0$ at $t = t_f^*$	$\boldsymbol{\lambda}_f^{**} = \boldsymbol{\lambda}^*(t_f^*) = \nabla_{\mathbf{x}}[S](\mathbf{x}_f^{**}, t_f^*)$ $(\mathcal{H} + S_t)_f^{**} = 0$ at $t = t_f^*$

Notation: $\mathbf{x}_f^* \equiv \mathbf{X}^*(t_f)$, $\mathbf{u}_f^* \equiv \mathbf{U}^*(t_f)$, $\boldsymbol{\lambda}_f^* \equiv \boldsymbol{\lambda}^*(t_f)$ *and* $\mathcal{H}_f^* \equiv \mathcal{H}(\mathbf{x}_f^*, \mathbf{u}_f^*, \boldsymbol{\lambda}_f^*, t_f)$ *in the case of fixed final time* t_f*, while* $\mathbf{x}_f^{**} \equiv \mathbf{X}^*(t_f^*)$, $\mathbf{u}_f^{**} \equiv \mathbf{U}^*(t_f^*)$, $\boldsymbol{\lambda}_f^{**} \equiv \boldsymbol{\lambda}^*(t_f^*)$ *and* $\mathcal{H}_f^{**} \equiv \mathcal{H}(\mathbf{x}_f^{**}, \mathbf{u}_f^{**}, \boldsymbol{\lambda}_f^{**}, t_f^*)$ *in the case of free final time with optimal final time* t_f^**.*

can not be independently specified but follow from the structure of the optimal control problem by the method of **calculus of variations** [40, 15, 163, 44].

The final conditions for the free terminal time and free terminal state case

$$\boldsymbol{\lambda}_f^{**} = \boldsymbol{\lambda}^*(t_f^*) = \nabla_{\mathbf{x}}[S](\mathbf{x}_f^{**}, t_f^*)\,, \tag{A.13}$$

$$0 = \mathcal{H}(\mathbf{x}_f^{**}, \mathbf{u}_f^{**}, \boldsymbol{\lambda}_f^{**}, t_f^*) + S_t(\mathbf{x}_f^{**}, t_f^*)\,, \tag{A.14}$$

in Table A.1 are a good example of the results the calculus of variations. The equation (A.13) is the final or **transversality** condition for the optimal Lagrange multipier that usually accompanies the stationary point Euler-Lagrange equations (A.7) for the optimal multiplier and (A.8) for the optimal control [44]. The Euler-Lagrange equations along with the dynamic constraint equation and initial condition (A.1) satisfy a two-point boundary value problem, also called a **final-initial** value problem.

Theorem A.2. ***Legendre-Clebsch sufficient conditions for regular control optimum:***
If the Hamiltonian $\mathcal{H}$ *(A.5) has continuous second order derivatives in the control vector* $\mathbf{u}$*, then the* ***sufficient condition*** *for a* ***regular point maximum*** *is that the Hessian matrix must be* ***negative definite****, i.e.,* $\mathcal{H}$ *is* ***concave*** *at the regular point,*

$$\mathcal{H}_{\mathbf{uu}}^* = \nabla_{\mathbf{u}}\left[\nabla_{\mathbf{u}}^\top[\mathcal{H}]\right](\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t), t) < \mathbf{0}\,. \tag{A.15}$$

and the ***sufficient condition*** *for a* ***regular control minimum*** *is that the Hessian matrix must be* ***positive definite****, i.e.,* $\mathcal{H}$ *is* ***convex*** *at the regular control,*

$$\mathcal{H}_{\mathbf{uu}}^* = \nabla_{\mathbf{u}}\left[\nabla_{\mathbf{u}}^\top[\mathcal{H}]\right](\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t), t) > \mathbf{0}\,. \tag{A.16}$$

These sufficient conditions are called the (strengthened) ***Legendre-Clebsch*** *conditions.*

The proof is a standard optimization proof in multivariate calculus ([263, 221, 44]) and is a general form of the so-called *2nd Derivative Optimum Test*.

If the Legendre-Clebsch conditions do not hold, then extra conditions usually are needed. For example, if $\mathcal{H}$ is linear in the control $\mathbf{u}$, then the control problem may be singular [24] and more basic optimization principles may be needed.

See the next section for how to handle some of these exceptions to **regular control** or **normal control** with the critical, stationary condition with respect to the control $\mathbf{u}$ here, using basic optimization principles in terms of a maximum or minimum principle.

Example A.3. ***Regular Control Problem:*** *This problem is a simplified fragment of a financial portfolio application. Let the dynamics be linear in the positive scalar state $X(t) > 0$, denoting the measure of the wealth at time t, but bi-linear in the control-state, such that*

$$\dot{X}(t) \equiv \frac{dX}{dt}(t) = (\mu_0 - U(t))X(t)\,, \quad X(0) = x_0 > 0\,, \quad 0 \le t \le t_f\,, \quad \text{(A.17)}$$

where μ_0 is a fixed mean production rate of the wealth and $U(t)$ is the control variable that is a measure of the rate of consumption of the wealth at time t. The consumption is constrained to be non-negative and bounded above

$$U^{(\min)} = 0 \le U(t) \le U^{(\max)}\,. \quad \text{(A.18)}$$

The objective is to maximize the cumulative utility of instantaneous consumption where the utility is a ***power law****,*

$$C(x,u,t) = u^\gamma/\gamma\,, \quad \text{(A.19)}$$

for positive powers $\gamma > 0$, but in the following analysis we will exclude the linear case $\gamma = 1$ to keep this a regular or normal control problem. In addition, let there be terminal wealth utility using the same power law,

$$S(x,t) = x^\gamma/\gamma\,. \quad \text{(A.20)}$$

Thus, this is a Bolza problem described above, but here the maximum utility is sought rather than the minimum cost. The difference between solving a maximum versus a minimum problem is trivial, as can be seen from the Legendre-Clebsch sufficient conditions, (A.15) and (A.16), with only a difference in the sign of the inequality.
Solution: *The Hamiltonian is then*

$$\mathcal{H}(x,u,\lambda,t) = u^\gamma/\gamma + \lambda(\mu_0 - u)x\,. \quad \text{(A.21)}$$

Hamilton's equations for a ***regular control*** *solution are*

$$\dot{X}^*(t) = +\mathcal{H}^*_\lambda = (\mu_0 - U^{(\mathrm{reg})}(t))X^*(t)\,, \quad \text{(A.22)}$$

$$\dot{\lambda}^*(t) = -\mathcal{H}^*_{\mathbf{x}} = -(\mu_0 - U^{(\mathrm{reg})}(t))\lambda^*(t)\,, \quad \text{(A.23)}$$

$$0 = \mathcal{H}^*_u = (U^{(\mathrm{reg})})^{\gamma-1}(t) - \lambda^*(t)X^*(t)\,, \quad \text{(A.24)}$$

the latter equation yields the regular control,

$$U^{(\mathrm{reg})}(t) = (\lambda^*(t)X^*(t))^{1/(\gamma-1)}\,, \quad \text{(A.25)}$$

provided that $\gamma \neq 1$, as promised, i.e., excluding the what is called the ***neutral risk*** *case. Since the control is a regular control, then, strictly speaking, $X^*(t) = X^{(\mathrm{reg})}(t)$ and $\lambda^*(t) = \lambda^{(\mathrm{reg})}(t)$.*

Before designating the regular control as the optimal control, the Legendre-Clebsch second derivative sufficient conditions are examined:

$$\mathcal{H}_{uu} = (\gamma - 1)u^{\gamma-2}\,, \tag{A.26}$$

it is seen from the Legendre-Clebsch sufficient condition for a maximum, that $\mathcal{H}$ is ***concave*** *or $(\mathcal{H}_{uu})^{(\mathrm{reg})} < 0$, is only satisfied for $\gamma < 1$, the "low"* ***risk adverse*** *case. Hence, $U^*(t) = U^{(\mathrm{reg})}$.*

However, for $\gamma > 1$ and ***risk-seeking*** *utility, the regular control (A.25) yields a minimum since $\mathcal{H}$ is* ***convex*** *or $(\mathcal{H}_{uu})^{(\mathrm{reg})} > 0$, but it would not be rational to get a minimum utility. If maximizing the utility is needed when $\gamma > 1$ then the control constraints must be used. See Exercise 6 for how to obtain the proper maximum utility solution when $\gamma > 1$.*

The first two of Hamilton's equations, though seemingly complicated, can be solved by dividing both sides of the equations and examining them in the phase plane without the time dependence,

$$\frac{dX^*}{d\lambda^*} = -\frac{X^*}{\lambda^*}\,, \tag{A.27}$$

which is just the product rule of differentiation, $(dX^*\lambda^*)/dt = 0$, slightly rearranged and the solution is

$$X^*\lambda^* = K\,, \tag{A.28}$$

where K is a constant of integration. Consequently, our optimal control is the regular control and must be a constant as well,

$$U^*(t) = U^{(\mathrm{reg})} = K^{1/(\gamma-1)} \equiv K_0\,, \tag{A.29}$$

provided $0 \leq U^{(\mathrm{reg})} \leq U^{(\max)}$. Constant control means that the state and co-state equations here are equations of simple exponential growth, so

$$X^*(t) = x_0 e^{(\mu_0 - K_0)t}\,, \tag{A.30}$$

$$\lambda^*(t) = \lambda^*(t_f)e^{-(\mu_0-K_0)(t-t_f)}\,, \tag{A.31}$$

where the constant K_0 and the final adjoint value $\lambda^*(t_f) = \lambda_f^*$ need to be determined. By the transversality condition in Table A.1 for t_f fixed and $X^*(t_f) = x_f^*$ unspecified,

$$\lambda_f^* = S_x(x_f^*, t_f) = (x_f^*)^{\gamma-1} = \left(x_0 e^{(\mu_0-K_0)t_f}\right)^{\gamma-1}\,, \tag{A.32}$$

using the derivative of the terminal utility $S(x,t)$ (A.20) and the state solution $X^*(t)$ in (A.30). Finally, the definitions of K in (A.28) and K_0 in (A.29) yield a nonlinear equation for the control constant $U^*(t) = K_0$ using (A.28-A.32),

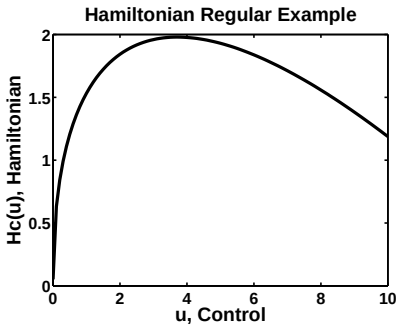
$$K_0 = K^{\frac{1}{\gamma-1}} = (x_f^*\lambda_f^*)^{\frac{1}{\gamma-1}} = (x_f^*)^{\frac{\gamma}{\gamma-1}} = \left(x_0 e^{(\mu_0-K_0)t_f}\right)^{\frac{\gamma}{\gamma-1}}\,, \tag{A.33}$$

in terms of the specified X_0, μ_0 and $\gamma < 1$.

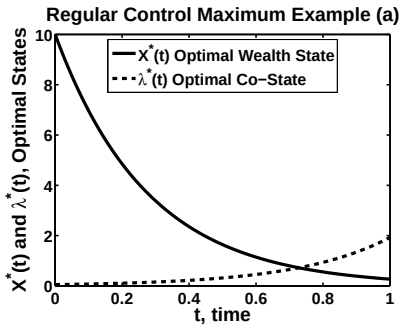
We are assuming that the control constraint $U^{(\max)}$ is sufficiently larger than K_0, so that the control remains regular. Control constraint violations, bang control and linear or singular control are treated in the next section.

Some sample optimal wealth state $X^*(t)$ and co-state $\lambda^*(t)$ solutions are displayed in Fig. A.1 for maximum utility with $\gamma = 0.5$ in Subfigure A.1(b).The terminal wealth at the terminal time $t_f = 1.0$ starting from $x_0 = 10.0$ is $S = 1.038$ for $\gamma = 0.5$. The mean production rate was $\mu_0 = 0.10$ or 10% in absence of consumption. MATLAB's modification of Brent's zero finding algorithm `fzero` [87] is used to find the control constant $U^*(t) = K_0$ whose approximate value is 3.715 when $\gamma = 0.5$ to accuracy of order 10^{-15} in satisfying (A.33).

For completeness and to provide a contrasting illustration with a non-regular, bang control case for a power utility with $\gamma = 2.0$, the Hamiltonian and optimal paths are displayed in Subfigures A.2(a)-A.2(b), respectively. The control constant $U^*(t)$ has an approximate value of 10.0 when $\gamma = 2.0$. The terminal wealth is $S = 5.02e\text{-}4$ at the terminal time $t_f = 1.0$ starting from $x_0 = 10.0$ for $\gamma = 2.0$. See Exercise 6 for obtaining proper maximum utility solution when $\gamma > 1$.



(a) Hamiltonian for regular maximum utility example for power $\gamma = 0.5$.



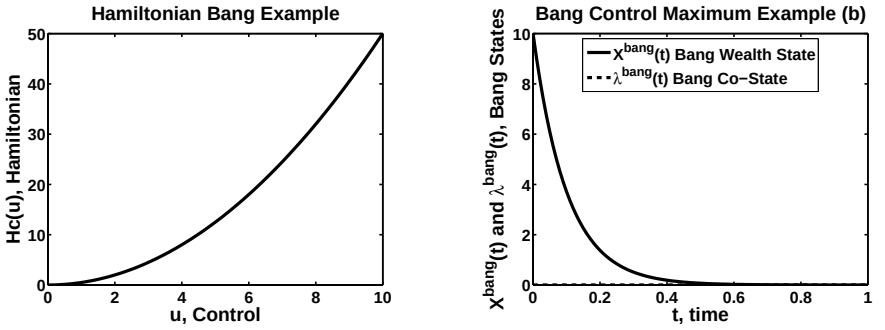
(b) Optimal paths for regular maximum utility example for power $\gamma = 0.5$.

Figure A.1. *Hamitonian and optimal solutions for regular control problem example from (A.30) for $X^*(t)$ and (A.31) for $\lambda^*(t)$. Note that the $\gamma = 0.5$ power utility is only for illustration purposes.*

Remark A.4. *Many control problems are not this easy, since they may require much more analysis, especially in multiple dimensions, and often numerical approximation is needed. For more information on optimal finance portfolios with consumption, see Section 10.4 in Chapter 10 on financial applications.*

A.1.1 Deterministic Computation and Computational Complexity

Except for simple or analytical homework problems, usually numerical discretization and iterations are required until the solution $(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t))$ converges to some



(a) Hamiltonian for endpoint maximum utility example for power $\gamma = 2.0$.

(b) Optimal paths for endpoint maximum utility example for power $\gamma = 2.0$.

Figure A.2. *Hamiltonian and optimal solutions for bang control problem example from (A.30) for $X^*(t)$ and (A.31) for $\lambda^*(t)$. Note that the $\gamma = 2.0$ power utility is only for illustration purposes.*

prescribed accuracy. If there are n_t discrete time nodes, $T_k = t_0 + (k-1)\Delta T$ for $k = 1 : N_t$ with $\Delta T = (t_f - t_0)/(N_t - 1)$, then the n_x dimensional state vector $\mathbf{X}^*(t)$ is discretized into $\mathbf{X}^*(T_k) = \mathbf{X}_k = [X_{i,k}]_{n_x \times N_t}$ or $n_x \cdot N_t$ discrete variables. For the three vector solution the **computational complexity** or the order of the **computational cost** [110] is

$$CC(n_x, n_t) = O(3n_x \cdot N_t) \tag{A.34}$$

per iteration, i.e., bi-linear in the dimension and number of time nodes, a very manageable computational problem, even for today's powerful personal computers.

In addition, MATLAB™ [210] has a good number of control Toolboxes to handle problems. There are also several good on-line tutorials available, such as Tilbury and Messner's [268, 205] *Control Tutorials for MATLAB And Simulink*.

Some early surveys on computational methods for optimal control problems are by Larson [182], Dyer and McReynolds [76], and Polak [227].

A.2 Optimum Principles: The Basic Principles Approach

For many problems, as discussed in Section B.15 of Chapter B of preliminaries, the unconstrained or regular control conditions expressed by Hamilton's equations (A.6, A.7, A.8) are in general inadequate. The inadequacy arises in problems for which the optima are not located at interior points but are located at the boundaries of the state and control domains, such as when the domains have bounded constraints in addition to dynamical constraints like (A.1). One exceptional case is the linear control problem. Another exception is when the optima are at interior points at which the derivatives in Hamilton's equations cease to exist, or any of the multitude of combinations of these exceptions depending on all or a subset of the components

of the variables involved.

Basic Optimum Principle: *Hence, for general optimization theory and its application, it is essential to return to basic optimization principles, that the* ***global minimum is the smallest*** *or that the* ***global maximum is the biggest****.*

Example A.5. ***Simple static example of state dependent control with quadratic costs and control constraints****:*
Consider the following static quadratic cost function with scalar control u *and state* x*,*

$$\mathcal{H}(x,u) = C(x,u) = 2 + x + \frac{1}{2}x^2 - xu + \frac{1}{2}u^2 = 2 + x + \frac{1}{2}(u-x)^2 \,, \quad \text{(A.35)}$$

with control constraints,

$$-1 \leq u \leq +1 \,, \quad \text{(A.36)}$$

but without any dynamical constraints like (A.1). The objective is to find the optimal control law and optimal cost.
Solution: *The control gradient or derivative is*

$$\frac{\partial C}{\partial u}(x,u) = -x + u \,,$$

yielding the critical, stationary point with respect to the control, called a ***regular control*** *in control theory,*

$$U^{(\mathrm{reg})}(x) = x \,,$$

which would be the global minimum in absence of control constraints since the second partial with respect to the control is positive, $C_{uu}(x,u) = +1 > 0$ *with corresponding regular cost*

$$C^{(\mathrm{reg})}(x) \equiv C(x, u^{(\mathrm{reg})}(x)) = 2 + x \,,$$

that is linear (affine) in the state variable.

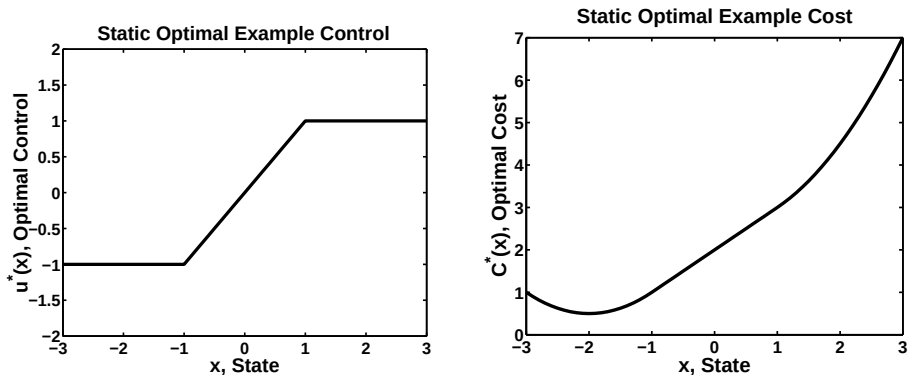
However, this example has control constraints (A.36) which forces the correct optimal control to assume the constrained values when the regular control goes beyond those constraints, i.e.,

$$U^*(x) = \left\{ \begin{array}{cc} -1, & x \leq -1 \\ x, & -1 \leq x \leq +1 \\ +1, & +1 \leq x \end{array} \right\} \,. \quad \text{(A.37)}$$

This type of optimal control could be called a ***bang-regular-bang*** *control, where the term* ***bang*** *signifies hitting the control constraints, the control boundaries becoming active. The corresponding correct optimal cost is*

$$C^*(x) = C(x,u^*(x)) = \left\{ \begin{array}{cc} 2 + x + \frac{1}{2}(x+1)^2, & x \leq -1 \\ 2 + x, & -1 \leq x \leq +1 \\ 2 + x + \frac{1}{2}(x-1)^2, & +1 \leq x \end{array} \right\} \,. \quad \text{(A.38)}$$

For this example, $C^*(x)$ is continuous and continuously differentiable, but not twice continuously differentiable. However, optimal controls and optimal costs of some problems can have much worse analytic properties. The optimal solution (A.38) for this simple, somewhat artificial, static optimal control problem is illustrated in Fig. A.3 with the optimal control in Subfigure A.3(a) and the optimal cost in Subfigure A.3(b). This simple example provides an example motivating why the stationary optimality condition (A.8) for the optimal control is not generally valid.



(a) Optimal control $u^*(x)$. (b) Optimal cost $C^*(x)$.

Figure A.3. *Optimal solutions for a simple, static optimal control problem represented by (A.35) and (A.36), respectively.*

The **basic optimum principle** is just the underlying principle for optimization, but the rigorous justification is beyond the scope of this text. In control theory the optimum principle is associated with the name **Pontryagin maximum principle** [226] in the Russian literature, where the Hamiltonian is formed with an extra multiplier λ_0 to include the objective functional as the 0th dynamical constraint

$$\dot{X}_0(t) = C(X(t), U(t), t),$$

so the maximum refers to the Hamiltonian when the objective is minimum costs and λ_0 must be non-positive (see also (A.39) below). Often the optimum principle is called the **minimum principle** in the English literature [163, 44, 258], particularly when dealing with minimum cost problems, though not exclusively. The difference between a maximum and a minimum principle is essentially a difference in the sign of the Hamiltonian and the fact that the conversion from a maximum objective to a minimum objective problem is quite simple:

$$\max_u[F(u)] = -\min_u[-F(u)]\,. \tag{A.39}$$

With regard to applications, which version of the optimum principle is used depends on the whether the optimal objective is minimum costs or maximum profit.

minimum energy or maximum energy, minimum time or maximum speed, and there are many other objective choices:

- Minimum time ($C = 1$ and $S = 0$).
- Minimum control ($C = |u|$ and $S = 0$).
- Minimum fuel ($C = |u|$, i.e., thrust measure of fuel consumption, and $S = 0$).
- Minimum energy ($C = u^2$, i.e., energy, and $S = 0$).
- Minimum net profit ($C = p_0 X - c_0$, i.e., profit less cost, and $S = 0$).
- Maximum utility of consumption ($C = \mathcal{U}(u)$, i.e., utility of consumption, and $S = \mathcal{U}(x)$, i.e., utility of portfolio wealth).
- Maximum thrust angle ($C = sin(\theta(t))$ and $S = 0$).
- Minimum distance.
- Minimum surface area.

Here, the maximum and minimum principles are only stated, but see the references at the end of the chapter for more information, such as Anderson-Moore [8], Athans-Falb [15], Bryson-Ho [44], Kirk [163], Pontryagin et al. [226] and Bell-Jacobson [24]. While the statement of the principle seems very intuitive, the rigorous proof is far from easy.

Theorem A.6. ***Optimum Principles:***
The ***necessary condition for a maximum*** *or* ***maximum principle****, is*

$$\mathcal{H}^* = \mathcal{H}(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t), t) \geq \mathcal{H}(\mathbf{X}^*(t), \mathbf{U}(t), \boldsymbol{\lambda}^*(t), t), \tag{A.40}$$

but the ***necessary condition for a minimum*** *or* ***minimum principle*** *is*

$$\mathcal{H}^* = \mathcal{H}(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t), t) \leq \mathcal{H}(\mathbf{X}^*(t), \mathbf{U}(t), \boldsymbol{\lambda}^*(t), t)\,, \tag{A.41}$$

in general replacing (A.8), where $\mathbf{X}^*(t)$ and $\boldsymbol{\lambda}^*(t)$ are candidates for optimal state or co-state, respectively. The optimal state $\mathbf{X}^*(t)$ must satisfy the dynamical constraint $\dot{\mathbf{X}}^*(t) = (\mathcal{H}_{\boldsymbol{\lambda}})^*$ (A.6) and the co-state $\boldsymbol{\lambda}^*(t)$ must satisfy the co-state equation $\dot{\boldsymbol{\lambda}}^*(t) = -(\mathcal{H}_{\mathbf{x}})^*$ (A.7). The optimal control $\mathbf{U}^*(t)$ is the argument of the corresponding maximum in (A.40) or minimum in (A.41).

Remarks A.7.

- *Note that the optimal principles (A.40) and (A.41) as in the basic optimizing principles are used as a general replacement for the necessary conditions for a regular point $\mathcal{H}_u^* = \mathbf{0}$ (A.8) and the Legendre-Clebsch second order sufficient conditions $\mathcal{H}_{uu}^* < 0$ (A.15) for a maximum and (A.16) $\mathcal{H}_{uu}^* > 0$ for a minimum. However, these first and second order derivative conditions are still valid for interior or regular points.*

- *In fact, Pontryagin et al. [226] justify briefly that the optimum principles are sufficient conditions as they are more basic conditions.*

- *If we let the control perturbation be*

$$\delta\mathbf{U}(t) \equiv \mathbf{U}(t) - \mathbf{U}^*(t)\,, \tag{A.42}$$

then the corresponding perturbation or variation in the Hamiltonian is

$$\begin{aligned}\Delta_u\mathcal{H}(\mathbf{X}^*(t),\mathbf{U}^*(t),\boldsymbol{\lambda}^*(t),t) \equiv \mathcal{H}(\mathbf{X}^*(t),\mathbf{U}^*(t)+\delta\mathbf{U}(t),\boldsymbol{\lambda}^*(t),t) \\ - \mathcal{H}(\mathbf{X}^*(t),\mathbf{U}^*(t),\boldsymbol{\lambda}^*(t),t)\end{aligned} \tag{A.43}$$

and the maximum principle can be reformulated as

$$\Delta_u\mathcal{H}(\mathbf{X}^*(t),\mathbf{U}^*(t),\boldsymbol{\lambda}^*(t),t) \leq 0\,, \tag{A.44}$$

while the minimum principle can be reformulated as

$$\Delta_u\mathcal{H}(\mathbf{X}^*(t),\mathbf{U}^*(t),\boldsymbol{\lambda}^*(t),t) \geq 0\,. \tag{A.45}$$

In the language of the calculus of variations, the optimum principles are that the first variation of the Hamiltonian is negative semi-definite for a maximum, while it is positive semi-definite for a minimum.

- *Concerning the simple static example A.5, the perturbation form of the minimum principle (A.45) can be used to justify the choice of the bang controls given in (A.37). The perturbation for the example is*

$$\Delta_u\mathcal{H}^* = (U^* - x)\delta U^* + \frac{1}{2}(\delta U^*)^2\,,$$

where only the linear term need be considered for its contribution to the non-negativity of the perturbation since the quadratic term is never negative. When there in minimal bang control, $U^* = -1$, then the perturbation δU^* must necessarily be non-negative, otherwise the control constraints (A.36) would be violated, so for non-negativity of the Hamiltonian perturbation the control perturbation coefficient $(-1-x)$ must also be non-negative or that $x \leq -1$. Similarly, when there is maximal bang control, $U^* = +1$, then the perturbation has to be non-positive, $\delta U^* \leq 0$, to avoid violating the control constraints. So $\Delta_u\mathcal{H}^* \geq 0$ (A.45) implies that the coefficient $(1-x)$ of δU^* must be non-positive or that $x \geq +1$.

- *Similar techniques work with the application of the optimum principles to the case where the Hamiltonian is linear in the control. For example, consider the scalar, **linear control** Hamiltonian,*

$$\mathcal{H}(x,u,\lambda,t) = C_0(x,t) + C_1(x,t)u + \lambda(F_0(x,t) + F_1(x,t)u)\,,$$

subject to control constraints,

$$U^{(\min)} \leq U(t) \leq U^{(\max)}\,,$$

and such that

$$\mathcal{H}_u(x,u,\lambda,t) = C_1(x,t) + \lambda F_1(x,t) = \mathcal{H}_u(x,0,\lambda,t)\,,$$

so no regular control exists. However, the perturbed Hamiltonian has the form,

$$\Delta_u\mathcal{H}(X^*,U^*,\lambda^*,t) = \mathcal{H}_u(X^*,0,\lambda^*,t)\delta U^*\,,$$

so optimal control is of the ***bang-bang*** *form, which for a minimum of $\mathcal{H}$ using $\Delta_u\mathcal{H} \geq 0$ yields the composite form,*

$$U^*(t) = \left\{ \begin{array}{ll} U^{(\min)}, & (\mathcal{H}_u)^* = C_1(X^*(t),t) + \lambda^*(t)F_1(X^*(t),t) > 0 \\ U^{(\max)}, & (\mathcal{H}_u)^* = C_1(X^*(t),t) + \lambda^*(t)F_1(X^*(t),t) < 0 \end{array} \right\}, \quad \text{(A.46)}$$

since for $(\mathcal{H}_u)^* > 0$ then $\delta U^* \geq 0$ or equivalently $U^*(t) = U^{(\min)}$. Similarly when $(\mathcal{H}_u)^* < 0$ then $\delta U^* \leq 0$ or equivalently $U^*(t) = U^{(\max)}$, but if $(\mathcal{H}_u)^* = 0$ no information on either δU^* or $U^*(t)$ can be determined.

Example A.8. ***Bang-Bang Control Problem:*** *Consider a simple lumped model of a leaky reservoir (after Kirk [163]) given by*

$$\dot{X}(t) = -aX(t) + U(t), \quad X(0) = x_0,$$

where $X(t)$ is the depth of the reservoir, $U(t)$ is the net inflow of water at time t and $a > 0$ is the rate of leakage as well as usage. The net inflow is constrained pointwise $0 \leq U(t) \leq M$ for all $0 < t \leq t_f$ and also cumulatively by

$$\int_0^{t_f} U(t)dt = K > 0, \quad \text{(A.47)}$$

where K, M and t_f are fixed constants, such that $K \leq M \cdot t_f$ for consistency. Find the optimal control law that maximizes the cumulative depth,

$$J[X] = \int_0^{t_f} X(t)dt$$

and optimal depth $X^*(t)$.

Solution: *The extra integral condition (A.47) presents a variation on our standard control problem, but can be treated nicely by extending the state space letting $X_1(t) = X(t)$ and $\dot{X}_2(t) = U(t)$ starting at $X_2(0) = 0$, so that $X_2(t_f) = K$ is precisely the constraint (A.47). Thus, the Hamiltonian is*

$$\mathcal{H}(x_1,x_2,u,\lambda_1,\lambda_2,t) = x_1 + \lambda_1(-ax_1 + u) + \lambda_2 u, \quad \text{(A.48)}$$

where λ_1 and λ_2 are Lagrange multipliers. The Hamilton equations for the optimal state and co-state solutions are

$$\begin{aligned} \dot{X}_1^*(t) &= \mathcal{H}_{\lambda_1}^* = -aX_1^*(t) + U^*(t), \quad X_1^*(0) = x_0; \\ \dot{X}_2^*(t) &= \mathcal{H}_{\lambda_2}^* = U^*(t), \quad X_2^*(0) = 0; \\ \dot{\lambda}_1^*(t) &= -\mathcal{H}_{x_1}^* = -1 + a\lambda_1^*(t); \\ \dot{\lambda}_2^*(t) &= -\mathcal{H}_{x_2}^* = 0. \end{aligned}$$

Consequently, $\lambda_2^*(t) = C_2$*, a constant, and* $X_2^*(t_f) = K$ *is fixed. Also,* $\lambda_1^*(t) = C_1 \exp(at) + 1/a$ *with the constant determined from the transversality condition* $\lambda_1^*(t_f) = 0$ *of Table A.1 with* $X_1^*(t_f)$ *free and no terminal cost, i.e.,* $S(x) \equiv 0$*, so* $C_1 = -\exp(-at_f)/a$ *and*

$$\lambda_1^*(t) = \frac{1}{a}\left(1 - e^{-a(t_f - t)}\right). \tag{A.49}$$

Since

$$\mathcal{H}_u^* = \lambda_1^*(t) + \lambda_2^*(t) \neq 0$$

in general, the usual critical point condition will not directly produce an optimal control $U^*(t)$*, but a bang-bang control will work. By applying the essential Pontryagin maximum principle (first derivative test) in the form (A.43-A.44) with* $\delta U(t) = U(t) - U^*(t)$*,*

$$\Delta_u \mathcal{H}(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t), t) = (\lambda_1^*(t) + \lambda_2^*(t))(U(t) - U^*(t)) \leq 0,$$

so if $(\lambda_1^*(t) + \lambda_2^*(t)) > 0$ *then* $U(t) - U^*(t) \leq 0$ *and* $U^*(t) = \max[U(t)] = M$*, but if* $(\lambda_1^*(t) + \lambda_2^*(t)) < 0$ *then* $U(t) - U^*(t) \geq 0$ *and* $U^*(t) = \min[U(t)] = 0$*. If* $(\lambda_1^*(t) + \lambda_2^*(t)) = 0$*, then* $U^*(t)$ *cannot be determined. Now,* $U^*(t)$ *can not be zero on all of* $[0, t_f]$ *or be* M *on all of* $[0, t_f]$*, because both options would violate the constraint (A.47) in the strict case* $K < M \cdot t_f$*. In this case and noting that* $\lambda_1^*(t)$ *is decreasing in time, there must be a switch time* t_s *on* $[0, t_f]$ *such that* $\lambda_1^*(t_s) + \lambda_2^*(t_s) = 0$*,* $C_2 = \lambda_2^*(t_s) = -\lambda_1^*(t_s) = -(1 - \exp(-a(t_f - t_s))/a < 0$ *and*

$$X_2^*(t_f) = K = \int_0^{t_s} M dt + \int_{t_s}^{t_f} 0 dt = M t_s,$$

so $t_s = K/M$*. The composite bang-bang control law is then*

$$U^*(t) = \left\{ \begin{array}{cc} M, & 0 \leq t < t_s \\ 0, & t_s < t \leq t_f \end{array} \right\}, \tag{A.50}$$

and the corresponding state trajectory is given by

$$X_1^*(t) = X^*(t) = x_0\ e^{-at} + \frac{M}{a}\left\{ \begin{array}{cc} (1 - e^{-at}), & 0 \leq t \leq t_s \\ e^{-at}\,(e^{+at_s} - 1), & t_s < t \leq t_f \end{array} \right\}. \tag{A.51}$$

The optimal control (A.50), state (A.51) and the switch time indicator multiplier sum (A.49), $\lambda_1^*(t) + \lambda_2^*(t)$*, are plotted together in Fig. A.4 with sample numerical parameter values.*

Example A.9. ***Singular Control Problem:*** *Consider the scalar dynamical system for a natural resource with state or mass* $X(t)$

$$\dot{X}(t) \equiv \frac{dX}{dt}(t) = (\mu_0 - U(t))X(t)\,, \quad X(t_0) = x_0 > 0\,, \quad t_0 \leq t \leq t_f\,, \tag{A.52}$$

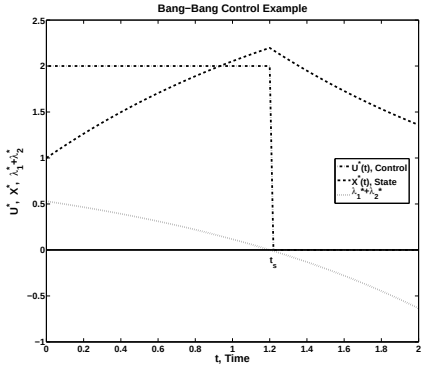


Figure A.4. *Optimal control, state and switch time multiplier sum are shown for bang-bang control example with sample parameter values* $t_0 = 0$, $t_f = 2.0$, $a = 0.6$, $M = 2$, $K = 2.4$ *and* $x_0 = 1.0$. *The computed switch time* t_s *is also indicated.*

where μ_0 *is the natural growth rate and* $U(t)$ *is the harvest rate or effort that will be taken as the control variable. Thus, (A.52) represents exponential growth of the resource whose growth rate is modified by the control. Let the running "cost" for the objective functional be*

$$C(x, u, t) = e^{-\delta_0 t} \max\left[p_0 x - c_0, 0\right] u(t)\,, \tag{A.53}$$

where $p_0 > 0$ *is the fixed price per unit effort per unit mass and* $c_0 > 0$ *is the fixed cost per unit effort, so* $p_0 X(t) - c_0$ *is the net instantaneous profit at time* t.

Note that only positive profit is considered to avoid the possibility of loss, so $X(t) > c_0/p_0$ *needs to be maintained. Since the objective concerns profit rather that costs, the objective will be the maximization of profit and the maximum version of the optimum principle is applicable here. The factor* $\delta_0 > 0$ *is the fixed discount rate or time value of money, but* $\delta_0 > \mu_0$ *is also assumed as a result of the analysis. There is no terminal cost* S. *Since real applications have constraints, let the control domain be defined by*

$$0 \le U(t) \le U^{(\max)}\,, \tag{A.54}$$

where $U^{(\max)}$ *is positive but whose value is left open for the moment. Since the dynamics are linear and the initial condition is positive, the state domain will also be positive values* $X(t) > 0$.

Solution: *In order to find the solution, the Hamiltonian is written*

$$\mathcal{H}(x, u, \lambda, t) = C(x, u, t) + \lambda \dot{X} = e^{-\delta_0 t}(p_0 x - c_0)u + \lambda(\mu_0 - u)x\,,$$

assuming a positive profit. Prior to applying basic optimization principles, we first seek critical, stationary solutions in the control dependence. The control derivative

is

$$\mathcal{H}_u(x,u,\lambda,t) = e^{-\delta_0 t}(p_0 x - c_0) - \lambda x \,, \tag{A.55}$$

which is independent of the control u and when set to zero for stationarity yields the optimal candidate for the adjoint variable, say,

$$\widehat{\lambda}(t) = e^{-\delta_0 t}\left(p_0 - c_0/\widehat{x}(t)\right) \,.$$

However, the other Hamilton's equations specify the potential optimal dynamics of the adjoint and state variables,

$$\dot{\widehat{\lambda}}(t) = -\mathcal{H}_x = -e^{-\delta_0 t} p_0 \widehat{u}(t) - \widehat{\lambda}(t)(\mu_0 - \widehat{u}(t)) \,, \tag{A.56}$$

$$\dot{\widehat{x}}(t) = \mathcal{H}_\lambda = (\mu_0 - \widehat{u}(t))\widehat{x}(t) \,. \tag{A.57}$$

So, combining the last three equations, it is found that the control terms $\widehat{u}$ cancel out exactly. Consequently, this yields a ***singular solution*** *for the state,*

$$X^{(\mathrm{sing})} = \widehat{x}(t) = (c_0/p_0)/(1 - \mu_0/\delta_0) \,. \tag{A.58}$$

This singular solution leads to the requirement that $\delta_0 > \mu_0$ to maintain the profit restriction that $X(t) > c_0/p_0$. Note that the singular solution in this case is also a constant. The solution (A.58) is called a singular solution, rather that a regular or normal solution, since (A.55) does not define a stationary point or regular control and by the way the control cancels out due to the linear dependence on control. However, the singular control can be recovered from inverting the state dynamics,

$$U^{(\mathrm{sing})} = \widehat{u}(t) = \mu_0 - \dot{X}^{(\mathrm{sing})}/X^{(\mathrm{sing})} = \mu_0 \,.$$

For the optimal solution, the control constraints and the initial condition $X(0) = x_0$ need to be considered.

If $U^{(\max)} \geq \mu_0$, then $U^*(t) = U^{(\mathrm{sing})} = \mu_0$ and $X^*(t) = X^{(\mathrm{sing})}$ on $0 < t^* \leq t \leq T_0^{(\max)}$ where $T_0^{(\max)}$ is a transition time where the initial trajectory connects to the singular trajectory at point that is called a ***corner****. The initial trajectory must be chosen using the control bound that allows the singular path to be reached and this control trajectory could be called a* ***bang control*** *trajectory.*

If $X^{(\mathrm{sing})} < x_0$ and $U^{(\max)} > \mu_0$, then $U^*(t) = U^{(\max)}$ on $[0, T_0^{(\max)}]$ where the maximal state trajectory starting from x_0 at $t = 0$ integrating (A.57) is

$$X_0^{(\max)}(t) = x_0 \exp\left(\left(\mu_0 - U^{(\max)}\right)t\right) \,, \quad 0 \leq t \leq T_0^{(\max)} \,,$$

$$T_0^{(\max)} = -\frac{\ln\left(X^{(\mathrm{sing})}/x_0\right)}{\left(U^{(\max)} - \mu_0\right)} > 0 \,.$$

If $X^{(\mathrm{sing})} > x_0$, then $U^*(t) = 0$ on $[0, T_0^{(\min)}]$ where the minimal state trajectory starting from x_0 at $t = 0$ integrating (A.57) is

$$X_0^{(\min)}(t) = x_0 e^{\mu_0 t} \,, \quad 0 \leq t \leq T_0^{(\min)} \,,$$

$$T_0^{(\min)} = +\frac{\ln\left(X^{(\mathrm{sing})}/x_0\right)}{\mu_0} > 0 \,.$$

At the final time the adjoint final or transversality condition must be used as a final value condition for the adjoint dynamics (A.56), which from the scalar version of the entry for fixed t_f and free $X(t_f)$ in Table A.1 on page A8 is

$$\lambda^*(t_f) = S_x(\mathbf{x}_f^*, t_f) = 0\,,$$

since there is no terminal value S in this example. Note that this is consistent with the maximum principle using the calculus of variation and that the regular, critical relation $\mathcal{H}_u = 0$ can not be used as it was for the singular path. Obviously, it is necessary to use the maximal control in (A.56) to reach the condition $\lambda^*(t_f) = 0$ from the singular path,

$$\lambda^{(\mathrm{sing})}(t) = e^{-\delta_0 t} p_0 \mu_0 / \delta_0\,,$$

since that leads to a positive running cost and the minimum control cannot be used to physically reach $\lambda^*(t_f) = 0$. Letting $\lambda_f(t) = \widehat{\lambda}(t)$ be the solution of the adjoint dynamics equation (A.56) with conditions $\widehat{\lambda}(T) = 0$ and connection or corner time T_f such that $\lambda_f(T_f) = \widehat{\lambda}(T_f) = \lambda^{(\mathrm{sing})}(T_f)$, thus

$$T_f = t_f + \frac{\ln\left(1 - \mu_0(\delta_0 + U^{(\max)} - \mu_0)/(\delta_0 U^{(\max)})\right)}{(\delta_0 + U^{(\max)} - \mu_0)}\,.$$

Given the value of T_f, the corresponding state trajectory is

$$X_f(t) = X^{(\mathrm{sing})} e^{-(U^{(\max)} - \mu_0)(t - T_f)}\,,$$

on $[T_t, t_f]$.

Thus, the composite optimal control might be called ***bang-singular-bang*** *with the form*

$$U^*(t) = \left\{ \begin{array}{ll} \left\{ \begin{array}{ll} U^{(\max)}, & 0 \le t \le T_0^{(\max)} \\ U^{(\mathrm{sing})} = \mu_0, & T_0^{(\max)} \le T_f \end{array} \right\}, & x_0 > X^{(\mathrm{sing})} \\ \left\{ \begin{array}{ll} 0, & 0 \le t \le T_0^{(\min)} \\ U^{(\mathrm{sing})} = \mu_0, & T_0^{(\min)} \le T_f \end{array} \right\}, & x_0 < X^{(\mathrm{sing})} \\ U^{(\max)}, & T_f \le t \le t_f \end{array} \right\} \quad \text{(A.59)}$$

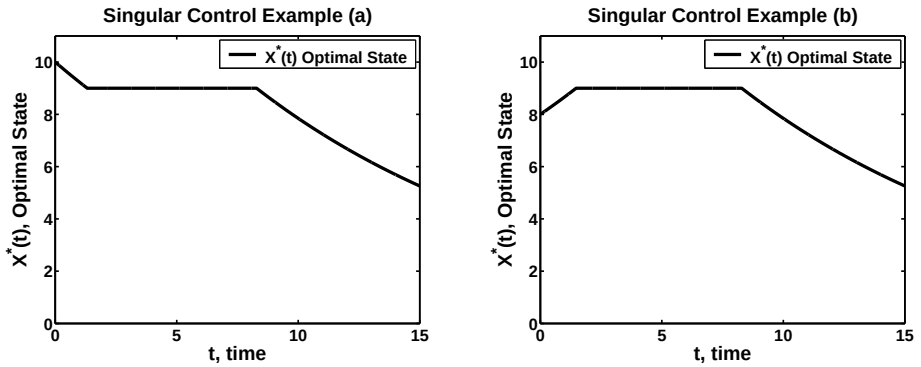
and a composite optimal state trajectory is

$$X^*(t) = \left\{ \begin{array}{ll} \left\{ \begin{array}{ll} X_0^{(\max)}(t), & 0 \le t \le T_0^{(\max)} \\ X^{(\mathrm{sing})} & T_0^{(\max)} \le T_f \end{array} \right\}, & x_0 > X^{(\mathrm{sing})} \\ \left\{ \begin{array}{ll} X_0^{(\min)}(t), & 0 \le t \le T_0^{(\min)} \\ X^{(\mathrm{sing})}, & T_0^{(\min)} \le T_f \end{array} \right\}, & x_0 < X^{(\mathrm{sing})} \\ X_f(t), & T_f \le t \le t_f \end{array} \right\}, \quad \text{(A.60)}$$

where it has been assumed for both $U^*(t)$ in (A.59) and $X^*(t)$ in (A.60) that $T_0^{(\min)} < T_f$ or $T_0^{(\max)} < T_f$ so that there is a non-trivial singular path. Thus, the possibility of a pure ***bang-bang*** *control is excluded, for example when a minimal bang path $X_0^{(\min)}(t)$ from x_0 intersects the maximal bang path $X_f(t)$ from x_f^* before hitting the singular path $X^{(\text{sing})}$.*

Note that this solution is for the case, $U^{(\max)} > \mu_0$, the case for $U^{(\max)} \leq \mu_0$ is left as an open problem in Exercise 7 for the reader, who should realize that some parameter values fail to lead to a control problem solution. One possible reason for this failure is the realistic assumption that the control is bounded does not allow the state to jump from the initial condition to the singular path and unbounded control that could do that is called ***impulse control****. Impulse control could be implemented as a Dirac delta function in the differential equation and more on this matter and similar examples can be found in Clark [56] and Bryson-Ho [44].*

Some sample results for this singular control example are displayed in Fig. A.5 using model parameters $\mu_0 = 0.08$, $\delta_0 = 0.144$, $p_0 = 5.0$, $c_0 = 12.0$, $t_0 = 0$ and $t_f = 15.0$. In Subfig. A.5(a) the optimal state trajectory starts out from $x_0 = 10.0$ at $t = 0$ using maximal bang control with $U^{(\max)} = 0.16$ moving down to reach the singular path at $X^{(\text{sing})} = 9.0$ below when $T_0^{(\max)} = 1.317$, proceeding along the singular path until the singular-bang final corner when $T_f = 8.285$ and then moving down the maximal bang path using $U^{(\max)}$ until reaching the end of the time horizon at $t = t_f = 15.0$. The trajectory displayed in Subfig. A.5(b) is similar except it starts at $x_0 = 8.0$ and moves up to the singular path until reaching the singular path at $(X^{(\text{sing})}, T_0^{(\min)}) = (9.0, 1.472)$; the rest of the path is the same for this example as for the maximal initial bang trajectory.



(a) Singular control optimal state $X^*(x)$ when $x_0 = 10.0$.

(b) Singular control optimal state $X^*(x)$ when $x_0 = 8.0$.

Figure A.5. *Optimal state solutions for singular control example leading to a bang-singular-bang trajectory represented by (A.60). Subfigure (a) yields a maximal bang trajectory from x_0 using $U^{(\max)}$, where as Subfigure (b) yields a minimal bang trajectory from x_0 using $U^{(\min)}$.*

A.3 Linear Quadratic (LQ) Canonical Models

The linear dynamics, quadratic costs or (LQ) problem has the advantage that the regular control can be found fairly explicitly in terms of the state or the co-state, thus avoiding the singular complications of linear control problems.

A.3.1 Scalar, Linear Dynamics, Quadratic Costs (LQ)

In the scalar, constant coefficient case the linear dynamics is given by

$$\dot{X}(t) = a_0 X(t) + b_0 U(t)\,, \quad t_0 \le t \le t_f\,, \quad X(t_0) = x_0 \neq 0\,, \tag{A.61}$$

where $a_0 \neq 0$ and $b_0 \neq 0$ are assumed so the dynamics is assumed to be non-trivial. The quadratic cost objective is given by

$$V[X,U,t_f](x_0,t_0) = \int_{t_0}^{t_f} C\left(X(t),U(t),t\right) dt + S\left(X(t_f),t_f\right)\,, \tag{A.62}$$

with the quadratic running cost in state and control,

$$C(x,u,t) = \frac{1}{2} q_0 x^2 + \frac{1}{2} r_0 u^2\,, \tag{A.63}$$

where $r_0 > 0$ for minimum costs and $q_0 \ge 0$, while the terminal quadratic cost is quadratic in the state only

$$S(x,t) = \frac{1}{2} s_0 x^2\,, \tag{A.64}$$

where $s_0 \ge 0$. It is assumed there are no bounds on the control $U(t)$ to preserve the nice canonical features of the LQ model. Otherwise the model features would have much more complexity.

Consequently, the Hamiltonian has the form,

$$\mathcal{H}(x,u,t) = \frac{1}{2} q_0 x^2 + \frac{1}{2} r_0 u^2 + \lambda(a_0 x + b_0 u)\,. \tag{A.65}$$

Without control constraints and with quadratic control costs, the regular control policy is the optimal one, governed by the corresponding Hamilton's equations,

$$\dot{X}^*(t) = +(\mathcal{H}_\lambda)^* = a_0 X^*(t) + b_0 U^*(t)\,, \tag{A.66}$$

$$\dot{\lambda}^*(t) = -(\mathcal{H}_x)^* = -q_0 X^*(t) - a_0 \lambda^*(t)\,, \tag{A.67}$$

$$0 = +(\mathcal{H}_u)^* = r_0 U^*(t) + b_0 \lambda^*(t)\,. \tag{A.68}$$

The Legendre-Clebsch 2nd order minimum condition is satisfied, since

$$(\mathcal{H}_{uu})^* = r_0 > 0 \tag{A.69}$$

by the positive definite assumption on r_0. Thus, the optimal control is

$$U^*(t) = U^{(\mathrm{reg})}(t) = -b_0 \lambda^*(t)/r_0\,, \tag{A.70}$$

while using (A.70) both the state and co-state optimal dynamics satisfies a linear first order matrix system of differential equations,

$$\dot{\mathbf{Z}}(t) \equiv \begin{bmatrix} \dot{X}^*(t) \\ \dot{\lambda}^*(t) \end{bmatrix} = M\mathbf{Z}(t) \equiv \begin{bmatrix} a_0 & -b_0^2/r_0 \\ -q_0 & -a_0 \end{bmatrix} \mathbf{Z}(t)\,. \tag{A.71}$$

The matrix differential equation (A.71) has the general eigen-solution,

$$\mathbf{Z}(t) = c_1 e^{\mu_1(t-t_0)} \begin{bmatrix} 1 \\ (a_0-\mu_1)r_0/b_0^2 \end{bmatrix} + c_2 e^{-\mu_1(t-t_0)} \begin{bmatrix} 1 \\ (a_0+\mu_1)r_0/b_0^2 \end{bmatrix} \tag{A.72}$$

where c_1 and c_2 are constants of integration, and

$$\mu_1 = \sqrt{a_0^2 + q_0 b_0^2/r_0} \tag{A.73}$$

is the principal eigenvalue of the matrix M defined in (A.71). This eigenvalue must be real by the coefficient assumptions, but $q_0 > -r_0 a_0^2/b_0^2$ would be a sufficient condition for for μ_1 to be real instead of the condition $q_0 > 0$.

The constants of integration (c_1, c_2) are determined by the initial condition

$$X^*(t_0) = x_0$$

from the first component of $\mathbf{Z}(t)$ in (A.72) and since t_f is fixed but not $X(t_f)$, the final or transversality condition in Table A.1 on page A8 provides a second condition,

$$\lambda^*(t_f) = S_x(X^*(t_f), t_f) = s_0 X^*(t_f) \tag{A.74}$$

from the second component of $\mathbf{Z}(t)$ in (A.72). Upon substitution of the constants of integration, the solution $(X^*(t), \lambda^*(t))$ can be found explicitly, say by symbolic computation systems such as Maple or *Mathematica*, but is too long and complicated to present here. However, an important property is that both $X^*(t)$ and $\lambda^*(t)$ are proportional to the initial state. The linear feedback relationship between the optimal control and the optimal state can be found from these two solutions and the linear relationship between the optimal control and the co-state in (A.70) yields a **linear feedback control law**,

$$U^*(t) = K(t)X^*(t)\,, \tag{A.75}$$

where

$$K(t) = -(b_0/r_0)\lambda^*(t)/X^*(t)\,, \tag{A.76}$$

which is called the **feedback gain** coefficient and is independent of the initial state x_0 since it cancels out of the co-state to state ratio. The linear feedback control law (A.75) with (A.76) is called feed back or **closed loop control** because it uses state information. However, if the control law is just time-dependent and state-independent, then the law would be called an **open loop control**.

If the plant manager is just concerned with what optimal control input is needed to achieve optimal control in the next time step, then only the feedback gain is required assuming the current state output $X^*(t)$ is known. This gain $K(t)$ (sometimes the control law is expressed with a minus sign, $U^*(t) = -\widehat{K}(t)X^*(t)$) can be found directly from a bilinear (quadratic) first order equation, called a **Riccati equation**,

$$\dot{K}(t) = -b_0 K^2(t) - 2a_0 K(t) + b_0 q_0/r_0 \,, \tag{A.77}$$

using a numerical differential equation solver backward in time, with just knowledge of the system and cost parameters, as well as the final condition,

$$K(t_f) = -b_0 s_0/r_0 \,, \tag{A.78}$$

from (A.76) and (A.74).

A.3.2 Matrix, Linear Dynamics, Quadratic Costs (LQ)

In general, LQ control problems will have time-dependent matrix coefficients, and will have both multi-dimensional vector states and controls. Again, let $\mathbf{X}(t)$ be n_x-dimensional and $\mathbf{U}(t)$ be n_u-dimensional. With some more effort the matrix form of the LQ problem can be solved, using the symbolic tools of Maple and *Mathematica* or the numerical tools of MATLAB.

Let the matrix form of the linear (L) state dynamics be

$$\dot{\mathbf{X}}(t) = A(t)\mathbf{X}(t) + B(t)\mathbf{U}(t) \,, \quad t_0 \le t \le t_f \,, \quad \mathbf{X}(t_0) = \mathbf{x}_0 \,, \tag{A.79}$$

where the coefficient matrices are $A(t) = [a_{i,j}]_{n_x \times n_x}$ and $A(t) = [b_{i,j}]_{n_x \times n_u}$, commensurate in matrix-vector multiplication. The quadratic (Q) cost objective is

$$V[\mathbf{X}, \mathbf{U}, t_f](\mathbf{x}_0, t_0) = \frac{1}{2}\int_{t_0}^{t_f} \left[\mathbf{X}^\top(t)Q(t)\mathbf{X}(t) + \mathbf{U}^\top(t)R(t)\mathbf{U}(t)\right] dt \tag{A.80}$$

$$+\frac{1}{2}\mathbf{X}^\top(t_f)S_f(t_f)\mathbf{X}(t_f) \,,$$

where the cost coefficient matrices are all symmetric, $n_x \times n_x$ state cost coefficients $Q(t)$ and $S_f(t)$ are positive semi-definite ($Q(t) \ge 0$, $S_f(t) \ge 0$), while the $n_u \times n_u$ control cost coefficients must be positive definite, $R(t) > 0$ to insure minimum costs. The Hamiltonian auxiliary objective is

$$\mathcal{H}(\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}, t) = \frac{1}{2}\left(\mathbf{x}^\top Q(t)\mathbf{x} + \mathbf{u}^\top R(t)\mathbf{u}\right) + \boldsymbol{\lambda}^\top \left(A(t)\mathbf{x} + B(t)\mathbf{u}\right) \,, \tag{A.81}$$

where $\boldsymbol{\lambda} = [\lambda_i]_{n_x \times 1}$ is the auxiliary co-state vector used to include the dynamical constraints to the running cost objective. In absence of control constraints and with $R(t) > 0$, the regular control is the optimal control and Hamilton's equations are

$$\dot{\mathbf{X}}^*(t) = +(\mathcal{H}_{\boldsymbol{\lambda}})^* = A(t)\mathbf{X}^*(t) + B(t)\mathbf{U}^*(t) \,, \tag{A.82}$$

$$\dot{\boldsymbol{\lambda}}^*(t) = -(\mathcal{H}_{\mathbf{x}})^* = -Q(t)\mathbf{X}^*(t) - A^\top(t)\boldsymbol{\lambda}^*(t) \,, \tag{A.83}$$

$$\mathbf{0} = (\mathcal{H}_{\mathbf{u}})^* = R(t)\mathbf{U}^*(t) + B^\top(t)\boldsymbol{\lambda}^*(t) \,, \tag{A.84}$$

where by the gradient peel theorem (B.133) the transposes of $A(t)$ and $B(t)$ multiply $\boldsymbol{\lambda}^*(t)$ in (A.83) and (A.84), respectively.

Since $R(t) > 0$, i.e., $R(t)$ is positive definite and has positive $R(t)$ eigenvalues, it is invertible (B.136). Hence, the optimal control in absence of control constraints is proportional to the co-state vector,

$$\mathbf{U}^*(t) = -R^{-1}(t)B^\top(t)\boldsymbol{\lambda}^*(t)\,. \tag{A.85}$$

As in the scalar case, we seek to show, as least formally, that the optimal control is also feedback control depending on the state vector $\mathbf{X}^*(t)$. Our approach will resemble the 2×2 scalar solution, but using $(2n_x)\times(2n_x)$ matrices partitioned into $n_x \times n_x$ sub-matrices to keep the analysis compact and close to the scalar case as much as possible. Thus, our system has the form

$$\dot{\mathbf{Z}}(t) = M(t)\mathbf{Z}(t)\,, \tag{A.86}$$

where the partitioned forms are

$$\mathbf{Z}(t) \equiv \begin{bmatrix} \mathbf{X}^*(t) \\ \boldsymbol{\lambda}^*(t) \end{bmatrix} \tag{A.87}$$

which has dimension $(2n_x)$ and

$$M(t) \equiv \begin{bmatrix} A(t) & -B(t)R^{-1}(t)B^\top(t) \\ -Q(t) & -A^\top \end{bmatrix} \tag{A.88}$$

which has dimension $(2n_x)\times(2n_x)$. The multiplication of partitioned matrices works essentially the same way that multiplication of non-partitioned matrices works.

Since the ordinary differential equation system in (A.87) for $\mathbf{Z}(t)$ is linear, then the usual exponential approximations works. So let a simple trial exponential solution form be

$$\mathbf{Z}(t) = Ce^{\mu t}\boldsymbol{\zeta}\,, \tag{A.89}$$

where C is a constant of integration, μ is a constant exponent coefficient and $\boldsymbol{\zeta}$ is a constant vector with the same $(2n_x)$ dimension as $\mathbf{Z}(t)$. Upon substitution into (A.87) yields the $(2n_x)$ dimensional eigenvalue problem (B.131)

$$M(t)\boldsymbol{\zeta} = \mu\boldsymbol{\zeta}\,, \tag{A.90}$$

so there should be $(2n_x)$ eigenvalues $[\mu_i]_{(2n_x)\times 1}$ and $(2n_x)$ associated eigenvectors

$$\boldsymbol{\zeta}_j = [\zeta_{i,j}]_{(2n_x)\times 1}$$

which are represented as columns of the matrix

$$\Psi = \left[\boldsymbol{\zeta}_j\right]_{1\times(2n_x)} \equiv [\zeta_{i,j}]_{(2n_x)\times(2n_x)}\,. \tag{A.91}$$

Linear superposition of these $(2n_x)$ eigen-solutions yields the general solution,

$$\mathbf{Z}(t) = \sum_{k=1}^{2n_x} C_k e^{\mu_k t} \zeta_k = (\Psi . * \mathbf{E}(t))\, \mathbf{C} \equiv \widehat{\Psi}(t)\mathbf{C}\,, \tag{A.92}$$

where $\mathbf{E}(t) \equiv [\exp(\mu_i t)]_{(2n_x)\times 1}$ is the exponential growth vector at the eigen-mode rate, the symbol pair $.*$ is MATLAB's dot-multiplication notation for element-wise multiplication (e.g., $\mathbf{x}.*\mathbf{y} = [x_i y_i]_{n_x \times n_x}$ for vector-vector multiplication or $A.*\mathbf{x} = [a_{i,j}x_j]_{n_x \times n_x}$ in matrix-vector multiplication), and

$$\widehat{\Psi}(t) = \begin{bmatrix} \widehat{\Psi}_{11}(t) & \widehat{\Psi}_{12}(t) \\ \widehat{\Psi}_{21}(t) & \widehat{\Psi}_{22}(t) \end{bmatrix} \equiv \Psi . * \mathbf{E}(t) = \begin{bmatrix} \Psi_{11}e^{\mu_1 t} & \Psi_{12}e^{\mu_2 t} \\ \Psi_{21}e^{\mu_1 t} & \Psi_{22}e^{\mu_2 t} \end{bmatrix}, \tag{A.93}$$

is a convenient abbreviation for the coefficient matrix of $\mathbf{C}$, also given partitioned into 4 $n_x \times n_x$ submatrices. The constant of integration vector

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} \tag{A.94}$$

is determined from the initial state condition

$$[Z_i(0)]_{n_x \times 1} = \widehat{\Psi}_{11}(0)\mathbf{C}_1 + \widehat{\Psi}_{12}(0)\mathbf{C}_2 = \mathbf{X}^*(0) = \mathbf{x}_0 \tag{A.95}$$

and the final co-state or transversality condition for free $\mathbf{X}^*(t_f)$ from Table A.1 on page A8,

$$\begin{aligned} [Z_{n+i}(t_f)]_{n_x \times 1} &= \widehat{\Psi}_{21}(t_f)\mathbf{C}_1 + \widehat{\Psi}_{22}(t_f)\mathbf{C}_2 \\ &= \boldsymbol{\lambda}^*(t_f) = \frac{1}{2}\nabla_{\mathbf{x}}\left[\mathbf{X}^\top S_f \mathbf{X}\right](t_f) = S_f(t_f)\mathbf{X}(t_f) \\ &= S_f(t_f)\left(\widehat{\Psi}_{11}(t_f)\mathbf{C}_1 + \widehat{\Psi}_{12}(t_f)\mathbf{C}_2\right). \end{aligned} \tag{A.96}$$

So this final condition is an algebraic equation that is homogeneous in $\mathbf{C}$. Upon rearranging the initial and final conditions, (A.95) and (A.96), the complete linear algebraic problem for $\mathbf{C}$ becomes

$$\begin{aligned} G\mathbf{C} &\equiv \begin{bmatrix} \widehat{\Psi}_{11}(0) & \widehat{\Psi}_{12}(0) \\ \widehat{\Psi}_{21}(t_f) - S_f(t_f)\widehat{\Psi}_{11}(t_f) & \widehat{\Psi}_{22}(t_f) - S_f(t_f)\widehat{\Psi}_{12}(t_f) \end{bmatrix} \mathbf{C} \\ &= \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{0} \end{bmatrix}. \end{aligned} \tag{A.97}$$

Assuming that the constant coefficient matrix G is invertible (this can be tested by one of the numerical or symbolic toolboxes), then the solution, using partitioning

and simplification due the homogeneity of the final condition, will formally be of the form:

$$\mathbf{C} = \widehat{G}^{-1} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \widehat{G}_{11}^{-1} & \widehat{G}_{12}^{-1} \\ \widehat{G}_{21}^{-1} & \widehat{G}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \widehat{G}_{11}^{-1} \\ \widehat{G}_{21}^{-1} \end{bmatrix} \mathbf{x}_0 \,, \tag{A.98}$$

where $\widehat{G}^{-1}$ is the inverse of G, i.e., $\widehat{G}^{-1}G = I_{2n_x \times 2n_x}$. The same relation does not necessarily hold for the $n_x \times n_x$ partitioned matrices, so $\widehat{G}_{i,j}^{-1}$ is not necessarily the inverse of $G_{i,j}$. Hence, the state and co-state solutions will be linear in the initial condition vector $\mathbf{x}_0$,

$$\mathbf{X}^*(t) = \left(\widehat{\Psi}_{11}(t)\widehat{G}_{11}^{-1} + \widehat{\Psi}_{12}(t)\widehat{G}_{21}^{-1} \right) \mathbf{x}_0 \,, \tag{A.99}$$

$$\boldsymbol{\lambda}^*(t) = \left(\widehat{\Psi}_{21}(t)\widehat{G}_{11}^{-1} + \widehat{\Psi}_{22}(t)\widehat{G}_{21}^{-1} \right) \mathbf{x}_0 \,. \tag{A.100}$$

Assuming that the coefficient matrix in (A.99) can be inverted so the backward evolution of the state is

$$\mathbf{x}_0 = \left(\widehat{\Psi}_{11}(t)\widehat{G}_{11}^{-1} + \widehat{\Psi}_{12}(t)\widehat{G}_{21}^{-1} \right)^{-1} \mathbf{X}^*(t) \,, \tag{A.101}$$

then the optimal control is a feedback control, i.e., linear in the state vector, and is given by

$$\mathbf{U}^*(t) = K(t)\mathbf{X}^*(t) \,, \tag{A.102}$$

where the gain matrix, using (A.85) with (A.99-A.102). The initial state thus far has been arbitrary and is

$$K(t) = -R(t)^{-1}B^\top(t) \left(\widehat{\Psi}_{21}(t)\widehat{G}_{11}^{-1} + \widehat{\Psi}_{22}(t)\widehat{G}_{21}^{-1} \right) \left(\widehat{\Psi}_{11}(t)\widehat{G}_{11}^{-1} + \widehat{\Psi}_{12}(t)\widehat{G}_{21}^{-1} \right)^{-1} . \tag{A.103}$$

Note that other texts may define the gain matrix differently, some using the state to co-state relation, but here we take the view that the user is the plant manager, who would be interested in the relation between the optimal control and the state. See Kalman [156] for justification of (A.103). An alternative to the eigenvalue problem approach to the solution of the dynamic equations, provided that the gain matrix is the main interest, is the Riccati differential equation approach. Using the state to co-state relation,

$$\boldsymbol{\lambda}^*(t) = J(t)\mathbf{X}^*(t) \,, \tag{A.104}$$

where the matrix $J(t)$ is defined so that

$$K(t) = -R^{-1}(t)B^\top J(t) \,, \tag{A.105}$$

and to avoid having to differentiate the variable coefficients. By differentiating both sides of (A.104) with respect to t, substituting for $\dot{\boldsymbol{\lambda}}^*(t)$ from (A.83), $\dot{\mathbf{X}}^*(t)$ from (A.82), $\boldsymbol{\lambda}^*(t)$ from (A.104) and $\mathbf{U}^*(t)$ from (A.85), and setting the common coefficient of $\mathbf{X}^*(t)$ equal to zero produces the quadratic, **matrix Riccati equation**,

$$\dot{J}(t) = \left[JBR^{-1}B^\top J - JA - A^\top J - Q\right](t) \tag{A.106}$$

with the final condition

$$J(t_f) = S_f(t_f) \tag{A.107}$$

from the final condition $\boldsymbol{\lambda}^*(t_f) = S_f(t_f)\mathbf{X}(t_f)$ in (A.96). Hence, $J(t)$ is just an extension of the terminal cost quadratic coefficient $S_f(t)$ for $0 \leq t < t_f$. This makes the Riccati problem (A.106) a final value problem rather than an initial value problem. It can be shown that $J(t)$ is symmetric from (A.106) and $S_f(t_f)$ is assumed to be symmetric, so only the upper or lower half of $J(t)$ plus its diagonal need be calculated. The control gain matrix $K(t)$ can be recovered using (A.105). Numerical approximation is almost always needed using methods of ordinary differential equations solvers in the numeric and symbolic computational toolboxes or elsewhere.

Once the feedback gain, either as $K(t)$ or $J(t)$, and the optimal state trajectory $\mathbf{X}^*(t)$ are obtained, the corresponding optimal control trajectory can be computed and then the optimal total cost value $v^*(\mathbf{x}_0, t_0) = \min_U[V[\mathbf{X}, \mathbf{U}](\mathbf{x}_0, t_0)]$ can be computed from (A.3) by integrating the running cost and adding the sum to the terminal cost term.

In the case where the cost function is a full quadratic polynomial in $\mathbf{x}$ and $\mathbf{u}$, i.e., with linear (affine) cost terms, then the control has $\mathbf{X}^*(t)$-independent terms requiring another companion ordinary differential equation for $J(t)$.

A.4 Deterministic Dynamic Programming (DDP)

Dynamic programming is another approach to the optimal control problem whose aim is to obtain the feedback optimal control $\mathbf{u}^*(\mathbf{x}, t)$ and the optimal value $v^*(\mathbf{x}, t)$, rather than primarily seeking the optimal trajectory set $\{\mathbf{X}^*(t), \boldsymbol{\lambda}^*(t), \mathbf{U}^*(t)\}$ using Hamilton's equations (A.6,A.7,A.8). The dynamic programming approach is principally due to Bellman [25] and begins with a slightly different formulation of the Bolza problem designed for better analytical manipulation using an arbitrary initial state $\mathbf{X}(t) = \mathbf{x}$ in the state domain. The deterministic dynamical system (A.1) is reformulated as

$$\frac{d\mathbf{X}}{ds}(s) = \mathbf{f}(\mathbf{X}(s), \mathbf{U}(s), s), \quad \mathbf{X}(t) = \mathbf{x}\,, \tag{A.108}$$

and the objective value functional as

$$V[\mathbf{X}, \mathbf{U}, t_f](\mathbf{x}, t) = \int_t^{t_f} C\left(\mathbf{X}(s), \mathbf{U}(s), s\right) ds + S\left(\mathbf{X}(t_f), t_f\right), \tag{A.109}$$

with total minimum costs or optimal value starting from $(\mathbf{x}, t)$

$$v^*(\mathbf{x}, t) = \min_{\boldsymbol{U}(t,t_f]} [V[\mathbf{X}, \mathbf{U}, t_f](\mathbf{x}, t)] \tag{A.110}$$

and optimal terminal value,

$$v^*(\mathbf{x}, t_f) = S(\mathbf{x}, t_f) . \tag{A.111}$$

When $t = t_f$ the running cost integral vanishes leaving only the terminal cost term and since the initial state is reduced to the final state when $t = t_f$ then the minimization is no longer operative. The $\mathbf{x}$ in (A.111) thus can be arbitrary, coinciding with the fact that $\mathbf{X}(t_f)$ is unspecified in this optimal control formulation.

A.4.1 Deterministic Principle of Optimality

Dynamic programming relies crucially on a recursion for the current optimal value in terms of a future optimal value called **Bellman's Principle of Optimality**. The basic concept is the assumption that the minimization operation in (A.110) can be decomposed over the control path $\mathbf{U}(s)$ for the time variable s on $(t, t_f]$, open on the left since the state $\mathbf{x}$ at time t is given, into a product over increments in time using the **minimization operator multiplicative decomposition** rule:

$$\min_{\boldsymbol{U}(t,t_f]} \stackrel{\text{op}}{=} \min_{\boldsymbol{U}(t,t+\Delta t]} \min_{\boldsymbol{U}(t+\Delta t,t_f]} . \tag{A.112}$$

for some positive time increment Δt such that $t < t + \Delta t < t_f$ and with an analogous rule for maximization. Using this rule and the fact that an integral has a corresponding additive decomposition rule:

$$\begin{aligned} \int_t^{t_f} C(\mathbf{X}(s), \mathbf{U}(s), s)ds = &\int_t^{t+\Delta t} C(\mathbf{X}(s), \mathbf{U}(s), s)ds \\ &+ \int_{t+\Delta t}^{t_f} C(\mathbf{X}(s), \mathbf{U}(s), s)ds . \end{aligned} \tag{A.113}$$

Application of the minimization and integration decompositions leads to

$$\begin{aligned} v^*(\mathbf{x}, t) &= \min_{\boldsymbol{U}(t,t+\Delta t]} \Bigg[\int_t^{t+\Delta t} C(\mathbf{X}(s), \mathbf{U}(s), s)ds \\ &\quad + \min_{\boldsymbol{U}(t+\Delta t,t_f]} \Bigg[\int_{t+\Delta t}^{t_f} C(\mathbf{X}(s), \mathbf{U}(s), s)ds \Bigg] + S(\mathbf{X}(t_f), t_f) \Bigg] \\ &= \min_{\boldsymbol{U}(t,t+\Delta t]} \Bigg[\int_t^{t+\Delta t} C(\mathbf{X}(s), \mathbf{U}(s), s)ds + v^*(\mathbf{X}(t + \Delta t), t + \Delta t) \Bigg] , \end{aligned} \tag{A.114}$$

where the optimal value $v^*(\mathbf{x}, t)$ definition (A.110,A.109) has been reused when starting at the future state $\mathbf{X}(t + \Delta t) = \mathbf{x} + \Delta\mathbf{X}(t)$ at time $t + \Delta t$. Thus, the following form of the optimality principle has been formally derived:

Lemma A.10. ***Bellman's Deterministic Principle of Optimality:*** *Under the assumptions of the operator decomposition rules (A.112, A.113),*

$$v^*(\mathbf{x},t) = \min_{\boldsymbol{U}(t,t+\Delta t]} \left[\int_t^{t+\Delta t} C(\mathbf{X}(s),\mathbf{U}(s),s)ds + v^*(\mathbf{x}+\Delta\mathbf{X}(t),t+\Delta t) \right] . \quad \text{(A.115)}$$

A.4.2 Hamilton-Jacobi-Bellman (HJB) Equation of Deterministic Dynamic Programming

In the derivation of the partial differential equation of deterministic dynamic programming or Hamilton-Jacobi-Bellman (HJB) equation, Bellman's principle of optimality is applied for small increments Δt, so $\Delta t$is replaced by the differential dt. The future state is approximated by a first order Taylor approximation,

$$\mathbf{X}(t+dt) \stackrel{\text{dt}}{=} \mathbf{X}(t) + \frac{d\mathbf{X}}{dt}(t)dt = \mathbf{x} + \frac{d\mathbf{X}}{dt}(t)dt \,, \quad \text{(A.116)}$$

provided the state vector $\mathbf{X}(t)$ is continuously differentiable. Consequently, the first order approximation for the optimal value $v^*(\mathbf{x},t)$ according to the principle of optimality with $\mathbf{X}(t)=\mathbf{x}$ is

$$v^*(\mathbf{x},t) \stackrel{\text{dt}}{=} \min_{\boldsymbol{U}(t,t+dt]} \Big[C(\mathbf{x},\mathbf{U}(t),t)dt + v^*(\mathbf{x},t) + v_t^*(\mathbf{x},t)dt \quad \text{(A.117)}$$
$$+\nabla_{\mathbf{x}}^\top [v^*](\mathbf{x},t)\cdot\mathbf{f}(\mathbf{x},\mathbf{U}(t),t)dt \Big] \,,$$

provided $v^*(\mathbf{x},t)$ is continuously differentiable in $\mathbf{x}$ and t and $C(\mathbf{x},\mathbf{u},t)$ is continuous so that $\mathrm{o}(dt)$ can be neglected. Note that the optimal value $v^*(\mathbf{x},t)$ appears alone on both sides of (A.117), so both of these $v^*(\mathbf{x},t)$ terms can be cancelled. Upon letting $\mathbf{U}(t)\equiv\mathbf{u}$ and replacing the vector set $\mathbf{U}(t,t+dt]$ by $\mathbf{u}$ the PDE of deterministic dynamic programming can be summarized as the following result:

Theorem A.11. ***Hamilton-Jacobi-Bellman Equation (HJBE) for Deterministic Dynamic Programming)***
If $v^*(\mathbf{x},t)$ is once differentiable in $\mathbf{x}$ and once differentiable in t, while the decomposition rules (A.112,A.113) are valid, then

$$0 = v_t^*(\mathbf{x},t) + \min_{\mathbf{u}} [\mathcal{H}(\mathbf{x},\mathbf{u},t)] \equiv v_t^*(\mathbf{x},t) + \mathcal{H}^*(\mathbf{x},t) \,, \quad \text{(A.118)}$$

where the ***Hamiltonian*** *(technically a* ***pseudo-Hamiltonian****) functional is given by*

$$\mathcal{H}(\mathbf{x},\mathbf{u},t) \equiv C(\mathbf{x},\mathbf{u},t) + \nabla_{\mathbf{x}}^\top [v^*](\mathbf{x},t)\cdot\mathbf{f}(\mathbf{x},\mathbf{u},t) \,. \quad \text{(A.119)}$$

The optimal control, if it exists, is given by

$$\mathbf{u}^*(\mathbf{x},t) = \operatorname*{argmin}_{\mathbf{u}} [\mathcal{H}(\mathbf{x},\mathbf{u},t)] \,. \quad \text{(A.120)}$$

This **Hamilton-Jacobi-Bellman** equation (HJBE) (A.118, A.119) is no ordinary PDE, but has the following properties or attributes:

Properties A.12.

- *The HJBE is a **functional PDE** due to the presence of the minimum operator* min.
- *The HJBE is a **scalar valued equation**, but solution output has dimension* $(nu + 1)$ *consisting of the scalar optimal value function* $v^* = v^*(\mathbf{x}, t)$ *and the optimal control vector* $\mathbf{u}^* = \mathbf{u}^*(\mathbf{x}, t)$ *as well. These **dual solutions** are generally tightly coupled in functional dependence. In general, this tight coupling requires a number of iterations between* v^* *and* $\mathbf{u}^*$ *to obtain a reasonable approximation to the* $(nu+1)$*-dimensional solution over the* (n_x+1)*-dimensional space of independent variables* $(\mathbf{x}, t)$*. However, it should be noted that the optimal control* $\mathbf{u}(\mathbf{x}, t)$ *in (6.18) is also feedback optimal control if the* $\mathbf{x}$ *dependence is genuine.*
- *In contrast to the Hamilton's equations formulation, the dynamic programming solution does not give the state trajectory directly but the state dynamics (A.108) must be solved using the feedback optimal control* $u^*(\mathbf{X}(t), t)$ *using (A.120). If the optimal control solution is computational, which is usual except for special or canonical problems, then the state dynamic solution would also be computational.*

A.4.3 Computational Complexity for Deterministic Dynamic Programming

The state-time vector valued form of the solution set, $\{v^*(\mathbf{x}, t), \mathbf{u}^*(\mathbf{x}, t)\}$, given independent state and time variables, $\mathbf{x}$ and t, makes the dynamic programming quite different from the Hamilton's equations for optimal time-dependent vector trajectories $\{\mathbf{X}(t), \boldsymbol{\lambda}(t), \mathbf{U}(t)\}$. If time is fixed at a single discrete value $T_k = t_0 + (k-1)\Delta T$ for some k where $k = 1 : N_t$ with $\Delta T = (t_f - t_0)/(N_t - 1)$, then the independent discretization of the n_x-dimensional state vector $\mathbf{x}$ is replaced by $\mathbf{X_j} = [X_{i,j_i}]_{n_x \times 1}$ where $\mathbf{j} = [j_i]_{n_x \times 1}$, $j_i = 1 : N_x$ for $i = 1 : n_x$ and N_x is the common number of state nodes, simply taken to be the same for each component (otherwise, N_x could be the geometric mean of n_x node counts N_i for $i = 1 : n_x$). However, $\mathbf{X_j}$ only represents one point in state space and there are a total $N_x^{n_x}$ numerical nodes or points in n_x state-dimensions. Thus, total numerical representation optimal value $v(\mathbf{x}, T_k)$ is

$$V^{(k)} = [V^{(k)}_{j_1, j_2, \ldots, j_{n_x}}]_{N_x \times N_x \times \cdots \times N_x} , \tag{A.121}$$

per time step k, so that the computational complexity is

$$CC(N_x, n_x) = O(N_x^{n_x}) = O(\exp(n_x \ln(N_x))) , \tag{A.122}$$

which by the law of exponents is exponential in the dimension with an exponent coefficient depending on the logarithm of the common number of nodes N_x, symbolizing the exponential computational complexity of Bellman's Curse of Dimensionality. This is also the exponential order of the complexity for solving multi-dimensional PDEs. For the optimal control vector, the order is n_x times this order, but that does not change the exponential order dependency. The deterministic dynamic programming exponential complexity (A.122) should be compared with the deterministic Hamilton's equation formulation in (A.34) with its linear or bilinear complexity $O(3n_x \cdot N_t)$.

Further, for second order finite difference errors, the total error for one state dimension ($n_x = 1$) will be by definition

$$E_T(N_x, 1) = O(N_x^{-2}). \tag{A.123}$$

So even if the order of the complexity is fixed in state dimension $n_x > 1$, i.e., $N = N_x^{n_x}$ is a constant, then $N_x(N) = N^{1/n_x}$ and

$$E_T(N_x(N), n_x) = O\left(N^{-2/n_x}\right) \to O(1) \tag{A.124}$$

as $n_x \to +\infty$ for fixed N and accuracy, i.e., diminishing accuracy in the limit of large dimension.

There are many other computational issues but there is not enough space here to discuss them. Many of these are covered in the author's computational stochastic dynamic programming chapter [108] and more recently in [110].

A.4.4 Linear Quadratic (LQ) Problem by Deterministic Dynamic Programming

The linear quadratic problem is also a good demonstration of the method of dynamic programming as it was as an application of Hamilton's equations and the optimum principle. Using the same formulation, but modified for dynamic programming analysis to start at an arbitrary time t rather than a fixed time t_0, with the dynamics linear in both the control vector $\mathbf{U}(t)$ and the state vector $\mathbf{X}(t)$, the state dynamics is given by

$$\dot{\mathbf{X}}(s) = A(s)\mathbf{X}(s) + B(s)\mathbf{U}(s)\,, \quad t \le s \le t_f\,, \quad \mathbf{X}(t) = \mathbf{x}. \tag{A.125}$$

The objective cost functional is given by

$$V[\mathbf{X}, \mathbf{U}, t_f](\mathbf{x}, t) = \frac{1}{2}\int_t^{t_f} \left[\mathbf{X}^\top(s)Q(s)\mathbf{X}(s) + \mathbf{U}^\top(s)R(s)\mathbf{U}(s)\right] ds \tag{A.126}$$

$$+\frac{1}{2}\mathbf{X}^\top(t_f)S_f(t_f)\mathbf{X}(t_f)\,.$$

The total minimum cost is again from (A.110)

$$v^*(\mathbf{x}, t) = \min_{\boldsymbol{U}(t,t_f]} \left[V[\mathbf{X}, \mathbf{U}, t_f](\mathbf{x}, t)\right]\,, \tag{A.127}$$

provided mainly that the quadratic cost matrix $R(t) > 0$, i.e., is positive definite. The HJB equation is

$$0 = v_t^*(\mathbf{x}, t) + \min_{\mathbf{u}} \left[\mathcal{H}(\mathbf{x}, \mathbf{u}, t)\right] , \tag{A.128}$$

where the pseudo-**Hamiltonian** functional simplifies to

$$\mathcal{H}(\mathbf{x}, \mathbf{u}, t) = \frac{1}{2}\left(\mathbf{x}^\top Q(t)\mathbf{x} + \mathbf{u}^\top R(t)\mathbf{u}\right) + \nabla_{\mathbf{x}}^\top [v^*](\mathbf{x}, t)\left(A(t)\mathbf{x} + B(t)\mathbf{u}\right) . \tag{A.129}$$

Comparing the dynamic programming pseudo-Hamiltonian (A.119) with the standard Hamiltonian in (A.81) shows that the optimal value gradient $\nabla_{\mathbf{x}}[v^*](\mathbf{x}, t)$ (the marginal value or shadow value in economics) plays the same role as the Lagrange multiplier vector $\boldsymbol{\lambda}$ in (A.81).

Although the decomposition of the optimal value can be rigorously proven, it is sufficient for the purposes here to propose the decomposition is a quadratic form,

$$v^*(\mathbf{x}, t) = \frac{1}{2}\mathbf{x}^\top J(t)\mathbf{x} , \tag{A.130}$$

and justify it heuristically, i.e., by showing the form (A.130) works. The quadratic coefficient $J(t)$ is a $(n_x \times n_x)$ matrix and since the quadratic form ignores the asymmetric part of the quadratic coefficient, $J(t)$ will be assumed to be symmetric. Thus, the optimal value gradient with respect to the state vector by (B.138) will be

$$\nabla_{\mathbf{x}}[v^*](\mathbf{x}, t) = J(t)\mathbf{x} . \tag{A.131}$$

In the case that the cost function is a general quadratic form with linear and zeroth degree terms, then the optimal value LQ decomposition (A.130) will have the same kind of terms.

It is also assumed that there are no constraints on the control to maintain the classical LQ problem form. Thus, stationary points of the pseudo-Hamiltonian are sought,

$$\nabla_{\mathbf{u}}[\mathcal{H}](\mathbf{x}, \mathbf{u}, t) = R(t)\mathbf{u} + B^\top(t)J(t)\mathbf{x} = \mathbf{0} , \tag{A.132}$$

using (B.133, B.138) and the fact that $R(t)$ is symmetric. So the unconstrained optimal control is the linear feedback control

$$\mathbf{u}^*(\mathbf{x}, t) = K(t) \equiv -R^{-1}(t)B^\top(t)J(t)\mathbf{x} , \tag{A.133}$$

where the inverse of the quadratic cost coefficient $R(t)$ exists since $R(t)$ is positive definite and where $K(t)$ is the same **gain matrix** as in (A.103) found from the Hamilton's equation formulation. Upon substitution into the HJB equation leads to a pure quadratic form using

$$v_t^*(\mathbf{x}, t) = \mathbf{x}^\top J'\mathbf{x} \tag{A.134}$$

and

$$\begin{aligned}\mathcal{H}^*(\mathbf{x}, t) &\equiv \mathcal{H}(\mathbf{x}, \mathbf{u}^*, t) \\ &= \mathbf{x}^\top \left[-\frac{1}{2}J(t)B(t)R^{-1}(t)B^\top(t)J(t) + J(t)A(t) + Q(t)\right]\mathbf{x} .\end{aligned} \tag{A.135}$$

Taking two partial derivatives or using the Hessian matrix formula (B.139) yields the **matrix Riccati equation**

$$J'(t) = -\frac{1}{2}J(t)B(t)R^{-1}(t)B^{\top}(t)J(t) - J(t)A(t) - A^{\top}(t)J(t) - Q(t) \,, \quad \text{(A.136)}$$

subject to the same final condition as well,

$$J(t_f) = S_f(t_f) \,. \quad \text{(A.137)}$$

For feedback optimal control and optimal value, the dynamic programming approach is a more direct approach and the algebra is more manageable than the approach through Hamilton's equations. However, the state trajectory is not produced directly by dynamic programming. The more general linear quadratic problem with jump-diffusion processes and other features will be treated in the next chapter.

A.5 Control of PDE Driven Dynamics (DPS)

Thus far, only the control of ODE driven systems have been considered. However, many dynamical systems are governed by partial differential equations (PDES), such as in fluid and solid dynamics. The PDE dynamics do greatly complicate the optimal control problem and there are many cases to consider. The control of PDE driven systems usually appears under the heading of *distributed parameter systems* (DPS) and the control is called *distributed control*, while ODE driven systems are classified as *lumped parameter systems* in contrast. For a more thorough but very applied approach to DPS control, the reader can consult Ahmed and Teo's [4] DPS book, Gunzberger's [101] recent monograph on flow control or consult the many applications in the DPS research directions proceedings [234]. See also the recent biomedical application to cancer drug delivery to the brain by Chakrabarty and Hanson [48] (briefly summarized in the biomedical application Section 11.2.2. Only one fairly general deterministic model will be presented here since the focus is on stochastic problems.

A.5.1 DPS Optimal Control Problem

Let $\mathbf{y}(\mathbf{x}, t)$ be a n_y-vector state variable in space-time where $\mathbf{x}$ is the n_x-dimensional space vector. The state dynamics for $\mathbf{y}(\mathbf{x}, t)$ satisfy a nonlinear reaction diffusion equation with drift,

$$\frac{\partial \mathbf{y}}{\partial t}(\mathbf{x}, t) = D\nabla_x^2[\mathbf{y}](\mathbf{x}, t) + C\nabla_x[\mathbf{y}](\mathbf{x}, t) + \mathbf{B}(\mathbf{y}(\mathbf{x}, t), \mathbf{x}, t) + A\mathbf{u}(\mathbf{x}, t) \,, \quad \text{(A.138)}$$

$\mathbf{x} \in \mathcal{D}_x$ and $t_0 < t \leq t_f$, with initial condition

$$\mathbf{y}(\mathbf{x}, t_0) = \mathbf{y}_0(\mathbf{x})$$

and mixed boundary condition

$$(\alpha(\widehat{\mathbf{n}}^{\top}\nabla_x)[\mathbf{y}] + \beta\mathbf{y} + \boldsymbol{\gamma})(\mathbf{x}, t) = \mathbf{0}$$

for $\mathbf{x}$ on the space domain boundary $\partial\mathcal{D}_x$ while $\widehat{\mathbf{n}}(\mathbf{x},t)$ is the outward normal to the boundary. Here $\mathbf{u}(\mathbf{x},t)$ is the n_u-dimensional space-time control variable in a linear control-dependent term. All coefficient functions are assumed to be bounded while being commensurate in multiplication and sufficiently differentiable as needed. In particular, the diffusion tensor $D = [D_i\delta_{i,j}]_{n_y\times n_y}$ is a positive-definite diagional matrix and the drift coefficient $C = [C_{i,k}\delta_{i,j}]_{n_y\times n_y\times n_x}$. The main *reaction* vector $\mathbf{B}(\mathbf{y}(\mathbf{x},t),\mathbf{x},t)$ is the only term assumed to be nonlinear since reaction terms are often naturally nonlinear. The control coefficient is $A = [A_{i,j}]_{n_y\times n_u}$ and is assumed to be constant but could depend on $(\mathbf{x},t)$, as could be C and D.

Further, let the space-time objective be in the form of the total quadratic costs,

$$V[\mathbf{y},\mathbf{u},t_f] = \frac{1}{2}\int_{t_0}^{t_f} dt \int_{\mathcal{D}_x} d\mathbf{x}\left(\mathbf{y}^\top Q\mathbf{y}+\mathbf{u}^\top R\mathbf{u}\right)(\mathbf{x},t)+\frac{1}{2}\int_{\mathcal{D}_x} d\mathbf{x}\left(\mathbf{y}^\top S\mathbf{y}\right)(\mathbf{x},t_f)\ , \tag{A.139}$$

where the quadratic control coefficient R is symmetric positive-definite, while Q and S are symmetric positive-semi-definite to ensure a minimum. Eqs. (A.138-A.139) provide the underlying formulation of the DPS optimal control problem.

A.5.2 DPS Hamiltonian Extended Space Formulation

For the formulation of the equations for the optimal solutions to the control problem, the dynamic and initial-boundary constraints need to be combined into a pseudo-Hamiltonian,

$$\begin{aligned}\mathcal{H}(\mathbf{y},\mathbf{u},\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{\nu}) = V[\mathbf{y},\mathbf{u},t_f]&\\ &+\int_{t_0}^{t_f} dt\int_{\mathcal{D}_x} d\mathbf{x}\ \boldsymbol{\lambda}^\top\big(\mathbf{y}_t - D\nabla_x^2[\mathbf{y}] - C\nabla_x[\mathbf{y}] - \mathbf{B} - A\mathbf{u}\big)(\mathbf{x},t)\\ &+\int_{t_0}^{t_f} dt\int_{\partial\mathcal{D}_x} d\boldsymbol{\Gamma}\boldsymbol{\mu}^\top\big(\alpha(\widehat{\mathbf{n}}^\top\nabla_x)[\mathbf{y}]+\beta\mathbf{y}+\boldsymbol{\gamma}\big)(\mathbf{x},t)\\ &+\int_{\mathcal{D}_x} d\mathbf{x}\ \boldsymbol{\nu}^\top\big(\mathbf{y}(\mathbf{x},t_0^+)-\mathbf{y}_0(\mathbf{x})\big)\ ,\end{aligned} \tag{A.140}$$

where $\{\boldsymbol{\lambda}(\mathbf{x},t),\boldsymbol{\mu}(\mathbf{x},t),\boldsymbol{\nu}(\mathbf{x})\}$ is a set of Lagrange multiplier that provide the mechanism for including the control problem constraints at the expense of extending the state-control space to higher dimension with

$$\mathbf{z}(\mathbf{x},t) \equiv \{\mathbf{y}(\mathbf{x},t),\mathbf{u}(\mathbf{x},t),\boldsymbol{\lambda}(\mathbf{x},t),\boldsymbol{\mu}(\mathbf{x},t),\boldsymbol{\nu}(\mathbf{x})\}$$

denoting the extended space-control vector. Next, assuming an optimal extended state $\mathbf{z}(\mathbf{x},t) = \mathbf{z}^*(\mathbf{x},t)$ exists under sufficient differentiability properties of $\mathcal{H}(\mathbf{z})$, perturb about this optimal extended state as $\mathbf{z}(\mathbf{x},t) = \mathbf{z}^*(\mathbf{x},t) + \delta\mathbf{z}(\mathbf{x},t)$ where $\delta\mathbf{z}(\mathbf{x},t)$ is the *variation* and then expand the pseudo-Hamiltonian about this variation,

$$\mathcal{H}(\mathbf{z}^*(\mathbf{x},t)+\delta\mathbf{z}(\mathbf{x},t)) = \mathcal{H}(\mathbf{z}^*(\mathbf{x},t)) + \delta\mathcal{H}(\mathbf{z}^*(\mathbf{x},t),\delta\mathbf{z}(\mathbf{x},t)) + \mathrm{O}(|\delta\mathbf{z}|^2(\mathbf{x},t)).$$

Neglecting quadratic order perturbation terms, including the second variation of $\mathcal{H}(\mathbf{z})$, then the first variation $\delta\mathcal{H}(\mathbf{z}^*(\mathbf{x},t),\delta\mathbf{z}(\mathbf{x},t))$ is found to be a linear function of the extended state perturbation $\delta\mathbf{z}(\mathbf{x},t)$ using (A.139-A.140). For this perturbation, the nonlinear reaction term $\mathbf{B}(\mathbf{y}(\mathbf{x},t),\mathbf{x},t)$ is assumed to be more than once differentiable so that

$$\mathbf{B}(\mathbf{y}^*+\delta\mathbf{y},\mathbf{x},t)=\mathbf{B}(\mathbf{y}^*,\mathbf{x},t)+(\delta\mathbf{y}^\top\nabla_y)[\mathbf{B}](\mathbf{y}^*,\mathbf{x},t)+\mathrm{O}(|\delta\mathbf{y}|^2)\,,$$

for example twice differentiable to guarantee the quadratic order error term. For simplicity, let $B^*\equiv\mathbf{B}(\mathbf{y}^*,\mathbf{x},t)$. Applying multiple Taylor approximations,

$$\begin{aligned}\delta\mathcal{H}(\mathbf{z}^*,\delta\mathbf{z})=&\int_{t_0}^{t_f}dt\int_{\mathcal{D}_x}d\mathbf{x}\left((\mathbf{y}^*)^\top Q\delta\mathbf{y}+(\mathbf{u}^*)^\top R\delta\mathbf{u}\right)(\mathbf{x},t)+\int_{\mathcal{D}_x}d\mathbf{x}\left((\mathbf{y}^*)^\top S\delta\mathbf{y}\right)(\mathbf{x},t_f)\\&+\int_{t_0}^{t_f}dt\int_{\mathcal{D}_x}d\mathbf{x}\,(\boldsymbol{\lambda}^*)^\top\left(\delta\mathbf{y}_t-D\nabla_x^2[\delta\mathbf{y}]-C\nabla_x[\delta\mathbf{y}]-(\delta\mathbf{y}^\top\nabla_y)[\mathbf{B}]^*-A\delta\mathbf{u}\right)(\mathbf{x},t)\\&+\int_{t_0}^{t_f}dt\int_{\mathcal{D}_x}d\mathbf{x}\,d\boldsymbol{\lambda}^\top\left(\mathbf{y}_t^*-D\nabla_x^2[\mathbf{y}^*]-C\nabla_x[\mathbf{y}^*]-\mathbf{B}^*-A\mathbf{u}^*\right)(\mathbf{x},t)\\&+\int_{t_0}^{t_f}dt\int_{\partial\mathcal{D}_x}d\boldsymbol{\Gamma}(\boldsymbol{\mu}^*)^\top\left(\alpha(\widehat{\mathbf{n}}^\top\nabla_x)[\delta\mathbf{y}]+\beta\delta\mathbf{y}\right)(\mathbf{x},t)\\&+\int_{t_0}^{t_f}dt\int_{\partial\mathcal{D}_x}d\boldsymbol{\Gamma}(\delta\boldsymbol{\mu}^\top\left(\alpha(\widehat{\mathbf{n}}^\top\nabla_x)[\mathbf{y}^*]+\beta\mathbf{y}^*+\boldsymbol{\gamma}\right)(\mathbf{x},t)\\&+\int_{\mathcal{D}_x}d\mathbf{x}\left((\boldsymbol{\nu}^*)^\top\delta\mathbf{y}(\mathbf{x},t_0^+)+\delta\boldsymbol{\nu}^\top\left(\mathbf{y}^*(\mathbf{x},t_0^+)-\mathbf{y}_0(\mathbf{x})\right)\right)\,.\end{aligned}\tag{A.141}$$

Obtaining the critical or optimal conditions requires the reduction of the highest order partial derivative terms, since under integration the perturbations $\delta\mathbf{y}_t(\mathbf{x},t)$ and $\nabla_x^2[\delta\mathbf{y}]$ are not independent of lower order derivatives and the higher order derivatives can be reduced by integration by parts techniques to lower order derivatives. Thus, using integration by parts

$$\int_{t_0}^{t_f}dt\,(\boldsymbol{\lambda}^*)^\top\delta\mathbf{y}_t(\mathbf{x},t)=(\boldsymbol{\lambda}^*)^\top\delta\mathbf{y}\Big|_0^{t_f}-\int_{t_0}^{t_f}dt\,\delta\mathbf{y}^\top\boldsymbol{\lambda}_t^*\,,$$

$$-\int_{\mathcal{D}_x}d\mathbf{x}\,(\boldsymbol{\lambda}^*)^\top C\nabla_x[\delta\mathbf{y}]=-\int_{\partial\mathcal{D}_x}d\boldsymbol{\Gamma}\delta\mathbf{y}^\top\widehat{\mathbf{n}}^\top C^\top\boldsymbol{\lambda}^*+\int_{\mathcal{D}_x}d\mathbf{x}\,\delta\mathbf{y}^\top\nabla_x^\top[C^\top\boldsymbol{\lambda}^*]\,.$$

where $C^\top\equiv[C_{k,i}\delta_{k,j}]_{n_xn_x\times n_y\times n_y}$ defines the transpose of a three subscript array, and finally using a double integration by parts [102]

$$\begin{aligned}-\int_{\mathcal{D}_x}d\mathbf{x}\,(\boldsymbol{\lambda}^*)^\top D\nabla_x^2[\delta\mathbf{y}]=&-\int_{\partial\mathcal{D}_x}d\boldsymbol{\Gamma}\left((\widehat{\mathbf{n}}^\top\nabla_x)[\delta\mathbf{y}^\top]D\boldsymbol{\lambda}^*-\delta\mathbf{y}^\top(\widehat{\mathbf{n}}^\top\nabla_x)[D\boldsymbol{\lambda}^*]\right)\\&-\int_{\mathcal{D}_x}d\mathbf{x}\,\delta\mathbf{y}^\top\nabla_x^2[D\boldsymbol{\lambda}^*]\,.\end{aligned}$$

Using these reduced forms in (A.141) and collecting terms as coefficients of like extended state perturbations produces a more useful form:

$$\begin{aligned}\delta\mathcal{H}(\mathbf{z}^*,\delta\mathbf{z}) = &\int_{t_0}^{t_f} dt \int_{\mathcal{D}_x} d\mathbf{x}\ (\delta\mathbf{y})^\top \left(-\boldsymbol{\lambda}_t^* - \nabla_x^2[D\boldsymbol{\lambda}^*] + \nabla_x^\top[C^\top\boldsymbol{\lambda}^*] + \nabla_y[\mathbf{B}^\top]^*\boldsymbol{\lambda}^* + Q\mathbf{y}^*\right)\\ &+\int_{t_0}^{t_f} dt \int_{\mathcal{D}_x} d\mathbf{x}\ (\delta\mathbf{u})^\top \left(R\mathbf{u}^* - A^\top\boldsymbol{\lambda}^*\right)(\mathbf{x},t)\\ &+\int_{t_0}^{t_f} dt \int_{\mathcal{D}_x} d\mathbf{x}\ (\delta\boldsymbol{\lambda})^\top \left(\mathbf{y}_t^* - D\nabla_x^2[\mathbf{y}^*] - C\nabla_x[\mathbf{y}^*] - \mathbf{B}^* - A\mathbf{u}^*\right)(\mathbf{x},t)\\ &+\int_{t_0}^{t_f} dt \int_{\partial\mathcal{D}_x} d\boldsymbol{\Gamma}(\widehat{\mathbf{n}}^\top\nabla_x)[\delta\mathbf{y}^\top]\left(\alpha^\top\boldsymbol{\mu}^* - D\boldsymbol{\lambda}^*\right)(\mathbf{x},t) \qquad \text{(A.142)}\\ &+\int_{t_0}^{t_f} dt \int_{\partial\mathcal{D}_x} d\boldsymbol{\Gamma}(\delta\mathbf{y})^\top \left((\widehat{\mathbf{n}}^\top\nabla_x)[D\boldsymbol{\lambda}^*] - \widehat{\mathbf{n}}^\top C^\top\boldsymbol{\lambda}^* + \beta^\top\boldsymbol{\mu}^*\right)(\mathbf{x},t)\\ &+\int_{t_0}^{t_f} dt \int_{\partial\mathcal{D}_x} d\boldsymbol{\Gamma}(\delta\boldsymbol{\mu})^\top \left(\alpha(\widehat{\mathbf{n}}^\top\nabla_x)[\mathbf{y}^*] + \beta\mathbf{y}^* + \boldsymbol{\gamma}\right)(\mathbf{x},t)\\ &+\int_{\mathcal{D}_x} d\mathbf{x}\ \delta\mathbf{y}^\top(S\mathbf{y}^* + \boldsymbol{\lambda}^*)(\mathbf{x},t_f)\\ &+\int_{\mathcal{D}_x} d\mathbf{x}\ \delta\mathbf{y}^\top(\boldsymbol{\nu}^* - \boldsymbol{\lambda}^*)(\mathbf{x},t_0) + \int_{\mathcal{D}_x} d\mathbf{x}\ \delta\boldsymbol{\nu}^\top\left(\mathbf{y}^*(\mathbf{x},t_0^+) - \mathbf{y}_0(\mathbf{x})\right)\ .\end{aligned}$$

A.5.3 DPS Optimal State, Co-State and Control PDEs

Our interest here is to present a usable formulation for those whose prime interest is obtaining concrete solutions for applications, so our approach is a formal applied mathematical one. If the interest of the reader is in existence and uniqueness properties of the solution rather than the solution itself, the reader should explore the references [4], [101], [234] and references therein for abstract notions of Hilbert spaces with related Sobolev spaces and functional derivatives. However, such abstract approaches have little utility in solving real problems.

The optimal state, co-state and control trajectory dynamics follow from setting to zero each of the independent state, co-state and control first variations in the first variation of the pseudo-Hamiltonian in (A.142), as well as any relevant boundary, initial and final values which are assumed to be independent of the space-time interior values.

The optimal state equation for $\mathbf{y}^*(\mathbf{x},t)$ follows from setting the critical coefficient of the co-state variation $\delta\boldsymbol{\lambda}(\mathbf{x},t)$ on each interior point of $\mathcal{D}_x\times(t_0,t_f)$ yielding

$$\mathbf{y}_t^*(\mathbf{x},t) = \left(D\nabla_x^2[\mathbf{y}^*] + C\nabla_x[\mathbf{y}^*] + \mathbf{B}^* + A\mathbf{u}^*\right)(\mathbf{x},t)\,, \qquad \text{(A.143)}$$

for $\mathbf{x}\in\mathcal{D}_x$ and $t_0 < t \le t_f$, while the intial condition $\mathbf{y}^*(\mathbf{x},t_0^+) = \mathbf{y}_0(\mathbf{x})$ follows from setting the coefficient of the initial condition co-state variation $\delta\boldsymbol{\nu}(\mathbf{x})$ to zero and the boundary condition

$$\left(\alpha(\widehat{\mathbf{n}}^\top\nabla_x)[\mathbf{y}^*] + \beta\mathbf{y}^* + \boldsymbol{\gamma}\right)(\mathbf{x},t) = \mathbf{0}\,, \quad \mathbf{x}\in\partial\mathcal{D}_x\,, \quad t_0 < t < t_f,$$

from setting the coefficient of the final condition co-state variation $\delta\boldsymbol{\nu}(\mathbf{x})$ to zero. The optimal state equation (A.143), of course, has the same form as that of the original state equation (A.138), which is a forward parabolic PDE for $D > 0$.

The optimal co-state equation for $\boldsymbol{\lambda}^*$ is derived by setting the state variation $\delta\mathbf{y}(\mathbf{x}, t)$ coefficient to zero, so that

$$\left(\boldsymbol{\lambda}_t^* + \nabla_x^2[D\boldsymbol{\lambda}^*] - \nabla_x^\top[C^\top\boldsymbol{\lambda}^*] - \nabla_y[\mathbf{B}^\top]^*\boldsymbol{\lambda}^* - Q\mathbf{y}^*\right)(\mathbf{x}, t) = \mathbf{0}\,, \quad \text{(A.144)}$$

for $\mathbf{x} \in \mathcal{D}_x$ and $t_0 \leq t < t_f$, noting that (A.144) is a backward parabolic PDE since the diffusion term has an opposite sign to that of the forward equation (A.143). So a final condition is needed by setting the coefficient of $\delta\mathbf{y}(\mathbf{x}, t_f)$ to zero, i.e.,

$$(\boldsymbol{\lambda}^* + S\mathbf{y}^*)(\mathbf{x}, t_f) = \mathbf{0}\,, \quad \text{(A.145)}$$

coupling the computed final condition of $\boldsymbol{\lambda}^*$ to the computed final value of $(-S\mathbf{y}^*)$. The boundary conditions follow from setting the coefficient of $\delta\mathbf{y}$ on the boundary to zero, so

$$\left((\widehat{\mathbf{n}}^\top\nabla_x)[D\boldsymbol{\lambda}^*] - \widehat{\mathbf{n}}^\top C^\top\boldsymbol{\lambda}^* + \beta^\top\boldsymbol{\mu}^*\right)(\mathbf{x}, t) = \mathbf{0}\,, \quad \mathbf{x} \in \partial\mathcal{D}_x\,, \quad \text{(A.146)}$$

giving rise to another complication, in that the boundary condition co-state $\boldsymbol{\mu}^*(\mathbf{x}, t)$ appears. However, the coefficient the normal gradient $(\widehat{\mathbf{n}}^\top\nabla_x)[\delta\mathbf{y}^\top]$ yields,

$$\left(\alpha^\top\boldsymbol{\mu}^* - D\boldsymbol{\lambda}^*\right)(\mathbf{x}, t) = \mathbf{0}\,, \quad \mathbf{x} \in \partial\mathcal{D}_x\,,$$

which, if $\alpha^\top$ is invertible, can be used to eliminate $\boldsymbol{\mu}^*$ on the boundary. Another co-state condition comes from the initial value of $\delta\mathbf{y}$ which gives

$$\boldsymbol{\nu}^*(\mathbf{x}) = \boldsymbol{\lambda}^*(\mathbf{x}, t_0)\,,$$

where $\boldsymbol{\lambda}^*(\mathbf{x}, t_0)$ is the terminal output of the backward integration of the prime optimal co-state PDE (A.144) starting from the final condition (A.145).

From the coefficient of the control variation $\delta\mathbf{u}(\mathbf{x}, t)$, the optimal control is given by

$$\left(R\mathbf{u}^* - A^\top\boldsymbol{\lambda}^*\right)(\mathbf{x}, t) = \mathbf{0}\,, \quad \mathbf{x} \in \mathcal{D}_x\,, \quad t_0 \leq t < t_f\,,$$

and since $R(\mathbf{x}, t)$ should be invertible due to its positive-definite property then

$$\mathbf{u}^*(\mathbf{x}, t) = \left(R^{-1}A^\top\boldsymbol{\lambda}^*\right)(\mathbf{x}, t)\,, \quad \text{(A.147)}$$

in absence of control constraints, else it is merely the regular optimal control $\mathbf{u}^{(\mathrm{reg})}(\mathbf{x}, t)$.

A numerical scheme developed in Chakrabarty and Hanson [48] for a biomedical application uses a forward state integration of (A.143) and backward co-state integration of (A.144) with subsequent iterations until the norm of the iteration difference is sufficiently small. The forward integration step for (A.143) requires a good starting guess for the optimal control space-time distribution in addition to using the specified state initial condition. The final time approximation to $\mathbf{y}^*(\mathbf{x}, t)$ is then used as the final condition to start the co-state $\boldsymbol{\lambda}^*(\mathbf{x}, t)$ backward integration of (A.144). The end approximation of the co-state space-time distribution of $\boldsymbol{\lambda}^*(\mathbf{x}, t)$ is used by (A.147) to update the optimal control distribution approximation $\mathbf{u}^*(\mathbf{x}, t)$, which in turn is used in the next state forward integration.

A.6 Exercises

1. For the deterministic linear first order dynamics,

$$\dot{X}(t) = -\mu_0 X(t) + \beta_0 U(t), t > 0, \quad \text{given } X(0) = x_0 \neq 0, \quad \mu_0 > 0, \quad \beta_0 \neq 0,$$

and quadratic performance measure,

$$V[U] = \frac{r_0}{2} \int_0^{t_f} U^2(t) dt, \quad r_0 > 0,$$

find the optimal state trajectory and optimal (unconstrained) control to bring the state from the initial state to the final state x_f in t_f seconds while minimizing the functional $V[U]$ with respect to the control u, with the answer depending on the parameter set $\{x_0, x_f, t_f, \mu_0, \beta_0, r_0\}$. *Note that the final state and time are fixed.*

2. Consider another simple lumped model of a **leaky reservoir** (after Kirk [163]) given by

$$\dot{X}(t) = -aX(t) + U(t), \quad X(0) = x_0,$$

where $X(t)$ is the depth of the reservoir, $U(t)$ is the net flow of water per unit time into the reservoir at time t and $a > 0$ is the rate of leakage and usage. The net inflow is constrained pointwise $0 \leq U(t) \leq M$ for all $0 < t \leq t_f$ and also cumulatively by

$$\int_0^{t_f} U(t) dt = K > 0,$$

where K, M and t_f are fixed constants, such that $K \leq M \cdot t_f$ for consistency. Find the optimal control law $U^*(t)$ that maximizes only the final depth,

$$J[X] = bX(t_f)$$

with $b > 0$, the optimal state $X^*(t)$, optimal final depth $J[X^*]$ and the optimal Hamiltonian $\mathcal{H}^*$.

3. **Pontryagin's Auxiliary Necessary Conditions for the Hamiltonian in the Special Case of No Explicit Dependence on Time:** Assume sufficient differentiability for the Hamiltonian and that

$$\mathcal{H}^* = \mathcal{H}(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t)),$$

so $\partial \mathcal{H}^* / \partial t \equiv 0$. Then show:

   (a) If the final time t_f is *fixed* and the Hamiltonian $\mathcal{H}$ does not depend explicitly on time then the Hamiltonian must be constant when evaluated on an locally (interior) extreme trajectory, i.e.,

$$\mathcal{H}^* = \mathcal{H}(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t)) = c, \tag{A.148}$$

where c is a constant.

Explain why fixed t_f and local extremes are needed; also, explain show why Example A.8 or Exercise 2 are counter-examples for the result A.148 if certain condition is not satisfied, stating what that condition is.

(b) If the final time t_f is *free* and both the Hamiltonian $\mathcal{H}$ and the final cost function $S = S(x)$ do not depend explicitly on time then the Hamiltonian must be zero when evaluated on an locally (interior) extreme trajectory, i.e.,

$$\mathcal{H}^* = \mathcal{H}(\mathbf{X}^*(t), \mathbf{U}^*(t), \boldsymbol{\lambda}^*(t)) = 0. \tag{A.149}$$

4. Solve the deterministic optimal control problem with wealth state dynamics,

$$dX(t) = (\mu_0 - U(t))X(t)dt,$$

for $0 \le t \le t_f$, $X(0) = x_0 > 0$, μ_0 is a constant mean rate and the wealth consumption is unconstrained $-\infty < U(t) < +\infty$. The objective is maximum cumulative utility, the running consumption is the risk-adverse utility $C(x, u, t) = 2\sqrt{u}$ and similarly the utility of final wealth is $S(x, t) = 2\sqrt{x}$.

(a) Formulate the Hamiltonian $\mathcal{H}(x, u, \lambda, t)$ and the associated Hamilton's equations;

(b) Show that the optimal Hamiltonian $\mathcal{H}^*$ is a maximum at the regular point
$(X^*(t), U^*(t), \lambda^*(t), t)$, where $\lambda^*(t)$ is the optimal co-state.

(c) Show that optimal trajectories satisfy $\lambda^*(t)X^*(t) = K$ and $U^*(t) = 1/K^2 \equiv K_0$, where K is a constant.

(d) Show that K_0 is satisfied by the nonlinear equation $K_0 x_0 \exp((\mu_0 - K_0)t_f) = 1$. {*Hint: the transversality condition*

$$\lambda^*(t_f) = (\partial S/\partial x)(X^*(t_f), t_f)$$

since $X^*(t_f)$ is free and t_f is fixed.}

5. Find the maximum discounted net profit with objective function

$$C(x, u, t) = e^{-\delta_0 t}[p_0 X(t) - c_0]u\,, \quad S(x, t) = \sigma_0 x\,,$$

subject to the linear control-state dynamics,

$$\dot{X}(t) = \mu_0 X(t) - U(t)\,, \quad X(t_0) = x_0\,, \quad t_0 \le t \le t_f\,,$$

where δ_0, p_0, c_0, $\sigma_0 < 1$,μ_0 and x_0 are fixed, positive constant parameters. Assume that $X(t) < c_0/p_0$. Find intervals in parameter space where there is a maximal control solution.

Discuss the difference between the solution to this problem and the solution to a similar problem in Example A.9 with a bi-linear control-state term $U(t)X(t)$ rather than just linear in the control $U(t)$.

6. For the regular control demonstration in Example A.3 with dynamics (A.17), utility of instantaneous consumption (A.19) and terminal wealth (A.20), but with the utility power $\gamma > 1$ (for example, $\gamma = 2$), solve the Bolza problem for the proper maximum utility objective by using bang control with the bounded control constraints (A.18). Recall that the regular control solution yields a minimum rather than a maximum solution.

7. For the singular control demonstration in Example A.9 with state dynamics (A.52), cost function $C(x, u, t)$ (A.53) and control constraints (A.54), analyze the case when the maximum control $U^{(\max)}$ exceeds the mean rate μ_0, i.e., $U^{(\max)} > \mu_0$. When the parameter values permit a control solution, then find the solution; otherwise list the parameter ranges in which there fails to be a control solution.

8. Find the minimal control $U^*(t)$ for the optimal performance

$$v^*(x_1, x_2, t) = \max_U [V[X_1, X_2, U](x_1, x_2, t)]$$

of the measure

$$V[X_1, X_2, U](x_1, x_2, t) = \frac{1}{2} \int_t^{t_f} (q_1 X_1^2(s) + q_2 X_2^2(s) + rU^2(s))ds \ ,$$

$q_1 > 0$, $q_2 > 0$, $r > 0$, subject to the dynamics,

$$\dot{X}_1(t) = a_{1,1}X_1(t) + a_{1,2}X_2(t) + s_1 U(t), \ s_1 > 0 \ ,$$

$$\dot{X}_2(t) = a_{2,1}X_1(t) + a_{2,2}X_2(t) + s_2 U(t), \ s_2 > 0 \ ,$$

and the control constraints

$$|U(t)| \leq K, \ K > 0 \ ,$$

formally solving for $U^*(t)$ in terms of t, (x_1, x_2), first order partial derivatives $v^*_{x_1}(x_1, x_2, t)$ and $v^*_{x_2}(x_1, x_2, t)$ using dynamic programming. Do not solve the partial differential equation of dynamic programming but only substitute the composite formulae for $U^*(t)$ into it.

What changes in the solution form if the optimum is specified as a maximum rather than a minimum?

Suggested References for Further Reading

- Ahmed and Teo, 1981 [4].
- Anderson and Moore, 1990 [8].
- Athans and Falb 1966 [15].
- Bell and Jacobson 1975 [24].

- Betts, 2001 [30].
- Bliss, 1946 [40].
- Bryson and Ho, 1975 [44].
- Chakrabarty and Hanson, 2005 [48].
- Clark, 1976 [56].
- Gunzberger, 2003 [101].
- Kirk, 1970 [163].
- Nise, 2000 [220].
- Pontryagin et al., 1962 [226].
- *Research Directions in Distributed Parameter Systems*, 2003 [234].
- Stengel, 1986 [258].

Appendix B

Preliminaries in Probability and Analysis (Online)

It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge ... The most important questions in life are, for the most part, only problems in probability.
—Pierre-Simon Laplace, Marquis de Laplace (1749-1827) in *Théorie Analytique des Probabilités*.

I cannot believe that God would choose to play dice with the universe.
—Albert Einstein (1879-1955).

I would suggest that nobody – not even God – would know what a phrase like playing dice would mean in this context.
—Niels Henrik David Bohr (1885-1962), reply to Einstein in 1949 on the occasion of Einstein's 70th birthday continuing their famous discussion on the basis of quantum mechanics.

It is so easy to see far and discover when standing on the shoulders of giants, who before us have developed prior knowledge.
—Sir Isaac Newton (1642-1727), as quoted in [233].

There is randomness and hence uncertainty in mathematics, just as there is in physics.
—Paul Davis.

This online appendix provides a practical common background for necessary applied probability concepts for continuous and discrete random variables. These concepts include conservation of probability, expectation, variance, higher moments and basic distributions of interest. Also treated are applied analysis concepts of discontinuity and non-smoothness for deterministic processes, i.e., regular functions of time, as they affect regular calculus concepts of Taylor approximations, asymptotics and optimality principles. There is more in this appendix than many readers would be expected to know, so it should be at least be browsed for familiarity and returned to as a reference.

B.1 Distributions for Continuous Random Variables

Variables in upper case, such as $X = X(\omega)$, denote random variables, which are in general functions of some underlying random parameter or variable ω defined on some standard sample space Ω. For notational simplicity, the dependence on the underlying or background random variable $\omega \in \Omega$ will often be suppressed. Variables in lower case letters, such as x, denote the actual sample variables or realizations associated with the random variables and are used as the dummy variables in integrals.

B.1.1 Probability Distribution and Density Functions

Definition B.1. Φ *denotes the corresponding* ***probability distribution*** *such that*

$$\Phi(x) \equiv \text{Prob}[X \leq x] \,, \tag{B.1}$$

in the case of a distribution on $-\infty < X < +\infty$. *Here, the notation* Prob *denotes the probability function for the probability of occurrence of events on a subset as the ratio relative to all events in the sample space. Elsewhere many other notations are used, such as the minimal* $\boldsymbol{P}$ *and* $\boldsymbol{Pr}$.

If the **distribution is proper** then $\Phi(+\infty) = 1$, i.e., we say probability is conserved. Also, $\Phi(-\infty) = +0$ and Φ is obviously continuous as long as the probability distribution contains no jumps in value. However, later in this book, we will consider more general random processes, in continuous time, that are composed of continuous processes as well as processes with jump discontinuities, possibly a countably infinite number of jumps. Thus, in general,

Properties B.2. ***Continuous Distribution Functions****,* $\Phi(x)$*:*

- Φ *is non-decreasing, since probabilities must be non-negative.*
- Φ *is continuous, by properties of integrals with non-negative integrands (assuming there are no probability point masses, i.e., discrete components).*
- $\Phi(-\infty) = +0$*, by properties of integrals and* $X > -\infty$.

- $\Phi(+\infty) = 1$, *if* Φ *is a* ***proper distribution****.*
- $\Phi(x+y) = \Phi(x) + \text{Prob}[x < X \leq x+y]$, $y > 0$, *by the additivity of probability over disjoint sets, which here are* $(-\infty, x]$ *and* $(x, x+y]$.

Definition B.3. *The symbol* ϕ *will denote a* ***probability density*** *such that*

$$\phi(x)dx = \text{Prob}[x < X \leq x + dx] \tag{B.2}$$

in terms of the probability for the continuous random variable X.

Properties B.4. ***Relation between Distribution and Density****:*

- *By the additivity of probability and definition of the distribution function,*

$$\phi(x)dx = \text{Prob}[x < X \leq x + dx] = \Phi(x + dx) - \Phi(x) \, .$$

- *Thus, for infinitesimal* dx *and* Φ *differentiable,*

$$\phi(x)dx = \Phi'(x)dx \, ,$$

so

$$\phi(x) = \Phi'(x) \, . \tag{B.3}$$

The differentiability of the distribution Φ *is not considered a serious restriction here, since differentiability in the generalized sense will be considered in Section B.12.*

- *The relationship between the distribution function and the density in integral form is*

$$\Phi(x) \equiv \text{Prob}[X \leq x] \equiv \int_{-\infty}^{x} \phi(y)dy \, , \tag{B.4}$$

in the case of a differentiable distribution on $-\infty < X < +\infty$.

- *Another more general form is*

$$\Phi(x) \equiv \text{Prob}[X \leq x] \equiv \int_{-\infty}^{x} d\Phi(y) \, ,$$

which is called a ***Stieltjes integral****. In abstract formulations, the differential is written* $d\Phi(y) = \Phi(dy)$ *as short hand notation for* $\Phi((y, y+dy])$*, in the half-open interval notation here.*

- *Sometimes it is useful to transform the random variable* X *to a more convenient random variable* Y*, where* $X = \psi(Y)$*, for example. Consequently, for clarity of notation, let* $\phi(x) = \phi_X(x)$ *and similarly* $\Phi(x) = \Phi_X(x)$*, adding an*

extra subscript to mark which random variable pertains to a given density or distribution function since the argument x is only a dummy variable. Thus, the ***change of distribution for a change of random variable*** *on the interval $(x_1, x_2]$ is written,*

$$\begin{aligned}\Phi_X(x_2) - \Phi_X(x_1) &= \int_{x_1}^{x_2} \phi_X(x)dx \\ &= \int_{y_1}^{y_2} \phi_Y(y)dy = \Phi_Y(y_2) - \Phi_Y(y_1)\,, \end{aligned} \tag{B.5}$$

where

$$\phi_Y(y) = \phi_X(x)\left|\frac{dx}{dy}\right| = \phi_X(x)\left|\psi'(y)\right|\,, \tag{B.6}$$

provided $\psi(y)$ is a differentiable monotonic function on (y_1, y_2), i.e., either $\psi'(y) > 0$ or $\psi'(y) < 0$, where, in either case, the limits of integration are given by

$$y_1 = \min[\psi^{-1}(x_1), \psi^{-1}(x_2)]$$

and

$$y_2 = \max[\psi^{-1}(x_1), \psi^{-1}(x_2)]\,.$$

B.1.2 Expectations and Higher Moments

In general, there are basic definitions for averaged quantities in the case of continuous distributions:

Definition B.5. *The* ***mean*** *or* ***expectation*** *of any continuously distributed random variable X is*

$$\mu \equiv \mathrm{E}[X] \equiv \int_{-\infty}^{+\infty} x\phi(x)dx\,, \tag{B.7}$$

provided the above integral converges absolutely. The symbol E *is the expectation operator. Similarly, the expectation of a function of X, $f(X)$, is*

$$\mathrm{E}[f(X)] \equiv \int_{-\infty}^{+\infty} f(x)\phi(x)dx\,, \tag{B.8}$$

provided the integral converges absolutely.

Properties B.6. ***Expectations****:*

- *The expectation operator is a linear operator:*

$$\mathrm{E}[c_1X_1 + c_2X_2] = c_1\mathrm{E}[X_1] + c_2\mathrm{E}[X_2], \tag{B.9}$$

provided the expectations exist, for random variables X_i and constants c_i, for $i = 1:2$ (using MATLAB notation for the range of i).

Definition B.7. *The **variance** or **mean square deviation** or **second central moment** for any continuously distributed random variable X is*

$$\sigma^2 \equiv \mathrm{Var}[X] \equiv \mathrm{E}[(X - \mathrm{E}[X])^2] = \int_{-\infty}^{+\infty} (y-\mu)^2 \phi(y) dy \,, \tag{B.10}$$

*provided the integral converges absolutely. The deviation and the central moments are defined relative to the mean μ. The square root of the variance σ is called the **standard deviation**.*

While the mean and the variance are the most often used moments of the distribution, i.e., of the density, sometimes some of the higher moments are useful for further characterizing the distribution.

Definition B.8. *The **third central moment** is defined here in the normalized form called the **skewness coefficient** [82] for the random variable X:*

$$\eta_3[X] \equiv \mathrm{E}[(X - \mathrm{E}[X])^3]/(\mathrm{Var}[X])^{3/2}, \tag{B.11}$$

such that the distribution is negatively skewed, symmetric or positively skewed, if $\eta_3[X]$ is negative, zero or positive, respectively (zero being the skew of the normal distribution as discussed in Subsection B.1.4).

Definition B.9. *The **fourth central moment** is a measure of **kurtosis** (peakedness) and is defined here in the normalized form called the **kurtosis coefficient** [82] for the random variable X:*

$$\eta_4[X] \equiv \mathrm{E}[(X - \mathrm{E}[X])^4]/(\mathrm{Var}[X])^2, \tag{B.12}$$

*such that the distribution is **platokurtic** or **leptokurtic**, if the **coefficient of excess kurtosis** ($\eta_4[X] - 3$) is negative or positive, respectively (3 is the value of $\eta_4[X]$ for the normal distribution, discussed in Subsection B.1.4).*

The property of kurtosis, from the Greek word for convexity, signifies more at the crown (as seen from the density) for the distribution or peakedness in the case of leptokurtic and a distribution with flatness in the case of platokurtic. The kurtosis property together with skewness is of particular interest in mathematical finance for characterizing non-normal properties of real market distributions.

The little book on statistical distributions of Evans, Hastings and Peacock [82] concisely lists principal formulae for skewness, kurtosis and many other properties for forty distributions. The book has more useful and easy to find information in it than other books on distributions, including those requiring several volumes.

B.1.3 Uniform Distribution

The most fundamental continuous probability distribution is the uniform distribution.

Definition B.10. *The **uniform density** on the finite interval $[a, b]$ is defined as*

$$\phi_u(x; a, b) \equiv \left\{ \begin{array}{ll} 1/(b-a), & x \in [a, b] \\ 0, & x \notin [a, b] \end{array} \right\} . \tag{B.13}$$

Definition B.11. *The **uniform distribution** is defined by integrating the uniform density:*

$$\Phi_u(x; a, b) \equiv \int_{-\infty}^{x} \phi_u(y; a, b) dy = \left\{ \begin{array}{ll} 0, & x \leq a \\ (x-a)/(b-a), & a \leq x \leq b \\ 1, & b \leq x \end{array} \right\} , \tag{B.14}$$

$-\infty < x < +\infty$ *, so that* $\Phi_u(x; a, b) = 1$ *for* $b \leq x < +\infty$*, conserving total probability.*

Hence, the basic moments and other properties easily follow from simple integration:

Properties B.12. ***Uniform Distribution Moments**:*

- ***Conservation of probability**:* $\mathrm{E}_u[1] = 1$.
- ***Mean**:*

$$\mu = \mathrm{E}_u[X] = \int_a^b x \phi_u(x; a, b) dx = (b + a)/2 . \tag{B.15}$$

- ***Variance**:*

$$\sigma^2 = \mathrm{Var}_u[X] = \int_a^b (x - E_u[X])^2 \phi_u(x; a, b) dx = (b - a)^2/12 . \tag{B.16}$$

- ***Uniform domain correspondence to mean and variance**:* $a = \mu - \sqrt{3}\sigma$ *and* $b = \mu + \sqrt{3}\sigma$*.*
- ***Coefficient of skew**:* $\eta_3 = 0$.
- ***Coefficient of kurtosis**:* $\eta_4 = 1.8$ *or* $\eta_4 - 3 = -1.2$ *is the excess value over the normal value.*

Hence, the uniform distribution is platokurtic, signifying its obvious flatness compared to normal.

An important use of the uniform distribution is the numerical simulation of the distributions that can be transformed from the uniform distribution. The most basic random number generator is the standard uniform random number generator. The standard uniform random number generator is usually based on a deterministic generator called the linear congruential generator [230, 96] that is defined as nonzero

on the open interval $(0,1)$ instead of the closed interval $[0,1]$ as for the theoretical distribution $\phi_u(x;0,1)$, which is more convenient for numerical purposes and the endpoints do not contribute to the expectation integral anyway. Most computing systems, such as MATLAB™ [210], Maple™ [1] or *Mathematica*™ [284], and programming languages have a built-in uniform random number generator, but must be used with care considering that they use deterministic operations such as modular arithmetic, multiplication and division. These random number generators are more properly called **pseudo random number generators** since they generate only approximations to random numbers, which only exist exactly in theory. Pseudo random numbers should be carefully tested before using them in any computation. For instance, MATLAB's uniform generator is called `rand` (*note that MATLAB's functions and code fragment are typeset in* `typewriter` *style*) and can simulate an approximation to a scalar, vector or more general arrays of random numbers. Figure B.1 illustrates the histograms of a row vector with N simulations of uniform deviates for $\phi_u(x;0,1)$ using the form

$$\texttt{x} = \texttt{rand}(\texttt{N},\texttt{1})$$

or more generally

$$\texttt{y} = \texttt{a} + (\texttt{b} - \texttt{a}) * \texttt{rand}(\texttt{N},\texttt{1})$$

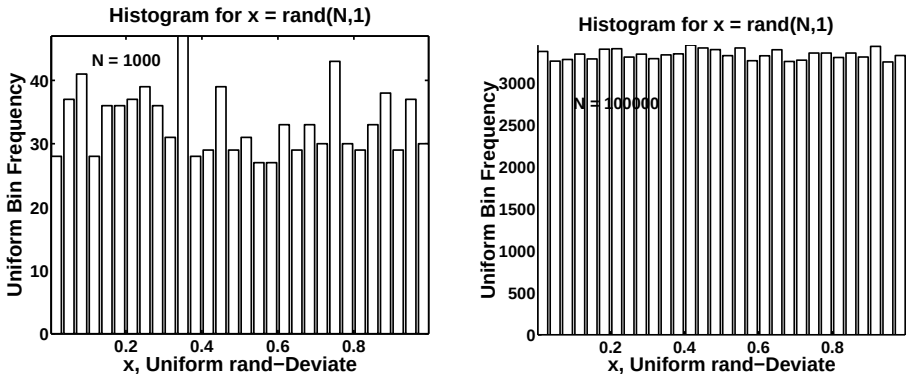
which simulates an $N-$vector sample uniform on (a,b) in MATLAB. Other computing systems may use a programming loop with N iterations may be needed. The approximate distribution displays with the bin-centered histogram function `hist(x)`. Scaling the bin frequencies upon normalizing by the average bin count `N/nbins`, where `nbins` is the number of bins, here 30 bins, would produce a scaled histogram more appropriate for approximating probability density, $\phi_u(x;0,1)$, of the theoretical uniform distribution. Thus, if f_i is the frequency associated with the ith bin $[x_i, x_i + \Delta x)$ `for i = 1 : nbins`, in MATLAB loop notation, of width Δx, then

$$\sum_{i=1}^{\texttt{nbins}} f_i = N \quad \text{or} \quad \frac{1}{N}\sum_{i=1}^{\texttt{nbins}} f_i = 1,$$

the latter in normalized form.

Clearly, the larger sample size simulation with $N = 100,000$ in Fig. B.1(b) is a much better approximation of the uniform approximation then the much cruder representation with $N = 1,000$ in Fig. B.1(a). The relative error for the sample mean is -0.24% for $N = 1,000$ and -0.43% for $N = 100,000$.

Note that the error in the sample mean did increase slightly with sample size, but these are only single samples and another set of samples could have been computed that would be used in the expected decreasing order with sample size, yet not realistic. These are just approximations to random samples, although it would be reasonable to expect that the average over repeated samples would be lower the higher the sample size, provided that the selected random number generator is sufficiently robust. The relative errors for the sample standard deviation (square root of the sample variance), are 0.95% for $N = 1,000$ and -0.20% for $N = 100,000$, which is more reasonable.



(a) Sample size $N = 10^3$. (b) Sample size $N = 10^5$.

Figure B.1. *Histograms of simulations of uniform distribution on* $(0, 1)$ *using MATLAB [210] for two different sample sizes* N*.*

The sample variance is obtained from the MATLAB function `var(x)`, which is normalized by number of degrees of freedom $(N-1)$ for the best estimate of the variance, correcting for conditioning due to the mean value, which in MATLAB is the function `mean(x)`.

For more sophisticated distribution validation tests, chi-square (χ^2) or Kolmogorov-Smirnoff [230] tests can be used. The two samples displayed in Fig. B.1 illustrate the problem of single samples requiring the averaging of several independent replications using a different random number generator initialization, called a **random seed** but now called a `state` in MATLAB (e.g., `rand('state', j)` sets `rand` in the `j`th `state`), so the error systematically decreases with sample size. Otherwise. the user can take a larger sample size. See Appendix C Section C.1 for the MATLAB figure code.

In this appendix, we present empirical representations of distributions by histograms derived from random number generation, rather than the purely mathematical graphs of the probability density as portrayed in probability and statistics texts. This is to emphasize that the distributions derived from real environments are not as ideal as the exact mathematical density functions. Another reason is to emphasize that sometimes computations are necessary when no exact solutions are available or useful when exact solutions are too complicated, beyond the expertise of the entry-level graduate student or advanced undergraduate student.

B.1.4 Normal Distribution and Gaussian Processes

A continuous distribution of interest for Gaussian processes and other applications is given in terms of the normal probability density, the derivative of the normal or Gaussian probability distribution.

Definition B.13. *The* ***normal density*** *with mean* $\mu = \mathrm{E}_n[X]$ *and* $\sigma^2 = \mathrm{Var}_n[X]$

is defined as

$$\phi_n(x;\mu,\sigma^2) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) , \quad -\infty < x < +\infty \, , \; \sigma > 0 \, , \tag{B.17}$$

where ϕ_n denotes the normal density function with argument x and parameters $\{\mu,\sigma^2\}$ following the semicolon. Here, X is called the ***normal random variate.***

Definition B.14. *The* ***normal distribution*** *is defined here through the density as*

$$\Phi_n(x;\mu,\sigma^2) \equiv \int_{-\infty}^{x} \phi_n(y;\mu,\sigma^2)dy \, , \quad -\infty < x < +\infty \, , \tag{B.18}$$

so that $\Phi_n(+\infty;\mu,\sigma^2) = 1$, conserving total probability.

Remark B.15. *The normal distribution can be computed using MATLAB, Maple or Mathematica computing systems, but the common special function that can be used, without resorting to special packages, is the* ***error function complement****,*

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-t^2} dt \, , \tag{B.19}$$

so that the normal distribution can be computed from these two identities

$$\Phi_n(x;\mu,\sigma^2) = \frac{1}{2}\mathrm{erfc}\left(\frac{\mu - x}{\sqrt{2}\sigma}\right) \tag{B.20}$$

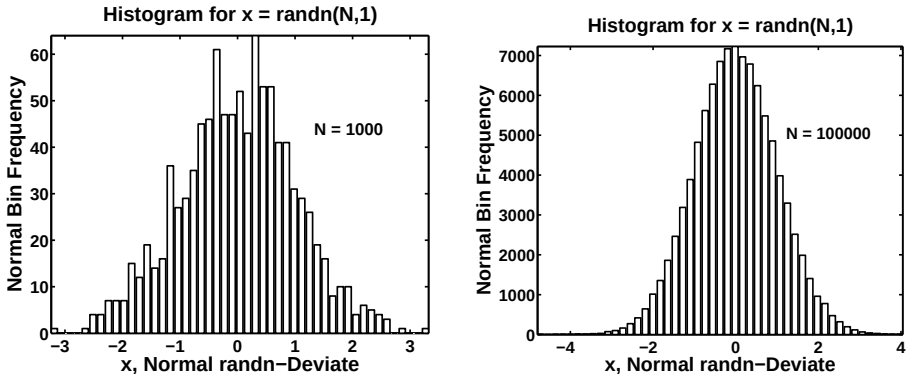
$$= 1 - \frac{1}{2}\mathrm{erfc}\left(\frac{x - \mu}{\sqrt{2}\sigma}\right) \, . \tag{B.21}$$

Properties B.16. ***Normal Distribution Skew and Kurtosis****:*

- *The normal distribution is* ***skewless****, since the coefficient of skew is $\eta_3[X] = 0$.*
- *The normal distribution has no* ***excess kurtosis****, since the coefficient of excess kurtosis is $(\eta_4[X] - 3) = 0$, where 3 is the coefficient of kurtosis of the normal distribution.*

As with the uniform distribution, the normal distribution is a theoretical idealization that is very useful in the analysis of stochastic problems. However, for practical computations numerical simulations are usually necessary. Since the normal density function is an exponential of a quadratic, direct transformation from a uniform random generator is not directly possible. However, the usual normal random number generating algorithm, called Box-Muller [230, 96], cleverly applies the uniform random generator to a polar coordinate version of a two-dimensional normal distribution, reminiscent of the classic technique of converting a normal probability integral on the infinite domain from one dimension to two

dimensions and polar coordinates to get exact integral values. In some computing systems there is a special built-in function for a normal random generator. In MATLAB [210] the function is called `randn`, also having vector or array capabilities in the vector form $\mathtt{x} = \mathtt{randn}(\mathtt{N}, 1)$ for a $N-$vector sample (more generally, $\mathtt{y} = \mathtt{mu} + \mathtt{sigma} * \mathtt{randn}(\mathtt{N}, 1)$ would simulate the density $\phi_n(y; \mathtt{mu}, \mathtt{sigma}^2)$ where `mu` is the specified mean and `sigma` is the specified standard deviation). (Note that MATLAB's functions, code variables such as `mu` and `sigma` and code fragments are set in `typewriter` style.) The simulated normal density is illustrated by the histogram in Fig. B.2 using two sample sizes, $N = 1,000$ and $100,000$. Clearly, the larger sample size in Fig. B.2(b) gives a better qualitative representation of the theoretical bell-shaped curve of the normal density $\phi_n(x; 0, 1)$. The percent relative errors in the mean and standard deviation are respectively -1.53% and -0.35% for $N = 1,000$, while the errors are 1.31% and -0.083% for the for $N = 100,000$ sample size. See Appendix C Section C.2 for the MATLAB figure code.



(a) Sample size $N = 10^3$. (b) Sample size $N = 10^5$.

Figure B.2. *Histograms of simulations of the standard normal distribution with mean* 0 *and variance* 1 *using MATLAB [210] with 50 bins for two sample sizes N. The histogram for the large sample size of* $N = 10^5$ *in Fig. B.2(b) exhibits a better approximation to the theoretical normal density* $\phi_n(x; 0, 1)$.

B.1.5 Simple Gaussian Processes

For later use, we will let $W(t)$ denote what is called a standard, mean zero Wn zero Wiener process with distribution

$$\Phi_{W(t)}(x) = \Phi_n(x; 0, t)\,, \quad -\infty < x < +\infty\,, \quad t > 0\,, \tag{B.22}$$

with corresponding probability density

$$\phi_{W(t)}(x) = \phi_n(x; 0, t)\,, \quad -\infty < x < +\infty\,, \quad t > 0\,. \tag{B.23}$$

A simple **Gaussian process** with linear mean growth in time,

$$X = G(t) = \mu t + \sigma W(t)\,, \tag{B.24}$$

has mean $\mathrm{E}[X] = \mu t$ and variance $\mathrm{Var}[X] = \sigma^2 t$, so that the distribution of this process is

$$\Phi_{G(t)}(x) = \Phi_n(x; \mu t, \sigma^2 t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{x} e^{-\frac{(y-\mu t)^2}{2\sigma^2 t}} dy \,, \tag{B.25}$$

on $-\infty < x < +\infty$, $t > 0$. The standard Wiener and Gaussian processes are also called diffusion processes, so they form models of the diffusion part of the jump-diffusion processes that are the main topic in this book. In order to see the connection between the stochastic Gaussian process and the deterministic diffusion process, let

$$u(x,t) = \Phi_{G(t)}(x)$$

and take partial derivatives of $u(x,t)$ with respect to t and x to derive the diffusion equation with **drift** $(-\mu)$ and **diffusion coefficient** $(\sigma^2/2)$,

$$u_t(x,t) = -\mu u_x(x,t) + \tfrac{\sigma^2}{2} u_{xx}(x,t) \,, \quad -\infty < x < +\infty \,, \quad t > 0 \,. \tag{B.26}$$

where the subscripts on u_t, u_x and u_{xx} denote partial derivatives and the equation is called a partial differential equation (PDE).

Remarks B.17.

- *Here we use the term Gaussian process as it is used in applied mathematics, science and engineering, i.e., for processes that are normally distributed. (For a more abstract view of Gaussian processes, see Mikosch [209]).*
- *There will be much more on the Wiener and Gaussian processes later, since they form the basic process for building the diffusion component of the jump-diffusion processes.*

B.1.6 Lognormal Distribution

Often in applications, such as in many linear financial models, the exponential of a normally distributed random variable arises and the distribution of this exponential is called a **lognormal distribution** since its logarithm produces the normally distributed exponent.

Theorem B.18. *Let*

$$X_{ln} = \exp\left(\mu + \sigma X_n\right) \tag{B.27}$$

be a ***lognormal variate*** *and let X_n be a standard normal variate, i.e., with zero mean and unit variance, Then* ***lognormal density*** *with mean $\mu_{ln} = \mathrm{E}[X_{ln}]$ and $(\sigma_{ln})^2 = \mathrm{Var}[X_{ln}]$ can be written in terms of the normal density ϕ_n (B.17) such that*

$$\phi_{ln}\left(x; \mu_{ln}, (\sigma_{ln})^2\right) \equiv x^{-1} \phi_n\left(\ln(x); \mu, \sigma^2\right) \quad 0 < x < +\infty \,,\ \sigma > 0 \,, \tag{B.28}$$

where ϕ_{ln} denotes the lognormal density function with argument x and parameters $\{\mu_n, (\sigma^2)_n\} = \{\mu, \sigma^2\}$ follow the semicolon. If $x = 0$, then define ϕ_{ln} as the limiting case:

$$\phi_{ln}\left(0; \mu_{ln}, (\sigma_{ln})^2\right) \equiv \phi_{ln}\left(0^+; \mu_{ln}, (\sigma_{ln})^2\right) = 0\,. \tag{B.29}$$

Proof. Let the realization variable $x > 0$ and note that $\sigma > 0$, while the natural logarithm is an increasing function. Consider the corresponding lognormal distribution definition, subsequently manipulated into the normal distribution:

$$\begin{aligned}
\Phi_{ln}\left(x; \mu_{ln}, (\sigma_{ln})^2\right) &= \text{Prob}\left[X_{ln} \le x\right] && \text{(B.30)}\\
&= \text{Prob}\left[\exp\left(\mu + \sigma X_n\right) \le x\right] && \text{(B.31)}\\
&= \text{Prob}\left[X_n \le (\ln(x) - \mu)/\sigma\right] && \text{(B.32)}\\
&= \Phi_n((\ln(x) - \mu)/\sigma; 0, 1) && \text{(B.33)}\\
&= \Phi_n(\ln(x); \mu, \sigma^2)\,. && \text{(B.34)}
\end{aligned}$$

The last step follows a normal distribution or density identity that allows transforming from the standard normal to non-standard normal with mean μ and variance σ^2 (see Exercise 9) on Page B72). Upon taking the derivatives of the first and the last of this chain of equations, using the chain rule to handle the logarithmic argument of the normal distribution, the relationship between the densities is

$$\begin{aligned}
\phi_{ln}\left(x; \mu_{ln}, (\sigma_{ln})^2\right) &= (\Phi_{ln})'\left(x; \mu_{ln}, (\sigma_{ln})^2\right)\\
&= x^{-1}(\Phi_n)'\left(\ln(x); \mu, \sigma^2\right)\\
&= x^{-1}\phi_n\left(\ln(x); \mu, \sigma^2\right)\,.
\end{aligned}$$

Note that as $x \to 0^+$ then

$$x^{-1}\exp\left(-(\ln(x) - \mu)^2/\left(2\sigma^2\right)\right) \to 0^+\,,$$

since the exponential approaches zero much faster than the reciprocal of x approaches infinity. Thus, since the singularity at zero is removable, we define the exception value of the lognormal density at zero to be

$$\phi_{ln}\left(0; \mu_{ln}, (\sigma_{ln})^2\right) \equiv \phi_{ln}\left(0^+; \mu_{ln}, (\sigma_{ln})^2\right) = 0\,.$$

□

In the above analytical manipulation of distribution probabilities, the general principle are embodied in the following lemma:

Lemma B.19. ***General Probability Inversion***
Let X and Y be two random variables with continuous densities $\phi_X(x)$ and $\phi_Y(y)$, respectively. Further, let the dependence between them be given by $X = g(Y)$, where $g(y)$ is continuously differentiable and increasing so that an inverse function

f exists, i.e., $y = f(x) = g^{-1}(x)$. Then the corresponding distributions are related by

$$\begin{aligned}\Phi_X(x) &= \text{Prob}[X \le x] = \text{Prob}[g(Y) \le x] \\ &= \text{Prob}[Y \le f(x)] = \Phi_Y(f(x))\end{aligned} \tag{B.35}$$

and the densities are related by

$$\phi_X(x) = f'(x)\phi_Y(f(x))\,. \tag{B.36}$$

If, instead, g is strictly decreasing, then

$$\Phi_X(x) = \text{Prob}[Y \ge f(x)] = 1 - \Phi_Y(f(x)) \tag{B.37}$$

and

$$\phi_X(x) = -f'(x)\phi_Y(f(x))\,. \tag{B.38}$$

Proof. Since f is the inverse function of g then with $x = g(y)$ and $y = f(x)$, $g(f(x)) = x$ and $g'(y)f'(x) = 1$, using the chain rule and the derivatives are reciprocals of each other. Further, the increasing property of g means f is also increasing, the signs of the derivatives must be the same. So if $x_1 \le x_2$ then $f(x_1) \le f(x_2)$, the direction of an inequality is preserved upon application of f. In the g decreasing case, the direction is reversed. Thus, Eq. (B.35) has been demonstrated in the increasing case. The decreasing case is similar, except for the change in inequality direction and a minor point in converting from probability to distribution function. The probability complement equivalent of $\text{Prob}[Y \ge f(x)]$ would strictly be $1 - \text{Prob}[Y < f(x)]$, but since the densities are continuous the probabilities assigned to an isolated point are zero, i.e., $\text{Prob}[Y < f(x)] = \text{Prob}[Y \le f(x)]$.

The densities follow upon differentiating by the chain rule,

$$\Phi'_X(x) = \phi_X(x) = f'(x)\Phi'_Y(f(x)) = f'(x)\phi_Y(f(x))$$

in the increasing case and the decreasing case is similar except for the minus sign in the density (B.38), which also preserves the non-negativity of the density, since $-f'(x) > 0$ in the negative case. The factor $\pm f'(x) > 0$ is the density conversion factor in either case. □

Properties B.20. ***Lognormal Distribution Moments:***

- ***Mean:***
$$\mu_{ln} = \text{E}_{ln}[X] = e^{\mu+\sigma^2/2}\,.$$
- ***Variance:***
$$\sigma_{ln} = \text{Var}_{ln}[X] = (\mu_{ln})^2\left(e^{\sigma^2} - 1\right)\,,$$

- ***Inverse, Normal from Lognormal:***

$$\sigma^2 = \ln\left(1 + \sigma_{ln}/\left(\mu_{ln}\right)^2\right)$$

and

$$\mu = \ln\left(\mu_{ln}\right) - \frac{1}{2}\sigma^2 \,.$$

- ***Coefficient of skewness:***

$$\eta_3^{(ln)}[X] = \left(e^{\sigma^2} + 2\right)\sqrt{e^{\sigma^2} - 1}\,.$$

- ***Coefficient of kurtosis:***

$$\eta_4^{(ln)}[X] = \left(e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3\right)\,.$$

Remark B.21. *The mean formula is justified using the logarithmic transformation, $y = (\ln(x) - \mu)/\sigma$, from lognormal back to normal along with completing the square method in the exponent,*

$$\begin{aligned}
\mathrm{E}_{ln}[X] &= \int_0^\infty \frac{\exp(-(\ln(x)-\mu)^2/(2\sigma^2))}{x\sqrt{2\pi\sigma^2}} x dx \\
&= \frac{1}{\sqrt{2\pi}} e^{\mu} \int_{-\infty}^{+\infty} e^{-y^2/2} e^{\sigma y} dy \\
&= \frac{1}{\sqrt{2\pi}} e^{\mu+\sigma^2/2} \int_{-\infty}^{+\infty} e^{-(y-\sigma^2)^2/2} dy = e^{\mu+\sigma^2/2}\,.
\end{aligned}$$

Then the rest of the moments rely on the same techniques.

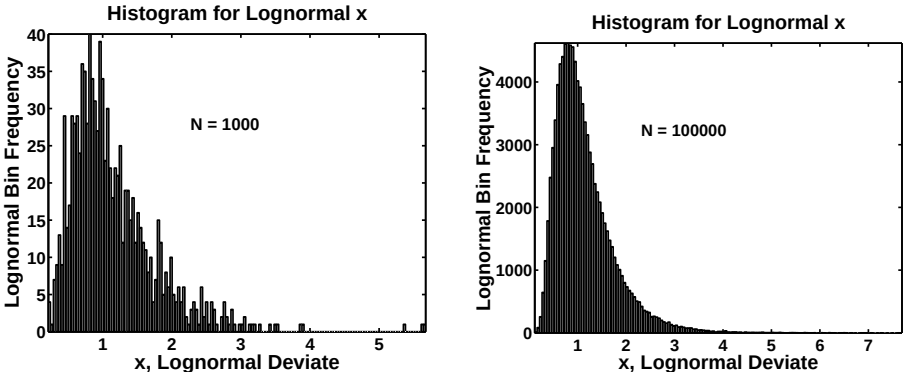
The simulation of the lognormal distribution relies on the fact (B.27) that the lognormal variate is the exponential of a normal variate, i.e., $X_{ln} = \exp(\mu + \sigma X_n)$. Thus the MATLAB approximation will be the set of simulations,

```
y = mu*ones(N,1) + sigma*randn(N,1);
x = exp(y);
```

where again `randn(N,1)` is MATLAB's normal random generator for a sample size of N while the `ones(N,1)` function produces an N-vector of ones preserving the vector form when adding the constant `mu`, with similar constructs in Maple and *Mathematica*. Eq. (B.28) for the density implies that the proper lognormal density will be obtained in theory.

The MATLAB graphical histogram output for two sample sizes, $N = 1,000$ and $100,000$, both sorted into `nbins`$= 150$, is given in Fig. B.3. The selected normal parameters are $\mu_n = \mu =$ `mu` $= 0.0$ and $\sigma_n = \sigma =$ `sigma` $= 0.5$, corresponding to lognormal parameters $\mu_{ln} \simeq 1.133$ and $\sigma_{ln} \simeq 0.3646$. The percent relative

errors in the lognormal mean and standard deviation are respectively -0.56% and -0.60% for $N = 1,000$, while the relative errors are -0.085% and -0.30% for the for $N = 100,000$ sample size. Again, the larger sample size Fig. B.3(b) gives a better qualitative representation of the theoretical shape of the lognormal density $\phi_{ln}(x; \mu_{ln}, \sigma_{ln})$. Both subfigures confirm that the density goes to zero as $x \to 0^+$. See Appendix C Section C.3 for the MATLAB figure code.



(a) Sample size $N = 10^3$. (b) Sample size $N = 10^5$.

Figure B.3. *Histograms of simulations of the lognormal distribution with mean $\mu_n = 0$ and variance $\sigma_n = 0.5$ using MATLAB [210] normal distribution simulations,* `x = exp(mu*ones(N,1) + sigma*randn(N,1))`*, with 150 bins for two sample sizes. The histogram for the large sample size of $N = 10^5$ in Fig. B.3(b) exhibits a better approximation to the theoretical lognormal density $\phi_n(x; 0, 1)$ than the one in Fig. B.3(a).*

B.1.7 Exponential Distribution

The continuous exponential density is closely related to the inter-arrival time of a Poisson process (discussed in Chapter 1).

Definition B.22. *The **exponential density** is given for some exponential random variate τ_e by*

$$\phi_e(t; \mu) \equiv \frac{1}{\mu} e^{-t/\mu} , \quad 0 \le t < +\infty , \; \mu > 0 , \tag{B.39}$$

*with mean μ, so the exponential distribution is called a **one-parameter distribution**. The explicit form of the **exponential distribution** is*

$$\Phi_e(t; \mu) = \text{Prob}[\tau_e \le t] = \left\{ \begin{array}{ll} 1 - e^{-t/\mu} , & t \ge 0 \\ 0 , & t < 0 \end{array} \right\} . \tag{B.40}$$

Properties B.23. ***Exponential Distribution Moments:***

- ***Conservation of probability:*** $\text{E}_e[1] = 1$.

- ***Mean:*** $\mu = \mathrm{E}_e[X]$ *, by selection of the exponential parameter.*
- ***Variance:*** $\sigma^2 = \mathrm{Var}_e[X] = \mu^2$ *, so the standard deviation is also* μ *.*
- ***Coefficient of skew:*** $\eta_3 = 2$ *, positive relative to the mean on* $[0, \infty)$ *.*
- ***Coefficient of kurtosis:*** $\eta_4 = 9$ *or* $\eta_4 - 3 = 6$ *is the excess value over the normal value.*

Hence, the exponential distribution defines a one-parameter family of distributions with the same mean and standard deviation, but also positively skewed by virtue of the semi-infinite domain and leptokurtic with clear pointedness.

Since the **exponential distribution** has such a simple form it can easily be transformed into the uniform distribution for use in practical simulations. Using **fundamental law** of **transformation** of **probabilities** [230] or as the **inverse transformation method** [96] for transforming the exponential density $\phi_e(x_e; \mu)$ to the standard $(0, 1)$ uniform density $\phi_u(x_u; 0, 1)$,

$$\phi_u\left(x_u; 0, 1\right) = \phi_e\left(x_e; \mu\right) \left| \frac{dx_e}{dx_u} \right| , \tag{B.41}$$

choosing the Jacobian sign negative, $dx_e/dx_u < 0$, because it leads to a faster computational form by eliminating a constant of integration, so that,

$$x_e = -\mu \ln\left(x_u\right) , \tag{B.42}$$

or in inverse form

$$x_u = \exp\left(-x_e/\mu\right) . \tag{B.43}$$

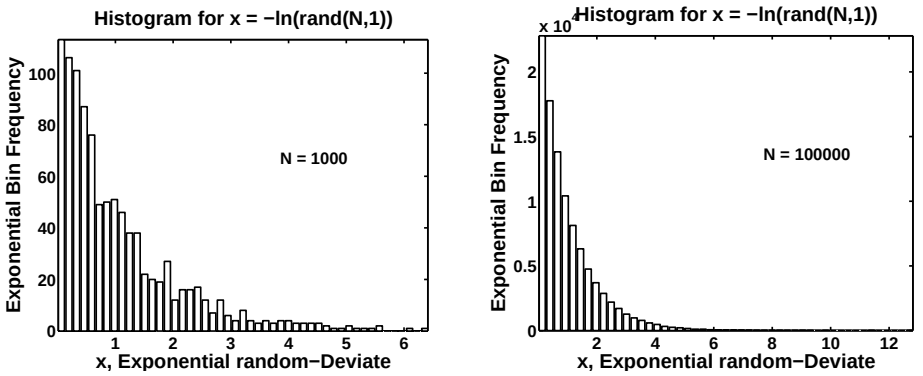
A prime prerequisite for random simulations is that the distribution is covered in the transformation, but the order of the covering does not matter so

$$\begin{aligned} \Phi_e(x_e; \mu) &= \mathrm{Prob}\left[0 \le X_e \le x_e\right] \\ &= \mathrm{Prob}\left[\exp\left(-x_e/\mu\right) \le X_u \le 1\right] \\ &= 1 - \Phi_u\left(\exp\left(-x_e/\mu\right); 0, 1\right) . \end{aligned}$$

works even though the uniform distribution here is covered from right to left starting from 1 while the exponential distribution is covered from left to right starting form $x_e = 0$. The interested reader can check that the general expectation $\mathrm{E}_e[f(X_e)] = \mathrm{E}_u[f(-\mu \ln(X_u))]$ is equivalent for any integrable function f (see Exercise 12).

Hence, `x = −mu * log(rand(N, 1))` leads to a MATLAB exponential random generator producing $N-$vector output, where `log` is the MATLAB natural logarithm function and `mu` is the input for the mean. The MATLAB graphical output for two sample sizes, $N = 1,000$ and $100,000$, is given in Figs. B.4(a) and B.4(b), respectively. The percent relative errors in the mean and standard deviation are respectively 7.94% and -0.71% for $N = 1,000$, while the errors are 3.81% and -0.54% for the for $N = 100,000$ sample size. See Appendix CSection C.4 for the MATLAB figure code.

Remarks B.24.



(a) Sample size $N = 10^3$. (b) Sample size $N = 10^5$.

Figure B.4. *Histograms of simulations of the standard exponential distribution, with mean taken to be* $\texttt{mu} = 1$*, using MATLAB's* `hist` *function [210] with 50 bins for two sample sizes* N*, generated by* $\texttt{x} = -\texttt{mu} * \texttt{log}(\texttt{rand}(\texttt{N}, 1))$ *in MATLAB . The histogram for the large sample size of* $N = 10^5$ *in Fig. B.4(b) exhibits a better approximation to the standard theoretical exponential density* $\phi_e(x; 1)$*.*

- *Alternatively, a more direct exponential to uniform transformation could have been selected,*

$$\hat{x}_u = 1 - \exp\left(-\hat{x}_e/\mu\right)$$

with inverse

$$\hat{x}_e = -\mu \ln\left(1 - \hat{x}_u\right) , \tag{B.44}$$

but that would not be as numerically efficient for large sample sizes N *as (B.42) which is more often used, since (B.42) requires one less floating point operation, not needing to subtract the uniform random sample from 1 per sample in (B.44). Typically random sample sizes are huge, so good representations of the distribution can be obtained.*

- *The probabilistic view of the difference between the two exponential to uniform transformations follows from Lemma B.19 on general probability inversion. In the direct case,* $\hat{g}(y) = -\mu \ln(1 - y)$ *and* $\hat{f}(x) = 1 - \exp(-x/\mu)$*, so* $g'(y) = +\mu/(1 - y) > 0$ *for* $0 < y < 1$*. Thus,*

$$\Phi_{\widehat{X}_e}(x) = \Phi_{\widehat{X}_u}(1 - \exp(-x/\mu))$$

by (B.35) and

$$\phi_{\widehat{X}_e}(x) = \frac{1}{\mu} \exp(-x/\mu) \phi_{\widehat{X}_u}(1 - \exp(-x/\mu))$$

by (B.36), which implies $\phi_{\widehat{X}_u}(1 - \exp(-x/\mu)) = 1$ *since its coefficient is* $\phi_{\widehat{X}_e}(x)$*. In the more useful case,* $g(y) = -\mu \ln(y)$ *and* $f(x) = \exp(-x/\mu)$*,*

so $g'(y) < 0$ *and*

$$\phi_{X_e}(x) = +\frac{1}{\mu}\exp(-x/\mu)\phi_{X_u}(\exp(-x/\mu))$$

by (B.38) and again $\phi_{X_u}(\exp(-x/\mu)) = 1$.

B.2 Distributions of Discrete Random Variables

In general, averaged quantities for discrete distributions involve sums rather than integrals used in the continuous distributions. (Note: the use of the term distribution is different for discrete and continuous cases.)

Definition B.25. *Let the* ***discrete distribution*** *be*

$$\pi_k = \text{Prob}[X = x_k] \tag{B.45}$$

for some countable set of values $\mathcal{X} = \{x_k | k = 0 : m\}$*, where* m *could be infinite (the* $1 : n$ *is MATLAB loop notation).*

Definition B.26. ***Colon or Loop Notation****:*
For compactness, the range of a discrete set will be in the MATLAB colon or loop notation [210, 142] with $k = m_1 : m_2$ *denoting that the index* k *ranges from integers* m_1 *to* m_2 *in steps of unity (1), meaning the same as the loosely defined* $k = m_1, m_1 + 1, \ldots, m_2 - 1, m_2$*, assuming* $m_1 < m_2$*. In the case of non-unit steps* Δm*, then* $k = m_1 : \Delta m : m_2$ *is used instead of* $k = m_1, m_1 + \Delta m, \ldots, m_2 - \Delta m, m_2$*, assuming the range* $m_2 - m_1$ *is a positive integer multiple of* Δm*.*

Properties B.27. ***Discrete Distributions*** π_k***:***

- ***Non-negativity****:* $\pi_k \geq 0$.
- ***Conservation of probability****:*

$$\sum_{k=0}^{m} \pi_k = 1\,. \tag{B.46}$$

The basic definitions in the discrete distribution case for averaged quantities are listed:

Definitions B.28.

- *The* ***mean*** *or* ***expectation*** *of the* ***discrete set*** $\mathcal{X} = \{x_k | k = 0 : m\}$ *is*

$$\mu = \text{E}[\mathcal{X}] \equiv \sum_{k=0}^{m} x_k \pi_k\,, \tag{B.47}$$

for any discretely distributed random variable provided the sum converges absolutely.

- *Similarly, the* ***expectation of a function $f(X)$ of X*** *is*

$$\mathrm{E}[f(X)] \equiv \sum_{k=0}^{m} f(x_k)\pi_k \,, \tag{B.48}$$

provided the sum converges absolutely.

Definition B.29. *The* ***variance*** *or* ***mean square deviation*** *of the discrete set $\mathcal{X}$ is*

$$\mathrm{Var}[\mathcal{X}] \equiv \mathrm{E}[(\mathcal{X} - \mathrm{E}[\mathcal{X}])^2] = \sum_{k=0}^{m} (x_k - \mu)^2 \pi_k \,, \tag{B.49}$$

for any discretely distributed random variable provided the sum converges absolutely, where the set difference $(\mathcal{X} - \mu) \equiv \{x_k - \mu | k = 0 : m\}$ for fixed μ.

B.2.1 Poisson Distribution and Poisson Process

Another important distribution is a **discrete distribution** and is called the Poisson distribution. It is useful for modeling jumps, especially for the jump component of jump-diffusions.

Definition B.30. *The* ***Poisson distribution*** *with Poisson variate ν and single Poisson parameter Λ is given by the probabilities*

$$p_k(\Lambda) \equiv \mathrm{Prob}[\nu = k] = e^{-\Lambda} \frac{(\Lambda)^k}{k!} \,, \tag{B.50}$$

for $k = 0, 1, 2, \ldots$ and $\Lambda \geq 0$, expressed as a simple Poisson distribution with continuous parameter Λ which serves as both mean

$$\mathrm{E}[\nu] = \Lambda \tag{B.51}$$

and variance

$$\mathrm{Var}[\nu] = \Lambda \tag{B.52}$$

of this one-parameter discrete distribution.

The mean and variance can conveniently be computed from the properties of the exponential series,

$$\sum_{k=0}^{\infty} \frac{u^k}{k!} = e^u = \exp(u) \,, \; -\infty < u < +\infty \,, \tag{B.53}$$

together with its derivatives such as its first derivative form

$$\sum_{k=0}^{\infty} k \frac{u^k}{k!} = u \frac{d}{du} e^u \,,$$

which can be used to compute the mean property from

$$\mathrm{E}[\nu] = e^{-\Lambda} \sum_{k=0}^{\infty} k \frac{(\Lambda)^k}{k!}$$

to derive (B.51) and its second derivative form

$$\sum_{k=0}^{\infty} k^2 \frac{u^k}{k!} = \left(u \frac{d}{du} \right)^2 e^u \,,$$

which can be used with the mean to compute the variance property from

$$\mathrm{Var}[\nu] = e^{-\Lambda} \sum_{k=0}^{\infty} (k - \Lambda)^2 \frac{(\Lambda)^k}{k!}$$

to derive (B.52) upon expanding the square in the sum.

From (B.50), it is simple to deduce that $p_k(0^+) = \delta_{k,0}$, where $\delta_{k,0}$ is defined:

Definition B.31.

$$\delta_{i,j} = \left\{ \begin{array}{lll} 1 & \text{if} & j = i \\ 0 & \text{if} & j \neq i \end{array} \right\} \tag{B.54}$$

is the ***Kronecker delta*** *or discrete delta function.*

Figure B.5 illustrates the Poisson distribution versus the Poisson counting variable k for four values of the Poisson parameter, $\Lambda = 0.2, 1.0, 2.0$ and 5.0. See Appendix C Section C.5 for the MATLAB figure code. For the smaller parameter

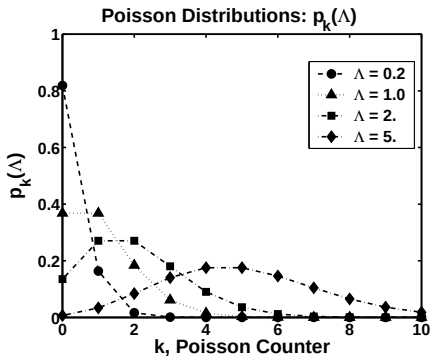


Figure B.5. *Poisson distributions with respect to the Poisson counter variable k for parameter values $\Lambda = 0.2, 1.0, 2.0$ and 5.0. These represent discrete distributions, but discrete values are connected by dashed, dotted and dash-dotted lines only to help visualize the distribution form for each parameter value.*

value, $\Lambda = 0.2$, the distribution resembles a discretized version of the exponential

distribution, while as Λ increases to 2.0 the distribution is beginning to resemble the normal distribution around the peak. For large values of the parameter Λ it can be shown (Feller, [83]) that the Poisson distribution has a normal approximation.

For later use, let $P(t)$ denote the simple Poisson process with linear time dependent parameter $\Lambda = \lambda t$ is a jump process with unit jumps, hence also characterized as a counting process. It can be shown (see Çinlar [55], for instance) that $P(t)$ discrete distribution is

$$p_k(\lambda t) \equiv \text{Prob}[P(t) = k] = e^{-\lambda t}\frac{(\lambda t)^k}{k!}\,. \tag{B.55}$$

If the random variable T_k is the time of the kth Poisson unit jump for $k = 0 : +\infty$, then time between jumps or inter-arrival time can be shown to be distributed exponentially:

$$\begin{aligned}\text{Prob}[T_{k+1} - T_k \le t \,|\, T_k] &= 1 - \text{Prob}[T_{k+1} - T_k > t \,|\, T_k] \\ &= 1 - \text{Prob}[P(T_k + t) - P(T_k) = 0 \,|\, T_k] \\ &= 1 - \text{Prob}[P(t) = 0] \\ &= 1 - e^{-\lambda t} = \Phi_e(t; 1/\lambda)\,,\end{aligned} \tag{B.56}$$

in the first step using conservation of probability to write the probability in terms of one minus the complement, in the second step using the fact that the probability that the inter-arrival time $\Delta T_k = T_{k+1} - T_k > t$ is the same as the probability that Poisson increment $P(T_k + t) - P(T_k) = 0$, in the third step using the stationarity property that $P(s+t) - P(s)$ and $P(t)$ have the same distribution (to be discussed later), and finally using Eq. (B.55) with $k = 0$.

Remark B.32. *More on the Poisson process will be presented in the main chapters of the text, since it serves as the basic process for building the jump component of the jump-diffusion processes.*

B.3 Joint and Conditional Distribution Definitions

In many part of this book, several properties of joint and conditional distributions will be useful and are summarized for two random variables here. These random variables can be combinations of discrete and continuous random variables, e.g., discrete for jump variables or continuous for diffusion variables. The definition forms are the forms that are useful in this text, but they are not necessarily the most general definitions. Many can be easily generalized from a couple to multiple random variables. For more general information see the long time standard reference of Feller [84] or the references of Karlin and Taylor [161, 265].

Definitions B.33. ***Jointly Distributed Random Variables***

- *The **joint probabilities** or **joint distribution functions** of two random variables X and Y depend on whether the random variables are discrete or continuous, leading to three cases:*

1. ***Two jointly distributed discrete random variables**, X and Y, have the **joint probability** or **joint distribution** function*

$$\pi_{X,Y}(x_i, y_j) \equiv \text{Prob}[X = x_i, Y = y_j] \,, \tag{B.57}$$

for specified discrete values x_i and y_j for integers i and j (in general, the discrete sets are assumed to be countable or denumerable) and such values will be assumed with the qualifications given here;

2. ***Two jointly distributed continuous random variables**, X and Y, have the **joint probability** or **joint distribution** function*

$$\Phi_{X,Y}(x, y) \equiv \text{Prob}[X \leq x, Y \leq y] \,; \tag{B.58}$$

3. ***Two jointly distributed mixed continuous and discrete random variables**, X and Y, have the hybrid **joint probability** or **joint distribution** function*

$$\Phi_{X,Y}(x, y_j) \equiv \text{Prob}[X \leq x, Y = y_j] \,, \tag{B.59}$$

for some discrete value y_j.

- *The **joint densities**, if they exist, of two jointly distributed random variables X and Y, are defined as*

1. ***Two jointly distributed discrete random variables**, X and Y, do not have a **joint density** in the usual way, but for an applied formulation, the generalized functions can be used (see Section B.12 on page B53).*

2. ***Two jointly distributed continuous random variables**, X and Y, have the **joint density** if the partial derivatives exist,*

$$\phi_{X,Y}(x, y) = \frac{\partial^2 \Phi_{X,Y}}{\partial x \partial y}(x, y) \,, \tag{B.60}$$

and then can be used to calculate the joint distribution;

$$\Phi_{X,Y}(x, y) = \int_{-\infty}^{x} d\xi \int_{-\infty}^{y} d\eta \, \phi_{X,Y}(\xi, \eta) \,. \tag{B.61}$$

3. ***Two jointly distributed mixed continuous and discrete random variables**, X and Y, have the **joint density** if only the x-partial derivative exists,*

$$\phi_{X,Y}(x, y_j) = \frac{\partial \Phi_{X,Y}}{\partial x}(x, y_j) \,, \tag{B.62}$$

*which is a **hybrid density-distribution** rather than a strict joint density, but then it can be used to calculate the joint distribution,*

$$\Phi_{X,Y}(x, y_j) = \int_{-\infty}^{x} d\xi \, \phi_{X,Y}(\xi, y_j) \,, \tag{B.63}$$

for some discrete value y_j.

- *The **marginal distributions** in one of two random variables X and Y are defined by summing or integrating over the other random variable:*

  1. ***Two jointly distributed discrete random variables**, X and Y, have the **marginal distributions***

$$\pi_X(x_i) = \sum_{j=1}^{\infty} \pi_{X,Y}(x_i, y_j)\,, \tag{B.64a}$$

$$\pi_Y(y_j) = \sum_{i=1}^{\infty} \pi_{X,Y}(x_i, y_j)\,; \tag{B.64b}$$

  2. ***Two jointly distributed continuous random variables**, X and Y, have the **marginal distributions***

$$\Phi_X(x) = \lim_{y\to+\infty} \Phi_{X,Y}(x,y) = \int_{-\infty}^{x} d\xi \int_{-\infty}^{+\infty} d\eta\, \phi_{X,Y}(\xi,\eta)\,, \tag{B.65a}$$

$$\Phi_Y(y) = \lim_{x\to+\infty} \Phi_{X,Y}(x,y) = \int_{-\infty}^{y} d\eta \int_{-\infty}^{+\infty} d\xi\, \phi_{X,Y}(\xi,\eta)\,, \tag{B.65b}$$

  provided the limits exist;

  3. ***Two jointly distributed mixed continuous and discrete random variables**, X and Y, have the **marginal distributions***

$$\Phi_X(x) = \int_{-\infty}^{x} d\xi \sum_{j=1}^{\infty} \phi_{X,Y}(\xi, y_j)\,, \tag{B.66a}$$

$$\pi_Y(y_j) = \int_{-\infty}^{+\infty} d\xi\, \phi_{X,Y}(\xi, y_j)\,, \tag{B.66b}$$

  provided the limit exists.

- *The **marginal densities** of two random variables, X and Y, are defined as*

  1. ***Two jointly distributed discrete random variables**, X and Y, do not have **marginal densities** in the usual way, but for an applied formulation, the generalized functions can be used (see Section B.12 on page B53);*

  2. ***Two jointly distributed continuous random variables**, X and Y, have the **marginal densities**,*

$$\phi_X(x) = \int_{-\infty}^{+\infty} d\eta\, \phi_{X,Y}(x,\eta)\,, \tag{B.67a}$$

$$\phi_Y(y) = \int_{-\infty}^{+\infty} d\xi\, \phi_{X,Y}(\xi,y)\,; \tag{B.67b}$$

3. ***Two jointly distributed mixed continuous and discrete random variables***, X *and* Y, *has the* ***marginal density*** *for the continuous random variable* X,

$$\phi_X(x) = \sum_{j=1}^{\infty} \phi_{X,Y}(x, y_j)\,, \tag{B.68}$$

and the marginal distribution $\pi_Y(y_j)$ *is given in (B.66b).*

- *The* ***expectation function*** $f(X,Y)$ ***of joint random variables***, X *and* Y, *is defined as:*

  1. ***Two jointly distributed discrete random variables***, X *and* Y, *have the* ***joint expectation*** *of* $f(X,Y)$, *providing the sums or integrals exist,*

$$\mathrm{E}_{X,Y}[f(X,Y)] = \sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty} f(x_i, y_j)\pi_{X,Y}(x_i, y_j)\,; \tag{B.69}$$

  2. ***Two jointly distributed continuous random variables***, X *and* Y, *have the* ***joint expectation*** *of* $f(X,Y)$,

$$\mathrm{E}_{X,Y}[f(X,Y)] = \int_{-\infty}^{+\infty} d\xi \int_{-\infty}^{+\infty} d\eta\, f(\xi,\eta)\phi_{X,Y}(\xi,\eta)\,; \tag{B.70}$$

  3. ***Two jointly distributed mixed continuous and discrete random variables***, X *and* Y, *have the* ***joint expectation***,

$$\mathrm{E}_{X,Y}[f(X,Y)] = \int_{-\infty}^{+\infty} d\eta \sum_{j=1}^{\infty} f(\xi, y_j)\phi_{X,Y}(\xi, y_j)\,, \tag{B.71}$$

  where $\phi_{X,Y}(x, y_j)$ *is the hybrid density-distribution given by (B.62).*

- *The* ***covariance of two jointly distributed random variables***, X *and* Y, *for all three cases, is defined as*

$$\mathrm{Cov}[X,Y] \equiv \mathrm{E}_{X,Y}[(X - \mathrm{E}_X[X])(Y - \mathrm{E}_Y[Y])]\,, \tag{B.72}$$

provided the expectations exist. Hence,

$$\mathrm{Cov}[X,Y] = \mathrm{E}_{X,Y}[X \cdot Y] - \mathrm{E}_X[X] \cdot \mathrm{E}_Y[Y]. \tag{B.73}$$

- *The* ***variance of a sum or difference of two random variables***, X *and* Y,

$$\mathrm{Var}[X \pm Y] = \mathrm{Var}_X[X] \pm 2\mathrm{Cov}[X,Y] + \mathrm{Var}_Y[Y]\,, \tag{B.74}$$

since upon expansion using (B.72) and the definition of variance twice,

$$\begin{aligned} \mathrm{Var}[X \pm Y] = &\quad \mathrm{E}[(X - \mathrm{E}[X] \pm (Y - \mathrm{E}[Y]))^2] \\ = &\quad \mathrm{Var}_X[X] \pm 2\mathrm{Cov}_{X,Y}[X,Y] + \mathrm{Var}_Y[Y]\,. \end{aligned}$$

Remarks B.34.

- *The subscript on the expectation symbol is often omitted, but can be used in multivariate expectation to precisely specify which variable or variables are the arguments of the expectation operator and avoid confusion.*
- *The integral notations are equivalent,*

$$\int_{x_1}^{x_2} dx \int_{y_1}^{y_2} dy f(x,y) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x,y) dy dx \,,$$

the former, having the element of integration follow the integration sign, makes it easy to see the order of integration and which limits of integration go with what elements of integration.

Definitions B.35. ***Independently Distributed Random Variables****:*

- *The* ***joint distribution of two independent random variables****, X and Y, is the product of the marginal distributions:*
  1. ***Two discrete random variables****, X and Y, are* ***independent*** *if their joint distribution is*

$$\pi_{X,Y}(x_i, y_j) = \pi_X(x_i) \cdot \pi_Y(y_j) \,; \tag{B.75}$$

  2. ***Two continuous random variables****, X and Y, are* ***independent*** *if their joint distribution is*

$$\Phi_{X,Y}(x,y) = \Phi_X(x) \cdot \Phi_Y(y) \,; \tag{B.76}$$

  3. ***Two mixed continuous discrete random variables****, X and Y, are* ***independent*** *if their joint distribution is*

$$\Phi_{X,Y}(x, y_j) = \Phi_X(x) \cdot \pi_Y(y_j) \,. \tag{B.77}$$

- *The* ***joint density of two independent random variables****, X and Y, is the product of the marginal densities:*
  1. ***Two discrete random variables****, X and Y, do not have a* ***joint density*** *in the usual way;*
  2. ***Two continuous random variables****, X and Y, are* ***independent*** *if their joint distribution is*

$$\phi_{X,Y}(x,y) = \phi_X(x) \cdot \phi_Y(y) \,; \tag{B.78}$$

  3. ***Two mixed continuous and discrete random variables****, X and Y, are* ***independent*** *if their hybrid density-distribution is*

$$\phi_{X,Y}(x, y_j) = \phi_X(x) \cdot \pi_Y(y_j) \,; \tag{B.79}$$

assuming densities exist where relevant.

- *The **joint expectation of the product** $f(X) \cdot g(Y)$ in two independent random variables, X and Y, is the **product of the expectations**,*

$$\mathrm{E}_{X,Y}[f(X) \cdot g(Y)] = \mathrm{E}_X[f(X)] \cdot \mathrm{E}_Y[g(Y)] \,, \tag{B.80}$$

covering all three cases.

- *The **covariance of two independent random variables**, X and Y, is **zero**,*

$$\mathrm{Cov}[X, Y] \equiv \mathrm{E}[(X - E[X])(Y - E[Y])] = 0 \,, \tag{B.81}$$

since by the separability of the expectation in (B.80),

$$\mathrm{Cov}[X, Y] = \mathrm{E}_X[(X - E[X])] \cdot \mathrm{E}_Y[(Y - E[Y])] = 0 \cdot 0 = 0.$$

Note that the converse is not true. If $\mathrm{Cov}[X, Y] = 0$, then the random variables are not necessarily independent.

B.3.1 Conditional Distributions and Expectations

Definitions B.36.

- *The **conditional probability** and **conditional distribution** of the random variable X given the random variable Y is defined such that*

  1. ***If X and Y are both discrete random variables**,*

$$\pi_{X|Y}(x_i|y_j) \equiv \mathrm{Prob}[X = x_i \,|\, Y = y_j] = \frac{\mathrm{Prob}[X = x_i, Y = y_j]}{\mathrm{Prob}[Y = y_j]} \,, \tag{B.82}$$

  provided the marginal distribution $\pi_Y(y_j) = \mathrm{Prob}[Y = y_j] \neq 0$ from (B.64).

  2. ***If X and Y are both continuous random variables**,*

$$\Phi_{X|Y}(x|y) \equiv \mathrm{Prob}[X \leq x \,|\, Y = y] = \frac{\int_{-\infty}^{x} d\xi \phi_{X,Y}(\xi, y)}{\phi_Y(y)} \,, \tag{B.83}$$

  provided marginal density $\phi_Y(y) \neq 0$ from (B.67). See Karlin and Taylor [161].

  Remarks B.37.

  - *Since we can write*

$$\mathrm{Prob}[Y \in [y, y + dy]] \stackrel{dy}{=} \phi_Y(y) dy \,,$$

  i.e., in precision-dy, the formula (B.83) can be rewritten in probabilities,

$$\mathrm{Prob}[X \leq x \,|\, Y = y] = \frac{\mathrm{Prob}[X \leq x, Y \in [y, y + dy]]}{\mathrm{Prob}[Y \in [y, y + dy]]} \,,$$

  provided $\mathrm{Prob}[Y \in [y, y + dy]] > 0$.

- *Regarding (B.83), note that if Y is a continuous random variable, then* $\text{Prob}[Y = y] = 0$ *since a single point has no probability mass with*

$$\lim_{\delta \to 0} \int_y^{y+\delta} \phi_Y(\eta) d\eta = 0.$$

- *The reader can confirm the consistency of these conditional probability formulas when X and Y are independent random variables.*

3. ***If X is a continuous and Y is a discrete random variable,***

$$\Phi_{X|Y}(x|y_j) \equiv \text{Prob}[X \leq x \mid Y = y_j] = \frac{\text{Prob}[X \leq x, Y = y_j]}{\text{Prob}[Y = y_j]} \qquad \text{(B.84)}$$
$$= \frac{\int_{-\infty}^{x} d\xi \phi_{X,Y}(\xi, y_j)}{\text{Prob}[Y = y_j]},$$

provided marginal distribution $\pi_Y(y_j) = \text{Prob}[Y = y_j] \neq 0$ from (B.66b), where $\phi_{X,Y}(\xi, y_j)$ is the ***hybrid density-distribution*** *in (B.62).*

- ***Iterated probability*** *uses the definitions of conditional probability in reverse to evaluate* ***joint probability*** *for the random variables X and Y ,*

1. ***If X and Y are both discrete random variables,***

$$\text{Prob}[X = x_i, Y = y_j] = \text{Prob}[X = x_i \mid Y = y_j] \cdot \text{Prob}[Y = y_j], \quad \text{(B.85)}$$

provided the conditional distribution $\text{Prob}[X = x_i \mid Y = y_j]$ *exists.*

2. ***If X and Y are both continuous random variables,***

$$\text{Prob}[X \leq x, Y \in [y, y + dy]] = \int_{-\infty}^{x} d\xi \phi_{X,Y}(\xi, y) dy$$
$$= \text{Prob}[X \leq x \mid Y = y] \cdot \phi_Y(y) dy, \quad \text{(B.86)}$$

provided the conditional distribution $\text{Prob}[X \leq x \mid Y = y]$ *exists, but if not then $\phi_Y(y) = 0$ should cover the case.*

3. ***If X is a continuous and Y is a discrete random variable,***

$$\text{Prob}[X \leq x, Y = y_j] = \text{Prob}[X \leq x \mid Y = y_j] \cdot \text{Prob}[Y = y_j], \quad \text{(B.87)}$$

provided marginal distribution $\pi_Y(y_j) = \text{Prob}[Y = y_j] \neq 0$ from (B.66b), where $\phi_{X,Y}(\xi, y_j)$ is the hybrid density-distribution in (B.62).

Remark B.38. *These forms are convenient for decomposing joint probability calculations into simpler parts.*

- *The **conditional density** is*

$$\phi_{X|Y}(x|y) = \frac{\partial \Phi_{X|Y}(x|y)}{\partial x} , \tag{B.88}$$

provided X is a continuous random variable and Y is either continuous or discrete.

- *The **conditional expectation** of X given $Y = y$ is defined as*

$$\mathrm{E}_X[X|Y = y] = \int_{-\infty}^{+\infty} x\phi_{X|Y}(x|y)dx \tag{B.89}$$

provided X is a continuous random variable and Y is either continuous or discrete; else

$$\mathrm{E}_X[X|Y = y_j] = \sum_{i=1}^{\infty} x_i \pi_{X|Y}(x_i|y_j) \tag{B.90}$$

when both X and Y are discrete random variables with a similar form for $\mathrm{E}_X[X|Y = y]$ if X is discrete but Y is continuous.

- *Similarly, the **expectation for a function** $f(X, Y)$ given $Y = y$ is*

$$\mathrm{E}_X[f(X, Y)|Y = y] = \int_{-\infty}^{+\infty} f(x, y)\phi_{X|Y}(x|y)dx .$$

provided X is a continuous random variable and Y is either continuous or discrete; else

$$\mathrm{E}_X[f(X, Y)|Y = y_j] = \sum_{i=1}^{\infty} f(x_i, y_j)\pi_{X|Y}(x_i|y_j)$$

when both X and Y are discrete random variables.

Properties B.39. ***Conditional Expectations**:*

- $\mathrm{E}[f(X)|X] = f(X)$ *for some function f.*
- $\mathrm{E}_Y[\mathrm{E}_{X|Y}[X|Y]] = \mathrm{E}_{X,Y}[X]$, *but* $\mathrm{E}_Y[\mathrm{E}_{X|Y}[X|Y]] = E_X[X]$ *if X and Y are independent random variables.*
- $\mathrm{E}[c_1X_1 + c_2X_2|Y] = c_1\mathrm{E}[X_1|Y] + c_2\mathrm{E}[X_2|Y]$, *provided the conditional expectations exist for random variables Y and X_i, and constants c_i, for $i = 1 : 2$, i.e., the conditional expectation is a linear operation.*
- *If X and Y are random variables, then the **iterated expectation** is*

$$\mathrm{E}_{X,Y}[f(X, Y)] = \mathrm{E}_Y\left[\mathrm{E}_X[f(X, Y)|Y]\right] , \tag{B.91}$$

provided the expectations exist, i.e., that $f(x,y)$ is sufficiently integrable with respect to any density. This is also a general form of the law of total probability given the next section.

Proof. *In the case that X and Y are both continuous random variables, the justification is built upon the basic definition of the conditional distribution in (B.83) which leads to the conditional density according to (B.88) upon differentiation,*

$$\phi_{X|Y}(x|y) = \phi_{X,Y}(x,y)/\phi_Y(y)$$

assuming $\phi_Y(y) > 0$. Further, $\phi_Y(y) > 0$ will be assumed on $-R \leq y \leq R$ for some $R > 0$, since $\phi_Y(y) \to 0^+$ as $y \to +\infty$ for conservation of probability through integrability at infinity. For convenience, the limit as $R \to +\infty$ will be ignored in the following formally justifying chain of equations:

$$\begin{aligned}
\mathrm{E}_{X,Y}[f(X,Y)] &= \int_{-\infty}^{+\infty} dy \int_{-\infty}^{+\infty} dx \phi_{X,Y}(x,y) f(x,y) \\
&= \int_{-\infty}^{+\infty} dy \int_{-\infty}^{+\infty} dx \left(\phi_{X|Y}(x|y)\phi_Y(y)\right) f(x,y) \\
&= \int_{-\infty}^{+\infty} dy \phi_Y(y) \int_{-\infty}^{+\infty} dx \phi_{X|Y}(x|y) f(x,y) \\
&= \mathrm{E}_Y\left[\mathrm{E}_X[f(X,Y)|Y]\right] .
\end{aligned}$$

The other random variable cases are similar with sums where discrete random variables are concerned. □

- *If X and Y are independent, then $\mathrm{E}[X|Y] = \mathrm{E}[X]$ and in general*

$$\mathrm{E}[f(X)g(Y)|Y] = \mathrm{E}[f(X)]g(Y),$$

provided the expectations exist.

See Mikosch [209] for more conditional expectation properties in a more abstract setting.

B.3.2 Law of Total Probability

Properties B.40. ***Law of Total Probability****:*

- *When X is a* ***discrete random variable*** *and given a countable set of* ***mutually independent discrete random variables****, $\{Y_1, Y_2, \ldots, Y_i, \ldots\}$, and the conditional probabilities $\mathrm{Prob}[X|Y_i]$ for $i = 1, 2, \ldots$, then the* ***law of total probability*** *(see Taylor and Karlin [265]) in this completely discrete case is*

$$\mathrm{Prob}[X] = \sum_{i=1}^{\infty} \mathrm{Prob}[X|Y_i]\mathrm{Prob}[Y_i] \,, \tag{B.92}$$

i.e., an extension of the law of additive probabilities for disjoint events.

- *When X is a **continuous random variable**, the corresponding law of total probability for the **probability distribution** $\Phi_X(x)$ is*

$$\Phi_X(x) = \sum_{i=1}^{\infty} \Phi_{X|Y}(x|Y_i)\text{Prob}[Y_i] \,. \tag{B.93}$$

- *Providing the density exists in the continuous random variable case, the corresponding law of total probability for the **probability density** of $\phi_X(x)$ is*

$$\phi_X(x) = \sum_{i=1}^{\infty} \phi_{X|Y}(x|Y_i)\text{Prob}[Y_i] \,. \tag{B.94}$$

- *Finally, the **expectation** corresponding to the law of total probability is*

$$\text{E}[f(X)] = \sum_{i=1}^{\infty} \text{E}_X[f(X)|Y_i]\text{Prob}[Y_i] \,. \tag{B.95}$$

*for either discrete or continuous X case and assuming the expectations of $f(X)$ exist. This is a special case of the **iterated expectations** given previously in (B.91).*

Example B.41. *An interesting financial example of (B.95) derived from [265] is the statistics for the daily stock price return observed on a transaction by transaction basis. Let the transaction price return be $\xi_i = \Delta S_i = S_{i+1} - S_i$, where S_i is the price of the ith transaction, with S_0 the initial price such as that from the previous day's closing. Suppose the returns are independent identically distributed (I.I.D.) random variables with common mean $\text{E}_\xi[\xi_i] = \mu$ and variance $\text{Var}_\xi[\xi_i] = \sigma^2$. Assume the current total daily stock return after N transactions is*

$$X = \sum_{i=0}^{N} \xi_i \,,$$

where N is Poisson distributed, i.e., N is a counting process such that $\text{Prob}[N = n] = p_n(\Lambda)$ with Λ being the Poisson parameter in (B.50), so $\text{E}_N[N] = \Lambda = \text{Var}_N[N]$. Starting from the law of total probability, the expectation of the daily return is decomposed as

$$\begin{aligned}
\text{E}_X[X] &= \sum_{n=0}^{\infty} \text{E}_{X|N}[X|N=n]p_n(\Lambda) = \sum_{n=0}^{\infty} \text{E}_{\xi|N}\left[\sum_{i=0}^{N} \xi_i \,\middle|\, N=n\right] p_n(\Lambda) \\
&= \sum_{n=0}^{\infty} \text{E}_\xi\left[\sum_{i=0}^{n} \xi_i\right] p_n(\Lambda) = \sum_{n=0}^{\infty}\sum_{i=0}^{n} \text{E}_\xi[\xi_i]p_n(\Lambda) \\
&= \sum_{n=0}^{\infty}\sum_{i=0}^{n} \mu p_n(\Lambda) = \mu \sum_{n=0}^{\infty} n p_n(\Lambda) = \mu\Lambda \,,
\end{aligned}$$

where the independence and identically distributed properties of the ξ_i random variables, as well as the mean properties of N, have been used.

The variance of X is more complicated but follows from similar techniques, except that terms are collected by ***completing the square*** *in the ith return deviation from the mean $(\xi_i - \mu)$ with several applications of the independence assumption:*

$$\begin{aligned}
\mathrm{Var}_X[X] = \mathrm{E}_X[(X-\Lambda\mu)^2] &= \sum_{n=0}^{\infty} \mathrm{E}_{\xi|N}\left[\left(\sum_{i=0}^{N}\xi_i - \Lambda\mu\right)^2 \middle| N=n\right] p_n(\Lambda) \\
&= \sum_{n=0}^{\infty} \mathrm{E}_{\xi}\left[\left(\sum_{i=0}^{n}(\xi_i-\mu) + (n-\Lambda)\mu\right)^2\right] p_n(\Lambda) \\
&= \sum_{n=0}^{\infty} \mathrm{E}_{\xi}\left[\sum_{i=0}^{n}\sum_{j=0}^{n}(\xi_i-\mu)(\xi_j-\mu) + 2(n-\Lambda)\mu\sum_{i=0}^{n}(\xi_i-\mu) + (n-\Lambda)^2\mu^2\right] p_n(\Lambda) \\
&= \sum_{n=0}^{\infty} \mathrm{E}_{\xi}\left[\sum_{i=0}^{n}(\xi_i-\mu)^2 + \sum_{i=0}^{n}\sum_{j\neq i}^{n}(\xi_i-\mu)(\xi_j-\mu) + (n-\Lambda)^2\mu^2\right] p_n(\Lambda) \\
&= \sum_{n=0}^{\infty}\left[\sum_{i=0}^{n}\mathrm{E}_{\xi}[(\xi_i-\mu)^2] + (n-\Lambda)^2\mu^2\right] p_n(\Lambda) = \sum_{n=0}^{\infty}[n\sigma^2 + (n-\Lambda)^2\mu^2]p_n(\Lambda) \\
&= \Lambda\sigma^2 + \Lambda\mu^2 = \Lambda(\sigma^2+\mu^2)\,,
\end{aligned}$$

such that the ith return variance is augmented by the mean squared.

B.4 Probability Distribution of a Sum: Convolutions

Combinations of random variables play an important role in the analysis of stochastic processes, especially in the sum of two stochastic processes. Consider the following result:

Theorem B.42. ***Convolution for Sums of Random Variables*** *If X and Y are independent random variables with densities $\phi_X(x)$ and $\phi_Y(y)$, respectively, then the* ***distribution of the sum*** *is*

$$\Phi_{X+Y}(z) \equiv \mathrm{Prob}[X+Y \leq z] = \int_{-\infty}^{+\infty} \Phi_Y(z-x)\phi_X(x)dx\,, \tag{B.96}$$

provided the integral exists, where

$$\Phi_Y(y) = \int_{-\infty}^{y} \phi_Y(\eta)d\eta\,.$$

Proof. By the independence of the variables X and Y, the joint density is separable, $\phi_{X+Y}(x,y) = \phi_X(x)\phi_Y(y)$. Thus, using the properties of the Heaviside step

function,

$$H(x) = \left\{ \begin{array}{ll} 0, & x < 0 \\ 1, & x \geq 0 \end{array} \right\} , \tag{B.97}$$

then

$$\begin{aligned} \mathrm{Prob}[X + Y \leq z] &= \mathrm{E}_{X+Y}[H(z - X - Y)] \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} H(z - x - y)\phi_X(x)\phi_Y(y)dydx \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} \phi_Y(y)dy\phi_X(x)dx \\ &= \int_{-\infty}^{+\infty} \Phi_Y(z - x)\phi_X(x)dx \\ &= \mathrm{E}_X[\Phi_Y(z - X)] , \end{aligned}$$

where iterated integrals have been freely interchanged by the theorem of Fubini which asserts that if an integral exists as a 2-dimensional integral then the two iterative integrals can be interchanged, i.e., the order of integration does not matter. **Fubini's theorem** is often used in probability theory [84, 168]. □

Since it has been assumed that the densities exist, then differentiation of the sides of the equation in (B.96), but under the integral sign for those on the right, yields the formula for the probability density of a sum:

Corollary B.43.

$$\phi_{X+Y}(z) = \int_{-\infty}^{+\infty} \phi_Y(z - x)\phi_X(x)dx , \tag{B.98}$$

The particular functional product forms of (B.96,B.98) are called convolutions [84]:

Definition B.44. *Let the* ***convolution*** *of a distribution or density $f(y)$ and a density $\phi(x)$ be*

$$(f * \phi)(z) \equiv \int_{-\infty}^{+\infty} f(z - x)\phi(x)dx , \tag{B.99}$$

provided the integral exists. Consequently, we have the following properties, including the reformulation of the above sum rules:

Properties B.45. Convolutions*:*

- *The* ***convolution of densities is symmetric*** $(f * \phi)(z) = (\phi * f)(z)$ *, upon change of variables in the integrand.*

- $\phi_{X+Y}(z) = (\phi_Y * \phi_X)(z) = (\phi_X * \phi_Y)(z)$.
- $\Phi_{X+Y}(z) = (\Phi_X * \phi_Y)(z) = (\Phi_Y * \phi_X)(z)$.
- *The form for n mutually independent random variables, all with given densities, is*

$$\begin{aligned}\phi_{X_1+X_2+\cdots+X_n}(z) &= (\phi_{X_1} * \phi_{X_2} * \cdots * \phi_{X_n})(z) \qquad \text{(B.100)}\\ &= \left\{ \begin{array}{l} ((\cdots((\phi_{X_1} * \phi_{X_2}) * \phi_{X_3}) \cdots * \phi_{X_{n-1}}) * \phi_{X_n})(z) \\ (\phi_{X_1} * (\phi_{X_2} * (\phi_{X_3} * \cdots (\phi_{X_{n-1}} * \phi_{X_n}) \cdots)))(z) \end{array} \right\},\end{aligned}$$

the latter forms depending on whether the convolution expansion is from the right or from the left, respectively.

Remark B.46. *The particular form depends on which particular inductive definition is used, i.e., the right and left convolution expansion forms, respectively, are*

$$\phi_{\sum_{i=1}^{n+1} X_i}(z) = \left\{ \begin{array}{l} \left(\phi_{\sum_{i=1}^{n} X_i} * \phi_{X_{n+1}}\right)(z) \\ \left(\phi_{X_1} * \phi_{\sum_{i=2}^{n+1} X_i}\right)(z) \end{array} \right\},$$

as can be shown by mathematical induction.

Lemma B.47. ***Convolution of Normal Densities is Normal****:*
If X and Y are normally distributed random variables, with probability densities $\phi_X(x) = \phi_n(x; \mu_x, \sigma_x^2)$ and $\phi_Y(y) = \phi_n(y; \mu_y, \sigma_y^2)$, respectively, then, letting $Z = X + Y$,

$$\begin{aligned}\phi_Z(z) &= (\phi_X * \phi_Y)(z) \qquad \text{(B.101)}\\ &= \int_{-\infty}^{+\infty} \phi_X(z-y)\phi_Y(y)dy \qquad \text{(B.102)}\\ &= \phi_n(z; \mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)\,. \qquad \text{(B.103)}\end{aligned}$$

***Maple Proof*:**

```
> phi:=(x,m,s)->exp(-(x-m)^2/(2*s^2))/sqrt(2*pi*s^2);
```

$$\phi := (x, m, s) \rightarrow \frac{e^{(-1/2\,\frac{(x-m)^2}{s^2})}}{\sqrt{2\,\pi\,s^2}}$$

```
> interface(showassumed=0); assume(sx>0); assume(sy>0);
> phi_Z:=simplify(int(phi(z-y,mx,sx)*phi(y,my,sy),
```

```
> y=-infinity..infinity));
```

$$phi\_Z := \frac{1}{2}\,\frac{e^{\left(-\frac{(z-mx-my)^2}{2\,(sy^2+sx^2)}\right)}\sqrt{2}\sqrt{\pi}}{\pi\sqrt{sy^2+sx^2}}$$

For more general results see the Exercises (16, 17, 18).

B.5 Characteristic Functions

Often it convenient to transform distributions or densities so that moments can be generated more systematically leading to a class of generating functions. Here, the emphasis will be on one class that is more useful for both positive and negative random variables, called *characteristic functions*.

Definition B.48. *The characteristic function of a random variable X is defined in general as*

$$\mathcal{C}_X(u) \equiv \mathrm{E}\left[e^{iuX}\right], \tag{B.104}$$

where $i = \sqrt{-1}$ is the imaginary unit constant, u is the characteristic function argument, assumed real here, the complex exponential is

$$e^{iux} = \cos(ux) + i\sin(ux)$$

by Euler's formula with complex conjugate $z^* = (x+iy)^* \equiv x - iy$ so

$$(\exp(iux))* = \exp(-iux)$$

and modulus (absolute value) $|z| \equiv \sqrt{(}x^2+y^2)$ so

$$\left|e^{iux}\right| = \sqrt{cos^2(ux) + sin^2(ux)} = 1$$

according to Pythagorus' theorem (summarizing almost all of the complex algebra that will be needed here). Only three main forms for $\mathcal{C}_X(u)$ are listed here:

- *if X is a continuous random variable with proper probability distribution function $\Phi_X(x)$ then*

$$\mathcal{C}_X(u) = \int_{-\infty}^{\infty} e^{iux} d\Phi_X(x), \tag{B.105}$$

  which is called a Fourier-Stieltjes transform;

- *if X is a continuous random variable and there exists a density corresponding to $\Phi_X(x)$, then*

$$\mathcal{C}_X(u) = \int_{-\infty}^{\infty} e^{iux} \phi_X(x) dx, \tag{B.106}$$

  which is just an ordinary Fourier transform;

- *if X is a discrete random variables with distribution function $\pi_k = \text{Prob}[X = x_k]$ for all non-negative integers k, then*

$$\mathcal{C}_X(u) = \sum_{k=0}^{\infty} \pi_k e^{iux_k}, \tag{B.107}$$

which is called a Fourier exponential series.

Properties B.49. ***Characteristic Functions**:*

- ***Moment Properties**:*
  - *$\mathcal{C}_X(0) = 1$ by conservation of probability;*
  - *$\mathcal{C}'_X(0) = \text{E}_X[X]$ by differentiation of integrand;*
  - *By induction for $k = 0, 1, 2, \cdots$,*

$$\frac{d^k \mathcal{C}_X}{du^k}(0) = i^k \text{E}_X\left[X^k\right].$$

- ***Relationship to Standard Generating Function**:*

$$G_X(z) \equiv \text{E}\left[z^X\right], \tag{B.108}$$

so letting $z^x = e^{iux}$, then $z = e^{iu}$, $u = -i\ln(z)$, $G_X(z) = \mathcal{C}_X(-i\ln(z))$ and $\mathcal{C}_X(u) = G_X(e^{iu})$.

- ***Complex Properties**: By Euler's formula, the resolution into real and imaginary parts:*

$$\mathcal{C}_X(u) = C_X(u) + iS_X(u),$$

where the real part is the cosine transform

$$C_X(u) = \int_{-\infty}^{\infty} \cos(ux)\phi_X(x)dx$$

and the imaginary part is the sine transform

$$S_X(u) = \int_{-\infty}^{\infty} \sin(ux)\phi_X(x)dx,$$

so the complex conjugate is

$$\mathcal{C}^*_X(u) = C_X(u) - iS_X(u).$$

- ***Reality and Symmetric Densities**: The characteristic function $\mathcal{C}_X(u)$ is real if and only if the corresponding probability density is symmetric, i.e. $\phi_X(-x) = \phi_X(x)$. Note that $\mathcal{C}_X(u)$ is real if the imaginary part $S_X(u)$ is zero*

and $\mathcal{C}_X(-u) = \mathcal{C}_X^*(u) = C_X(u) - iS_X(u)$ *(*$\exp(-iux) = \cos(ux) - i\sin(ux)$*), so*

$$\begin{aligned} iS_X(u) &= 0.5(\mathcal{C}_X(u) - \mathcal{C}_X(-u)) = 0.5\int_{-\infty}^{\infty}\left(e^{iux} - e^{-iux}\right)\phi_X(x)dx \\ &= 0.5\int_{-\infty}^{\infty} e^{iux}\left(\phi_X(x) - \phi_X(-x)\right)dx, \end{aligned}$$

then $\phi_X(x)$ *symmetric implies* $S_X(u) = 0$ *and* $S_X(u) = 0$ *implies* $\phi_X(x)$ *symmetric.*

- ***Upper Bound****:* $|\mathcal{C}_X(u)| \leq 1$*, since by Euler's formula and trigonometric identities*

$$\begin{aligned} |\mathcal{C}_X(u)|^2 &= \left(\int_{-\infty}^{\infty}\cos(ux)\phi_X(x)dx\right)^2 + \left(\int_{-\infty}^{\infty}\sin(ux)\phi_X(x)dx\right)^2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\cos(ux)\cos(uy) + \sin(ux)\sin(uy))\phi_X(x)\phi_X(y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cos(u(x-y))\phi_X(x)\phi_X(y)dxdy \\ &\leq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi_X(x)\phi_X(y)dxdy = 1. \end{aligned}$$

- ***Sums of Random Variables and Convolutions****: Let* $\{X_k; k = 1 : N\}$ *be a set of independent random variables, then* $\mathcal{C}_{X_1+X_2}(u) = \mathcal{C}_{X_1}(u)\cdot\mathcal{C}_{X_1}(u)$ *since by the convolution property (B.98)*

$$\begin{aligned} \mathcal{C}_{X_1+X_2}(u) &= \int_{-\infty}^{\infty} e^{iux}\phi_{X_1+X_2}(x)dx = \int_{-\infty}^{\infty} e^{iux}\left(\phi_{X_1} * \phi_{X_2}\right)(x)dx \\ &= \int_{-\infty}^{\infty} e^{iux}\int_{-\infty}^{\infty}\phi_{X_2}(x-y)\phi_{X_1}(y)dydx \\ &= \int_{-\infty}^{\infty} e^{iuy}\phi_{X_1}(y)\int_{-\infty}^{\infty} e^{iu(x-y)}\phi_{X_2}(x-y)dxdy \\ &= \mathcal{C}_{X_1}(u)\cdot\mathcal{C}_{X_1}(u), \end{aligned}$$

assuming integral interchange is permitted. Further, for a set of of N *independent random variables,*

$$\mathcal{C}_{\sum_{k=1}^N X_k}(u) = \prod_{k=1}^{N}\mathcal{C}_{X_k}(u).$$

- ***Uniqueness****: The characteristic function* $\mathcal{C}_X(u)$ *is uniquely related to its corresponding distribution* $\Phi_X(x)$ *and vice versa (see Feller [84] for justification and more information on characteristic and other generating functions, as well as the inverse Fourier transform that is beyond the simple complex variables that is assumed here).*

Examples B.50. ***Characteristic Functions for Common Distributions:***

- ***Normal Distribution:***

$$\mathcal{C}_n(u;\mu,\sigma^2) = \int_{-\infty}^{\infty} e^{iux}\phi_n(x;\mu,\sigma^2)dx = e^{-0.5\sigma^2u^2+i\mu u}.$$

- ***Exponential Distribution*** ($\mu > 0$)***:***

$$\mathcal{C}_e(u;\mu) = \int_0^{\infty} e^{iux}\phi_e(x;\mu)dx = \frac{1}{1-i\mu u} = \frac{1+i\mu u}{1+\mu^2u^2}.$$

- ***Uniform Distribution*** ($a < b$)***:***

$$\mathcal{C}_u(u;a,b) = \frac{1}{b-a}\int_a^b e^{iux}dx = \frac{e^{iub}-e^{iua}}{i(b-a)u}.$$

- ***Double Exponential (Laplace) Distribution*** ($\mu > 0$)***:***

$$\mathcal{C}_{de}(u;a,\mu) = \frac{1}{2\mu}\int_0^{\infty} e^{iux}e^{-|x-a|/\mu}dx = \frac{e^{iau}}{1+\mu^2u^2}.$$

- ***Poisson Distribution*** ($\Lambda > 0$***,*** $x_k = k$)***:***

$$\mathcal{C}_p(u;\Lambda) = \sum_{k=0}^{\infty} e^{iuk}p_k(\Lambda) = \sum_{k=0}^{\infty} e^{iuk}e^{-\Lambda}\frac{\Lambda^k}{k!} = e^{-\Lambda}\sum_{k=0}^{\infty}\frac{\left(e^{iu}\Lambda\right)^k}{k!} = e^{\Lambda\left(e^{iu}-1\right)}.$$

Characteristic functions are also used to define Lévy processes, which are basically a generalization of jump-diffusion processes to include processes with infinite jump rates. Thus, characteristic functions are essential for including such singular behavior. For references on Lévy processes see the cited sources on Lévy processes or jump-diffusion references that emphasize Lévy processes [12, 59, 223].

Another application is to financial option pricing for jump-diffusions with stochastic volatility (i.e., stochastic variance) where the characteristic function formulation and its inverse Fourier transform offer certain advantages for computation (see Carr et al. [47] or Yan and Hanson [287]).

B.6 Sample Mean and Variance: Sums of IID Random Variables

Just as there is no such thing as a truly random variable in practice, although the theory of random variables is very useful, there is no such thing as a continuously sampled random variable in practice. Typically, we sample discretely from a theoretical continuous distribution and assume that the samples are independently sampled.

Definition B.51. ***Independent, Identically Distributed Random Variables (I.I.D. or IID):***

*A set of n random variables $\{X_k|k = 1 : n\}$ is **independent, identically distributed** (I.I.D. or IID) if the X_k have the same distribution, i.e.,*

$$\Phi_{X_k}(x) = \Phi_{X_j}(x) \,,$$

for all $k, j = 1 : n$ and X_k is independent of X_j when $k \neq j$, i.e.,

$$\mathrm{Cov}[X_k, X_j] = \mathrm{Var}[X_j]\delta_{k,j} \,.$$

Definition B.52. ***Sample Mean and Variance:***
*Let $\{X_k|k = 1 : n\}$ be a sample of n random variables, then the **sample mean** is defined as*

$$m_n = \frac{1}{n}\sum_{k=1}^{n} X_k \,, \tag{B.109}$$

*and the **sample variance** or population variance is*

$$s_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - m_n)^2 \,, \tag{B.110}$$

*but the **unbiased estimate** of the **sample variance** is*

$$\widehat{s}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - m_n)^2 \,. \tag{B.111}$$

*An estimate $\widehat{Y}$ of a quantity y is called an **unbiased estimate** if*

$$\mathrm{E}\left[\widehat{Y}\right] = y.$$

Theorem B.53. ***IID Sample Mean and Variance:***
Let $\{X_k|k = 1 : n\}$ be a set of IID random variables, such that $\mathrm{E}[X_k] = \mu$ and $\mathrm{Var}[X_k] = \sigma^2$ for all k, then

$$\mathrm{E}[m_n] = \mu \,, \tag{B.112}$$

$$\mathrm{E}[s_n^2] = \sigma^2 \,, \tag{B.113}$$

$$\mathrm{E}[\widehat{s}_n^2] = \frac{n}{n-1}\sigma^2 \,, \tag{B.114}$$

$$\mathrm{Var}[m_n] = \frac{1}{n}\sigma^2 \,. \tag{B.115}$$

Remarks B.54.

- *These sample moments and more are left as Exercises (13, 15, 14). The first is trivial, but the other two rely heavily on the independence property so it is very helpful to collect all terms as deviations from the mean forms like* $(X_k - \mu)$. *Also, split up multiple sums into a single sum for equal indices (say* $j = k$*) and the product of an outer sum by an inner sum when the inner index in not equal to the outer index (say* $j \neq k$*) . Note that for large* n*, the difference between the regular and unbiased estimates of the variance will be small.*

- *Since* m_n *is a sum of random variables, then its distribution will be a nested convolution of the common distribution of the* X_k *variates. Convolutions are defined earlier in (B.96) of Section B.4.*

- *Later, the relevant limit theorems will be discussed. The Law of Large Numbers (B.116) says that the sample mean will approach the distribution mean and the Central Limit Theorem B.57, discussed later, says that the sample distribution will approach the normal limiting distribution for large sample sizes.*

- *For properties of powers of partial sums of zero-mean IID random variables* *see Lemma 5.15 on page 149.*

B.7 Law of Large Numbers

When applying probability to real applications, the user may need to compare the statistical properties of the practical sample with the ideal concepts of probability theory. For instance, when comparing the sample mean to an ideal distribution mean, some justification comes partly from the law of large numbers, a weak and a strong form are given here suitable for this appendix of preliminaries (see also Feller [83]or Karlin and Taylor [161]) .

B.7.1 Weak Law of Large Numbers (WLLN)

Theorem B.55. ***Law of Large Numbers*** *(weak form):*
Let $\{X_1, X_2, \dots X_i, \dots\}$ *be a sequence of independent identically distributed random variables (i.e., I.I.D. R.V.s or mutually independent random variables with common distribution* $\Phi(x)$*) with common mean* $\mu = \mathrm{E}[X_i]$ *for all* i*. Let* $S_n = \sum_{i=1}^{n} X_i$ *be a sequence of partial sums such that* S_n *is the sum of* n *of these sample measurements, so that the sample mean is* $m_n = S_n/n$*. Then for every* $\epsilon > 0$*.*

$$\mathrm{Prob}[|m_n - \mu| > \epsilon] \longrightarrow 0 \quad \text{as} \quad n \to +\infty \,. \tag{B.116}$$

Thus, if the sample size is large enough, the sample mean will approximate the distribution mean.

B.7.2 Strong Law of Large Numbers (SLLN)

Theorem B.56. ***Law of Large Numbers*** *(strong form):*
Let $\{X_1, X_2, \dots X_i, \dots\}$ be a sequence of independent identically distributed random variables (i.e., I.I.D. R.V.s or mutually independent random variables with common distribution $\Phi(x)$) with common mean $\mu = \mathrm{E}[X_i]$ for all i. Let $S_n = \sum_{i=1}^{n} X_i$ be a sequence of partial sums such that S_n is the sum of n of these sample measurements, so that the sample mean is $m_n = S_n/n$. Then

$$\begin{gathered} \mathrm{Prob}[\lim_{n\to\infty} m_n = \mu] = 1\,, \\ \text{i.e., } m_n \to \mu \text{ with probability one as } n \to +\infty\,. \end{gathered} \tag{B.117}$$

B.8 Central Limit Theorem

The central limit theorem is much more powerful than the law of large numbers. Again, a simple form is given for I.I.D. R.V.s [83].

Theorem B.57. ***Central Limit Theorem:***
Let $\{X_1, X_2, \dots X_i, \cdots\}$ be a sequence of independent identically distributed random variables (i.e., I.I.D. R.V.s or mutually independent random variables with common distribution $\Phi(x)$) with common mean $\mu = \mathrm{E}[X_i]$ and variance $\sigma^2 = \mathrm{Var}[X_i]$ for all i. Let $S_n = \sum_{i=1}^{n} X_i$ be the sum of n of these sample measurements, so that the sample mean is $m_n = S_n/n$. Then for every fixed ξ,

$$\mathrm{Prob}\left[\frac{m_n - \mu}{\sigma/\sqrt{n}} \le \xi\right] \longrightarrow \Phi_n(\xi; 0, 1)\,, \tag{B.118}$$

as $n \to +\infty$, where $\Phi_n(\xi; 0, 1)$ is the standard normal distribution defined in (B.1.4), when $\mu = 0$ and $\sigma^2 = 1$.

Thus, if the sample size is large enough, the deviation of the sample mean from the distribution mean, scaled by $\sigma/\sqrt{n}$, will be asymptotically normally distributed with mean 0 and variance 1.

For stronger versions of the central limit theorem see the many probability references listed at the end of this appendix.

B.9 Matrix Algebra and Analysis

Many important distributions, stochastic processes and control problems are multivariate, rather than scalar. Here matrix algebra and matrix analysis are summarized. Many of the given properties can be computed symbolically using Maple and *Mathematica*, or numerically using MATLAB.

- **Vector Notation**: $\mathbf{x} = [x_i]_{n\times 1}$, in boldface, denotes an n-vector, where the number x_i is the ith component. Let $\mathbf{y} = [y_i]_{n\times 1}$ be another n-vector. In this book vectors are column vectors, unless transposed. Numbers are also called scalars here.

- **Matrix or Array Notation**: $A = [a_{i,j}]_{n\times n}$ denotes an $n \times n$ square matrix (literally a table) or array, where the number $a_{i,j}$ is an element of the ith row and jth column. Sometimes we say that A is an order n matrix. Nonsquare matrices would be $Q = [q_{i,j}]_{m\times n}$ or $R = [r_{i,j}]_{n\times p}$. Matrix elements may also be functions.

- **Matrix equality**: $B = A$ means that all matrix elements are equal, $b_{i,j} = a_{i,j}$ for $i = 1 : n$ and $j = 1 : n$. The negation of the equality only requires one pair of unequal elements, $b_{k,\ell} \neq a_{k,\ell}$ for some (k, ℓ).

- **Matrix Identity**:

$$I_n \equiv [\delta_{i,j}]_{n\times n} \,, \tag{B.119}$$

  where $\delta_{i,j}$ is the Kronecker defined in (B.54) and has the sum property that $\sum_{j=1}^{n} a_j \delta_{i,j} = a_i$ provided i is in the range of j, $j = 1 : n$.

- **Matrix Transpose**:

$$Q^\top = [q_{j,i}]_{n\times m} \,, \tag{B.120}$$

  i.e., transposing a real matrix is switching rows and columns. If there are complex elements, then the **Hermitian transpose** is used, $Q^H = [q^*_{j,i}]_{n\times m}$ where if $z = x + \widehat{i}y$ is a complex number then the complex conjugate is $z^* = x - \widehat{i}y$ and $\widehat{i} = \sqrt{-1}$ is the imaginary unit such that $\widehat{i}^2 = -1$. Although this book is exclusively about real problems, there are important methods and even real problems that introduce complex numbers into the analysis.

- **Inner or Dot or Scalar Product of two Vectors**:

$$\mathbf{x}^\top \mathbf{y} = \mathbf{x}\bullet\mathbf{y} = \mathbf{x}^\top \mathbf{y} \equiv \sum_{i=1}^{n} x_i y_i \,, \tag{B.121}$$

  provided $\mathbf{y}$ is also an n-vector. If there are complex vector elements or components, then the Hermitian inner product is used,

$$\mathbf{x}^H \mathbf{y} \equiv \sum_{i=1}^{n} x_i^* y_i.$$

- **Matrix Trace**:

$$\text{Trace}[A] \equiv \sum_{i=1}^{n} a_{i,i}. \tag{B.122}$$

- **Matrix-Vector Product**:

$$Q\mathbf{x} \equiv \left[\sum_{j=1} q_{i,j} x_j\right]_{m\times 1} \,, \tag{B.123}$$

  i.e., the ith component is $(Q\mathbf{x})_i = \sum_{j=1} q_{i,j} x_j$.

- **Matrix-Matrix Product**:

$$QR \equiv \left[\sum_{k=1} q_{i,k} r_{k,j}\right]_{m \times p} , \tag{B.124}$$

so for two matrices to be **commensurate** or consistent in multiplication the number of columns of the pre-multiplier Q must be the same as the number of rows of the post-multiplier R.

- **Transpose of a Matrix Product**: $(QR)^\top = R^\top Q^\top$.

- **Matrix Inverse**: For square matrices A, the inverse A^{-1} has the property

$$A^{-1}A = I_n = AA^{-1} \tag{B.125}$$

whenever A^{-1} exists and this property provides a set of algebraic equation for determining the elements of the inverse. See the MATLAB, Maple and *Mathematica* packages.

- **Vector Norm**:

$$||\mathbf{x}||_p \equiv \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \tag{B.126}$$

is the pth norm with the properties that

1. $||\mathbf{x}||_p \geq 0$;
2. $||\mathbf{x}||_p = 0$ if and only if $\mathbf{x} = \mathbf{0}$;
3. $||s\mathbf{x}||_p = |s|||\mathbf{x}||_p$ if s is a scalar;
4. $||\mathbf{x} + \mathbf{y}||_p \leq ||\mathbf{x}||_p + ||\mathbf{y}||_p$, called the **triangular inequality**;
5. $||\mathbf{x}^\top \mathbf{y}||_p \leq ||\mathbf{x}||_p ||\mathbf{y}||_p$, called the **Cauchy inequality**.

Common norms are the

1. 1-norm, $||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i|$;
2. infinity-norm, $||\mathbf{x}||_\infty = \max_{i=1:n}[|x_i|]$;
3. 2-norm, $||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\mathbf{x}^\top \mathbf{x}}$ if $\mathbf{x}$ real, but $||\mathbf{x}||_2 = \sqrt{\mathbf{x}^H \mathbf{x}}$ if complex.

- **Matrix Norm**: Matrix norms are defined on the more basic vector norms,

$$||A||_p \equiv \max_{||\mathbf{x}||_p \neq 0} [||A\mathbf{x}||_p / ||\mathbf{x}||_p] = \max_{||\mathbf{u}||_p \neq 1} [||A\mathbf{u}||_p] , \tag{B.127}$$

and they satisfy properties analogous to the vector norm properties above. Usual values are $p = 1$, 2, or ∞.

- **Matrix Condition Number**:

$$\text{cond}_p[A] \equiv ||A||_p||A^{-1}||_p \tag{B.128}$$

is the pth condition number, bounded below by $\text{cond}_p[A] \geq 1$ and is scale-invariant since $\text{cond}_p[sA] = |s|\ \text{cond}_p[A]$ if s is a non-zero scalar. Implicit in the definition is that the inverse A^{-1} exists.

- **Matrix Determinants**: If A is a square matrix, then the determinant $\text{Det}[A]$ has a scalar value that can be computed by recursion from smaller determinants, expanding by either a row or a column. For instance,

  1. If $n = 1$, then $\text{Det}[a_{1,1}] = a_{1,1}$.
  2. If $n = 2$, then

$$\text{Det}\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} = a_{1,1}\text{Det}[a_{2,2}] - a_{1,2}\text{Det}[a_{2,1}].$$

  3. If $n = 3$, then

$$\begin{aligned}\text{Det}\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} &= a_{1,1}\text{Det}\begin{bmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{bmatrix} \\ &\quad -a_{1,2}\text{Det}\begin{bmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{bmatrix} \\ &\quad +a_{1,3}\text{Det}\begin{bmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{bmatrix}.\end{aligned}$$

  4. And so forth.

  Some useful properties are $\text{Det}[A^\top] = \text{Det}[A]$ since row and column expansions give the same result; the **Cauchy-Binet** formula that

$$\text{Det}[AB] = \text{Det}[A]\text{Det}[B] \tag{B.129}$$

provided A and B are commensurate; $\text{Det}[I_n] = 1$; a corollary is $\text{Det}[A^{-1}] = 1/\text{Det}[A]$ if $A^{-1}A = I_n$.

- **Systems of Linear Equations**:

$$A\mathbf{x} = \mathbf{b} \tag{B.130}$$

where the coefficient matrix A and $\mathbf{b} = [b_i]_{n\times 1}$ are given, and the object is to find the vector $\mathbf{x}$.

  1. In theory, a unique solution exists if $\text{Det}[A] \neq 0$, else if $\text{Det}[A] = 0$ then A is called singular.

2. In numerical practice, a nearly singular A usually has serious problems and the condition number cond $[A]$ due to its scale-invariance is a better measure of difficulties. If cond $[A]$ is of moderate size (not much bigger than O(1) say) then the problem is called **well-conditioned**, but if cond $[A]$ is very large than the problem is called ill-conditioned. In Gaussian elimination with back substitution, row pivoting with row scaling or full pivoting can reduce the conditioning problems and produce more reliable approximate solutions. The MATLAB, Maple and *Mathematica* systems provide either numerical or symbolic functions to solve $A\mathbf{x} = \mathbf{b}$.

- **Matrix Eigenvalue Problems**:

$$A\mathbf{x} = \lambda\mathbf{x} \tag{B.131}$$

is the eigenvalue problem statement, where the object is to find a set of characteristic values or eigenvalues λ_k and associated eigenvectors $\mathbf{x}_k$ that characterize the matrix A.

1. Since the algebraic problem $(A-\lambda_k I_n)\mathbf{x}_k = \mathbf{0}$ is equivalent to the original (B.131),
$$\text{Det}[A - \lambda I_n] = 0$$
is called the **characteristic** or **eigen equation**.
2. $(A - \lambda_k I_n)$ is a nth polynomial in λ_k,
$$P_n(\lambda) = \sum_{i=0}^{n} c_i \lambda^i \,,$$
where $c_0 = \text{Det}[A]$, $c_1 = -\text{Trace}[A] \cdots, c_n = (-1)^n$.
3. The characteristic equation is the condition for finding a non-trivial eigenvalue, $\mathbf{x}_k[x_{i,k}]_{n\times 1} \neq \mathbf{0}$.
4. Solving $\text{Det}[A - \lambda I_n] = 0$ yields n eigenvalues $[\lambda_i]_{n\times 1}$.
5. The eigenvectors can be found from a subset of the original problem, but are not unique.
6. If $\mathbf{x}_k$ is an eigenvector, then so is $\mathbf{y} = s * \mathbf{x}$, where s is an arbitrary, nonzero scalar.
7. A unit or normalized eigenvector is of the form $||\mathbf{u}_k||_p = 1$.
8. If A is real and symmetric, then the eigenvectors are **orthogonal**, $\mathbf{x}_j^\top \mathbf{x}_k = ||\mathbf{x}_k||_2^2 \delta_{j,k}$ or **orthonormal** if $||\mathbf{x}_k||_2 = 1$ in addition.
9. If A is not real and non-symmetric, then the left or adjoint eigen problem
$$\mathbf{y}_j^H A = \mu_j^* \mathbf{y}_j^H \quad \text{or} \quad A^H \mathbf{y}_j = \mu_j \mathbf{y}_j$$
would be needed for orthogonality conditions since $0 = (\lambda_k - \mu_j^*)\mathbf{y}_j^H \mathbf{x}_k$, so if $\mu_j^* \neq \lambda_k$ then $\mathbf{y}_j^H \mathbf{x}_k = 0$.

- **Gradient of a Scalar Valued Function of a Vector Argument**:

$$\nabla_{\mathbf{x}}[F](\mathbf{x}) = \frac{\partial F}{\partial \mathbf{x}}(\mathbf{x}) = F_{\mathbf{x}}(\mathbf{x}) \equiv \left[\frac{\partial F}{\partial x_i}(\mathbf{x})\right]_{n\times 1}, \tag{B.132}$$

so the gradient is a column vector with the same shape as $\mathbf{x}$ here. In some texts [44], the gradient may be a row vector, so matrix-vector products will be different there.

- **Gradient of a Matrix-Vector Product Transpose**:

$$\begin{aligned}\nabla_{\mathbf{x}}\left[(A\mathbf{x})^\top\right] &= \left[\frac{\partial}{\partial x_i}\sum_{k=1}^{n} a_{j,k}x_k\right]_{n\times n} = \left[\sum_{k=1}^{n} a_{j,k}\delta_{i,k}\right]_{n\times n} \\ &= [a_{j,i}]_{n\times n} = A^\top, \end{aligned} \tag{B.133}$$

so the gradient just peels off the pre-multiplied $\mathbf{x}^\top$, since $(A\mathbf{x})^\top = \mathbf{x}^\top A^\top$ (i.e., the **gradient peel theorem**).

- **Quadratic Forms**:

$$\mathcal{Q} = \mathbf{x}^\top A\mathbf{x} \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{i,j} x_j \tag{B.134}$$

which is a scalar, and since $\mathcal{Q}$ is a scalar and the transpose has no effect on scalars then

$$\mathcal{Q} = \mathcal{Q}^\top = \mathbf{x}^\top A^\top \mathbf{x} = \frac{1}{2}\left(\mathcal{Q} + \mathcal{Q}^\top\right) = \mathbf{x}^\top A^S \mathbf{x}, \tag{B.135}$$

where $A^S \equiv \frac{1}{2}(A + A^\top)$ is the symmetric part of A. Thus, for quadratic forms, the user might as well assume A to be symmetric or that $A^\top = A$.

- **Positive Definite Matrices**: The matrix A is **positive definite** if for every nonzero vector $\mathbf{x}$ ($\mathbf{x} \neq \mathbf{0}$) the quadratic form

$$\mathbf{x}^\top A\mathbf{x} > 0, \tag{B.136}$$

sometimes abbreviated as $A > 0$. Similarly, A is **positive semi-definite** if, for all $\mathbf{x} \neq \mathbf{0}$,

$$\mathbf{x}^\top A\mathbf{x} \geq 0, \tag{B.137}$$

or if so then we say $A \geq 0$. Further, A is positive definite if and only if all its eigen values are positive [67], so then A is invertible, i.e., A^{-1} exists.

- **Gradient of a Quadratic Form**:

$$\nabla_{\mathbf{x}}\left[\mathbf{x}^\top A\mathbf{x}\right] = 2A\mathbf{x}, \tag{B.138}$$

assuming A is symmetric, by two applications of the peel theorem, one on the left and another on the right by transposing first.

- **Hessian Matrix of a Scalar Valued Function**:

$$\nabla_{\mathbf{x}} \left[\nabla_{\mathbf{x}}^{\top} [F] \right] (\mathbf{x}) = \left[\frac{\partial^2 F}{\partial x_i \partial x_j} (\mathbf{x}) \right]_{n \times n} , \tag{B.139}$$

  so the matrix of second derivatives is a square $n \times n$ matrix.

- **Hessian Matrix of a Quadratic Form**:

$$\nabla_{\mathbf{x}} \left[\nabla_{\mathbf{x}}^{\top} [\mathbf{x}^{\top} A \mathbf{x}] \right] = \nabla_{\mathbf{x}} \left[2(A\mathbf{x})^{\top} \right] = 2 \nabla_{\mathbf{x}} \left[\mathbf{x}^{\top} A \right] = 2A \tag{B.140}$$

  by the peel theorem, assuming that A is symmetric.

B.10 Some Multivariate Distributions

The probability distributions, such as normal, exponential, and Poisson, previously considered have been functions of a single real sample variable representing a single random variate. However, some applications require multidimensional distributions representing jointly distributed multivariate random variables. The continuous multivariate normal (multinormal) distribution and the discrete multinomial distribution will serve as examples.

B.10.1 Multivariate Normal Distribution

Definition B.58. *The **multivariate normal distribution** for the real m-dimensional vector random variate $\mathbf{X} = [X_i]_{m \times 1} \in \mathbb{R}^m$ is defined by the density in matrix-vector notation as*

$$\phi_n(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \equiv \frac{1}{(2\pi)^{m/2} \sqrt{\mathrm{Det}[\boldsymbol{\Sigma}]}} \exp\left(-0.5 (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) , \tag{B.141}$$

*where $\boldsymbol{\mu} = [\mu_i]_{m \times 1} = \mathrm{E}[\mathbf{X}]$ is the **vector mean**,*

$$\boldsymbol{\Sigma} = [\sigma_{i,j}]_{m \times m} = \mathrm{E}\left[[(X_i - \mu_i)(X_j - \mu_j)]_{m \times m} \right]$$

*is the positive definite **variance-covariance matrix**, i.e., $\sigma_{i,i} \equiv \sigma_i^2 = \mathrm{Var}[X_i]$ for $i = 1 : m$, while $\sigma_{i,j} \equiv \mathrm{Cov}[X_i, X_j]$ if $j \neq i$ for $i, j = 1 : m$, and $\mathrm{Det}[\boldsymbol{\Sigma}]$ is the determinant of $\boldsymbol{\Sigma}$. The **correlation coefficient** is the normalized covariance,*

$$\rho_{i,j} \equiv \frac{\mathrm{Cov}[X_i, X_j]}{\sqrt{\mathrm{Var}[X_i] \mathrm{Var}[X_j]}} = \frac{\sigma_{i,j}}{\sigma_i \sigma_j} , \tag{B.142}$$

provided $\sigma_i,\ \sigma_j \neq 0$ and $i,\ j \neq 0$.

Total probability is conserved since

$$\int_{\mathbb{R}^m} \phi_n(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{x} = 1.$$

Theorem B.59. ***Correlation coefficient bounds:*** *Let X_1 and X_2 be two random variables, then*

$$|\rho(X_1, X_2)| \leq 1, \tag{B.143}$$

provided $\sigma_1 > 0$ and $\sigma_2 > 0$, but if $\rho(X_1, X_2) = \pm 1$, then

$$X_2/\sigma_2 = \pm X_1/\sigma_1 + C \tag{B.144}$$

for some constant C.

Proof. The proof is modeled after Feller's proof [83, p. 236]. Let $\rho = \rho(X_1, X_2)$ and using (B.74)

$$\begin{aligned} \mathrm{Var}[X_1/\sigma_1 \pm X_2/\sigma_2] = &\quad \mathrm{Var}[X_1/\sigma_1] \pm 2\mathrm{Cov}[X_1/\sigma_1, X_2/\sigma_2] + \mathrm{Var}[X_2/\sigma_2] \\ = &\quad 2(1 \pm \rho) \geq 0, \end{aligned}$$

since $\mathrm{Var}[X] \geq 0$, so $|\rho| \leq 1$.

If $\rho = 1$, then let $\pm 1 = -1$ and thus $X_1/\sigma_1 - X_2/\sigma_2 = C_1$ where C_1 is a constant, but if $\rho = -1$, then let $\pm 1 = +1$ and thus $X_1/\sigma_1 + X_2/\sigma_2 = C_2$ where C_2 is a constant. Combining these two cases leads to the form (B.144). □

Example B.60. *In the two-dimensional case, the* ***bivariate normal distribution****, with $\sigma_i > 0$ for $i = 1 : 2$, let $\sigma_{1,2} = \rho\sigma_1\sigma_2$ where $\rho = \rho_{1,2}$ is the correlation coefficient between state 1 and state 2 such that $-1 < \rho < +1$ to keep the density well defined. Thus,*

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \tag{B.145a}$$

$$\boldsymbol{\Sigma}^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2) \\ -\rho/(\sigma_1\sigma_2) & 1/\sigma_2^2 \end{bmatrix}. \tag{B.145b}$$

The $\boldsymbol{\Sigma}^{-1}$ follows upon calculating the two-dimensional inverse of $\boldsymbol{\Sigma}$, while substituting for $\boldsymbol{\Sigma}^{-1}$ and $\mathrm{Det}[\boldsymbol{\Sigma}] = (1-\rho^2)\sigma_1^2\sigma_2^2$ yields the more explicit density form:

$$\begin{aligned} \phi_n\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\Bigg(-\frac{0.5}{1-\rho^2}\Bigg[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 \\ -\frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\Bigg]\Bigg). \end{aligned} \tag{B.146}$$

Some of the first few moments are tabulated (results from the Maple™ symbolic computation system) in Table B.1:

Remark B.61. *The bivariate normal density becomes singular when $\sigma_1 \to 0^+$ or $\sigma_2 \to 0^+$ or $\rho^2 \to 1^-$ and the density becomes degenerate. If $\rho > 0$, then X_1 and*

Table B.1. *Some expected moments of bivariate normal distribution.*

Some Binormal Expectations
$\mathrm{E}[1] = 1$
$\mathrm{E}[x_i] = \mu_i,\ i = 1:2$
$\mathrm{Var}[x_i] = \sigma_i^2,\ i = 1:2$
$\mathrm{Cov}[x_1, x_2] = \rho\sigma_1\sigma_2$
$\mathrm{E}[(x_i - \mu_i)^3] = 0,\ i = 1:2$
$\mathrm{E}[(x_i - \mu_i)^4] = 3\sigma_i^4,\ i = 1:2$
$\mathrm{E}[(x_1 - \mu_1)^2(x_2 - \mu_2)^2] = (1 + 2\rho^2)\sigma_1^2\sigma_2^2$

*X_2 are **positively correlated**, while if $\rho < 0$, then X_1 and X_2 are **negatively correlated**.*

B.10.2 Multinomial Distribution

The multinomial distribution may be useful for studying discrete collections of samples from continuous distributions such as the bin frequencies of histograms and many other applications [83, 128].

Definition B.62. *Using m bins where π_k $(0 < \pi_k < 1)$ is the theoretical probability associated with the kth bin as well as a parameter of the distribution for $k = 1 : m$ bins such that*

$$\sum_{k=1}^{m} \pi_k = 1 \tag{B.147}$$

and f_k is the observed frequency (integer outcome count, $f_k \geq 0$) for the kth bin for a sample of N observations such that

$$\sum_{k=1}^{m} f_k = N\,, \tag{B.148}$$

*the **multinomial distribution** is given by the joint probability mass function*

$$p(\mathbf{f}; \boldsymbol{\pi}) = \mathrm{Prob}\left[\mathbf{F} = \mathbf{f} \,\middle|\, \mathbf{1}^T\boldsymbol{\pi} = 1, \mathbf{1}^T\mathbf{f} = N\right] = N! \prod_{k=1}^{m} \frac{\pi_k^{f_k}}{f_k!}\,, \tag{B.149}$$

where $\mathbf{f} = [f_i]_{m\times 1}$ is the frequency value vector, $\mathbf{F} = [F_i]_{m\times 1}$ is the random frequency vector, and $\mathbf{1} = [1]_{m\times 1}$ is the ones or summing vector.

Example B.63. *When $m = 2$, the multinomial distribution is called the **binomial distribution** and has probability function*

$$p(f_1, f_2; \pi_1, \pi_2) = \frac{N!\pi_1^{f_1}\pi_2^{f_2}}{f_1!f_2!} = \left(\begin{array}{c} N \\ f_1 \end{array}\right) \pi_1^{f_1}(1 - \pi_1)^{N - f_1}\,, \tag{B.150}$$

where the ***binomial coefficient***

$$\left(\begin{array}{c} n \\ k \end{array} \right) \equiv \frac{n!}{k!(n-k)!} \tag{B.151}$$

with the constraints $f_2 = N - f_1$ *and* $\pi_2 = 1 - \pi_1$ *used on the far right hand side. The binomial distribution is applicable to trials with just two outcomes, called* ***Bernoulli trials*** *(Feller [83]). Often these two outcomes or bins are identified as either a* ***success****, with probability* π_1*, or* ***failure****, for example, with probability* $\pi_2 = 1 - \pi_1$*. Feller [83] calls the binomial distribution, the normal distribution and the Poisson distribution the three principal distributions throughout probability theory.*

The ***binomial theorem*** *gives the* ***binomial expansion****,*

$$(\pi_1 + \pi_2)^N = \sum_{f_1=0}^{N} \left(\begin{array}{c} N \\ f_1 \end{array} \right) \pi_1^{f_1} \pi_2^{N-f_1} , \tag{B.152}$$

but the coefficients are precisely the binomial probability functions

$$(\pi_1 + \pi_2)^N = \sum_{f_1=0}^{N} p(f_1, N - f_1; \pi_1, \pi_2) , \tag{B.153}$$

which is why the distribution in (B.150) is called the Binomial distribution for binomial frequencies f_1 *for* $f_1 = 0 : N$ *(Feller [83]).*

Consequently, the binomial expectation for some function g *is given by*

$$\mathrm{E}[g(F_1)] = \sum_{f_1=0}^{N} g(f_1) p(f_1, N - f_1; \pi_1, 1 - \pi_1) .$$

Using parametric differentiation of the sums, with F_k *being the* k*th random variable and* f_k *being the* k*th given conditioned variable, it can be shown that*

- $\mathrm{E}[1] = 1$ *when* $g(f_k) = 1$ *(actually (B.152) or (B.153) with* $\pi_2 = \pi_1$*),*
- $\mathrm{E}[F_k] = N\pi_k$ *when* $g(f_k) = f_k$*,*
- $\mathrm{Var}[F_k] = N\pi_k(1 - \pi_k)$ *when* $g(f_k) = (f_k - N\pi_k)^2$*,*
- $\mathrm{Cov}[F_1, F_2] = -N\pi_1\pi_2 = -N\pi_k(1 - \pi_k) = -\mathrm{Var}[F_1]$ *when* $g(f_1) = (f_1 - N\pi_1)((N - f_1) - N(1 - \pi_1)) = -N(f_1 - N\pi_1)^2$*.*

As an illustration of an application of parametric differentiation to sum a finite number of terms, consider the first moment:

$$\begin{aligned} \mathrm{E}[F_1] &= \sum_{f_1=0}^{N} f_1 \left(\begin{array}{c} N \\ f_1 \end{array} \right) \pi_1^{f_1} (1 - \pi_1)^N - f_1 \\ &= \pi_1 \frac{d}{d\pi_1} \left[\sum_{f_1=0}^{N} \left(\begin{array}{c} N \\ f_1 \end{array} \right) \pi_1^{f_1} (\pi_2)^N - f_1 \right] \Bigg|_{\pi_2 = 1 - \pi_1} \\ &= \pi_1 \frac{d}{d\pi_1} \left[(\pi_1 + \pi_2)^N \right] \big|_{\pi_2 = 1 - \pi_1} = \pi_1 N \left[(\pi_1 + \pi_2)^{N-1} \right] \big|_{\pi_2 = 1 - \pi_1} = N\pi_1 . \end{aligned}$$

Similarly, forms with powers of $\{\pi_1, d/d\pi_1\}$ *can be used for higher moments.*

Figure B.6 illustrates the Binomial distributions as a function of the binomial frequency f_1 *when the total count is* $N = 10$ *for three values of the binomial probability parameter,* $\pi_1 = 0.25, 0.5$ *and 0.75. See Appendix C Section C.6 for the MATLAB figure code. These binomial distributions roughly resemble a discretized*

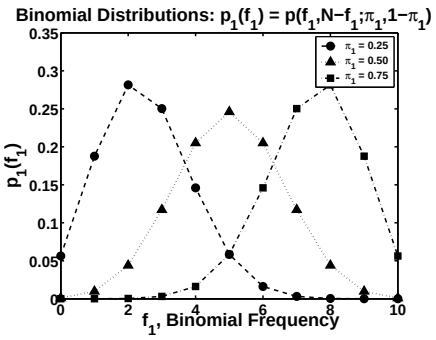


Figure B.6. *Binomial distributions with respect to the binomial frequency* f_1 *with* $N = 10$ *for values of the probability parameter,* $\pi_1 = 0.25, 0.5$ *and 0.75. These represent discrete distributions, but discrete values are connected by dashed, dotted and dash-dotted lines only to help visualize the distribution form for each parameter value.*

version of the normal distribution, except that they are skewed for $\pi_1 = 0.25$ *and 0.75 while the distribution for* $\pi_1 = 0.50$ *the distribution is symmetric. Feller [83] states that when* $N\pi_1(1-\pi_1)$ *is large that the binomial distribution can be approximated by the normal distribution with mean* $N\pi_1$ *and variance* $N\pi_1(1-\pi_1)$*, but when* N *is large and* π_1 *is the same order as* $1/N$ *then the binomial distribution can be approximated by the Poisson distribution with* $\Lambda = N\pi_1$ *order one. Since the Poisson can also be approximated by the normal approximation there is some overlap of the two approximations, but only the Poisson approximation is good when* $\Lambda = N\pi_1$ *is small.*

The multinomial distribution has the same basic moments as the binomial, but the constraints on the π_k and f_k also constrain the expectation summations. The multinomial distribution in (B.149) is in fact the terms in the multinomial expansion theorem,

$$\begin{aligned}\left(\sum_{k=1}^{m}\pi_k\right)^N &= N!\prod_{i=1}^{m-1}\left[\sum_{f_i=0}^{(N-\mathcal{F}_{i-1})}\frac{\pi_i^{f_i}}{f_i!}\right]\frac{\pi_m^{N-\mathcal{F}_{m-1}}}{(N-\mathcal{F}_{m-1})!}\,, \\ &= \prod_{i=1}^{m-1}\left[\sum_{f_i=0}^{(N-\mathcal{F}_{i-1})}\right]p(\mathbf{f};\boldsymbol{\pi})\Bigg|_{f_m=(N-\mathcal{F}_{m-1})}\,, \end{aligned} \tag{B.154}$$

which can be obtained from $(m-1)$ successive applications of the binomial expansion. It can be shown by induction upon replacing π_m by $(\pi_m + \pi_{m+1})$ in the

induction hypothesis above and using an additional application of the binomial expansion with the power $(N - \mathcal{F}_{m-1})$. Here, $\mathcal{F}_k \equiv \sum_{j=1}^{k} f_j$ is the partial sum of the first k frequencies, such that $\mathcal{F}_0 \equiv 0$. For application to the multinomial distribution, the constraints lead to the elimination formula $f_m = N - \mathcal{F}_{m-1}$ for the mth terms, so that the final fraction in (B.154) depends on the first $m-1$ sample frequencies f_k. In the case of the multinomial distribution, also the mth theoretical probability $\pi_m = 1 - \sum_{j=1}^{m-1} \pi_j$ can be eliminated by conservation of probability.

B.11 Basic Asymptotic Notation and Results

Definitions and Results B.64. *For purposes of a refined study of limits and asymptotic behaviors found in many stochastic problems, basic* ***asymptotic concepts*** *can be defined as*

- ***Equals big Oh*** *or* ***is the order of*** *symbol is such that* $f(x) = \mathrm{O}(g(x))$ *as* $x \to x_0$ *if* $f(x)/g(x)$ *is bounded as* $x \to x_0$ *provided* $g(x) \neq 0$ *in a deleted neighborhood of* $x = x_0$.

  For example: $8\sin(\epsilon/7) = \mathrm{O}(\epsilon)$ *as* $\epsilon \to 0$ *or* $(2N^2+3N+100)/(3N+5) = \mathrm{O}(N)$ *as* $N \to \infty$ *or* $\exp(-0.5\Delta t) = 1 - 0.5\Delta t + \mathrm{O}((\Delta t)^2)$ *as* $\Delta t \to 0$. *Also,* $\mathrm{O}(100\Delta t) = \mathrm{O}(\Delta t)$ *as* $\Delta t \to 0$, *since constants need not be considered. As alternate notation,* $\mathrm{O}((\Delta t)^2) = \mathrm{O}^2(\Delta t)$ *as* $\Delta t \to 0$.

- ***Equals little oh*** *or* ***is smaller order than*** *is such that* $f(x) = \mathrm{o}(g(x))$ *as* $x \to x_0$ *if* $f(x)/g(x) \to 0$ *as* $x \to x_0$ *provided* $g(x) \neq 0$ *in a deleted neighborhood of* $x = x_0$. *Also the notation* $f(x) \ll g(x)$ *is equivalent to* $f(x) = \mathrm{o}(g(x))$.

  For example: $\exp(-0.5\Delta t) = 1 - 0.5\Delta t + \mathrm{o}(\Delta t)$ *as* $\Delta t \to 0$ *or* $\int_t^{t+\Delta t} f(\tau)d\tau = f(t)\Delta t + \mathrm{o}(\Delta t)$ *as* $\Delta t \to 0$ *provided* $f(t)$ *is continuous. Note* $\mathrm{O}(\Delta t) + \mathrm{o}(\Delta t) = \mathrm{O}(\Delta t)$ *as* $\Delta t \to 0$.

- ***Equals*** ord *or* ***is the same order as*** *is such that* $f(x) = \mathrm{ord}(g(x))$ *as* $x \to x_0$ *if* $f(x) = \mathrm{O}(g(x))$ *but that* $f(x) \neq \mathrm{o}(g(x))$. *The relation* $f(x) \leq \mathrm{ord}(g(x))$ *is equivalent to* $f(x) = \mathrm{O}(g(x))$ *and* $f(x) < \mathrm{ord}(g(x))$ *is equivalent to* $f(x) = \mathrm{o}(g(x))$.

  For example: $(\Delta t)^2 < \mathrm{ord}(\Delta t)$ *as* $\Delta t \to 0$ *but* $\Delta t > \mathrm{ord}((\Delta t)^2)$ *as* $\Delta t \to 0$.

- *The* ***symbol*** $\sim$ *or* ***is asymptotic to*** *is such that* $f(x) \sim g(x)$ *as* $x \to x_0$ *if* $f(x)/g(x) \to 1$ *as* $x \to x_0$ *provided* $g(x) \neq 0$ *in a deleted neighborhood of* $x = x_0$.

  For example: $(1 - \exp(-0.425\Delta t))/\Delta t \sim 0.425$ *as* $\Delta t \to 0$.

  Remark B.65. *The symbol* $\sim$ *is commutative since if* $f(\epsilon) \sim g(\epsilon)$ *then* $g(\epsilon) \sim f(\epsilon)$ *as* $\epsilon \to 0$ *provided both* $f(\epsilon)$ *and* $g(\epsilon)$ *are not equal to zero in a*

*neighborhood of $\epsilon = 0$. Also, one should **never** say that $f(\epsilon) \sim 0$ (?) since according to our definition that would be dividing by zero.*

- *A sequence $\{\phi_n(x)\}$ for $n = 0 : \infty$ is an **asymptotic sequence** if $\phi_{n+1}(x) < \mathrm{ord}(\phi_n(x))$ as $x \to x_0$.*

  For example: $\phi_n(x) = (x - x_0)^n$ as $x \to x_0$ or $\phi_n(\Delta t) = (\Delta t)^{n/2}$ as $\Delta t \to 0^+$ for $n = 0 : \infty$.

- *An expansion $\sum_{n=0}^{\infty} a_n \phi_n(x)$, where a_n are coefficients constant in x and $\phi_n(x)$ are elements of an asymptotic sequence, is an **asymptotic expansion** which is asymptotic to a function $f(x)$ if*

$$f(x) - \sum_{n=0}^{N} a_n \phi_n(x) < \mathrm{ord}(\phi_N(x))$$

as $x \to x_0$ for all N, and if so then

$$f(x) \sim \sum_{n=0}^{\infty} a_n \phi_n(x)$$

as $x \to x_0$. As a corollary, the inductive algorithm for the coefficients follows starting with $a_0 = \lim_{x \to x_0} f(x)/\phi_0(x)$ and

$$a_{N+1} = \lim_{x \to x_0} \frac{f(x) - \sum_{n=0}^{N} a_n \phi_n(x)}{\phi_{N+1}(x)}$$

for $N = 0 : +\infty$, assuming that all limits exist.

For example, most convergent Taylor series, when considered under limiting conditions, are asymptotic expansions, or asymptotic power series in particular:

$$f(x) \sim \sum_{n=0}^{\infty} f^{(n)}(x_0)(x - x_0)^n / n!$$

as $x \to x_0$, but some asymptotic expansions can be divergent and still be useful if a finite number of terms are used, such as the expansion of the famous Stieltjes integral divergent asymptotic expansion example [28]

$$\int_0^{\infty} \frac{e^{-t} dt}{(1 + xt)} \sim \sum_{n=0}^{\infty} (-1)^n n! x^n$$

as $x \to 0$, which clearly diverges. For asymptotic applications, we are usually only interested in a few terms, whether the expansion is convergent or divergent, so the first few terms of a divergent expansion can be useful. Limits play a different role in asymptotic expansions then they do for Taylor series, in that limits of the independent variable (here, x) are used in asymptotics, while limits of the index (here, n) are used to test the convergence or divergence of Taylor series for a fixed value of the independent variable.

- *For integrals dominated by an exponential whose exponent, say $\phi(x)/\epsilon$, has a maximum at x^* within the interior of the range of integration (a, b) such that $\phi'(x^*) = 0$ and $\phi''(x^*) < 0$, i.e., $\phi(x) \sim \phi(x^*) + 0.5\phi''(x^*)(x - x^*)^2$, while $f(x) \sim f(x^*)$ is continuous and subdominant, as $x \to x^*$ and $0 < \epsilon << 1$,* ***Laplace's method*** *for* ***asymptotic evaluation of integrals*** *[28] leads to the asymptotic approximation,*

$$\int_a^b e^{\phi(x)/\epsilon} f(x)dx \sim \sqrt{\frac{2\pi\epsilon}{-\phi''(x^*)}} e^{\phi(x^*)/\epsilon} f(x^*), \tag{B.155}$$

as $\epsilon \to 0^+$. If $x^* = a$ or $x^* = b$, i.e., an end point maximum, then the integral is asymptotic to one half the above approximation.

For example, the general ***factorial function*** *or* ***gamma function*** *[2] for real x with $x > -1$,*

$$\begin{aligned} x! = \Gamma(x+1) = \int_0^\infty e^{-t} t^x dt = x^{x+1} \int_0^\infty e^{x(-y+\ln(y))} dy \\ \sim \sqrt{2\pi x} e^{-x} x^x \end{aligned} \tag{B.156}$$

as $x \to \infty$, after transforming the original integral to the Laplace form using $t = xy$ with $\phi(y) = -y + \ln(y)$ and $\epsilon = 1/x$, since the fast exponentially decaying coefficient function $\exp(-t)$ does not satisfy the subdominant requirement for Laplace's method. (Often, some transformation is necessary to fit a method.) The result is a correction to Stirling's (asymptotic) formula $\ln(x!) \sim x\ln(x)$, which is only the leading term of the exponent expansion of $x!$ as $x \to \infty$. Some authors refer to the leading term (B.156) of the full integral as Stirling's formula, e.g., Feller ([83]).

Remark B.66. Laplace and Probability:
Since Laplace was associated with the early foundational work in the analytical theory of probability in his treatise Théorie Analytique des Probabilités, it is likely that Laplace's method was developed for probability integrals, in particular normal probability integrals, which were not restricted to infinite or zero limits of integration and the integrals can be found exactly.

B.12 Generalized Functions: Combined Continuous and Discrete Processes

In stochastic problems, especially in extreme limits and distributions, representations beyond ordinary functions, such as generalized functions, are useful for the complete description of stochastic problems, such as combined continuous and discrete processes. While there are alternative abstract representations, generalized functions are very helpful in motivating stochastic models and solutions to associated stochastic problems as they are in the study of differential equations. Many

generalized functions are only defined under integration, but can be constructed as the limit of a sequence of ordinary functions.

Definitions B.67.

- *The **Heaviside step function**, $H(x)$, is a generalized function with the property that*

$$\int_{-\infty}^{+\infty} f(x)H(x-x_0)dx = \int_{x_0}^{+\infty} f(x)dx \tag{B.157}$$

for some integrable function $f(x)$ on $(-\infty, +\infty)$.

- ***Heaviside Step Function**:*
*One pointwise definition of the **Heaviside step function** is*

$$H(x) = \left\{ \begin{array}{ll} 0, & x < 0 \\ 1, & x \geq 0 \end{array} \right\}, \tag{B.158}$$

which is right-continuous, but another version takes the average value at zero so that it has better numerical properties,

$$H_a(x) = \left\{ \begin{array}{ll} 0, & x < 0 \\ 1/2, & x = 0 \\ 1, & x > 0 \end{array} \right\}, \tag{B.159}$$

although the Heaviside function is often left undefined at $x = 0$ since a single isolated point does not contribute to an ordinary or Riemann integral. For generalized functions, the averaged one, $H_a(x)$ is better for underlying numerical approximations.

- *For intervals on the real line, the right-continuous Heaviside step function is related to the **indicator function** for some set A,*

$$\mathbf{1}_{x\in A} \equiv \left\{ \begin{array}{ll} 1, & x \in A \\ 0, & x \notin A \end{array} \right\}, \tag{B.160}$$

so that

$$\mathbf{1}_{x\in(0,+\infty)} = H(x),$$

using the above Heaviside step function definition.

For example, the probability distribution can be written

$$\Phi_X(\xi) = \mathrm{E}_X[H(\xi - X)] = \mathrm{E}_X[\mathbf{1}_{X\in(-\infty,\xi]}], \tag{B.161}$$

provided the density is sufficiently continuous. Note that $\mathbf{1}_{(y-x)\in[a,b)} = \mathbf{1}_{x\in(y-b,y-a]}$, by definition, is a technique which becomes more useful in calculating multivariate probability distributions.

Definition B.68. ***Dirac Delta Function:***
The ***Dirac delta function****, $\delta(x)$, is a generalized function with the property that*

$$\int_{-\infty}^{+\infty} f(x)\delta(x-x_0)dx = f(x_0)\,, \tag{B.162}$$

for any continuous function $f(x)$ defined for x on $\mathbb{R}$ and some point x_0 on $\mathbb{R}$ (see B. Friedman [88]).

Remark B.69. *The generalized function $\delta(x-x_0)$ is not a regular function and it only has meaning in the integrand of an integral. Since $\delta(x-x_0)$ picks out a single value of the function $f(x)$, it must be concentrated at a point, i.e., for any $\epsilon > 0$,*

$$\int_{x_0-\epsilon}^{x_0+\epsilon} f(x)\delta(x-x_0)dx = f(x_0)\,.$$

Hence, for $\epsilon \to 0^+$, this integral will give the same answer $f(x_0)$, whereas for an ordinary integral of calculus and $f(x)$ continuous the answer will be $\mathrm{O}(\epsilon)$ as $\epsilon \to 0^+$ and thus zero in the limit. Consequently, the integral with $\delta(x-x_0)$ can be ignored away from the point of concentration x_0. The delta function, $\delta(x-x_0)$ is also called an ***impulse function*** *when it is used to impart an impulse to drive a differential equation.*

A simple constructive approximation that in the limit leads to the delta function $\delta(x)$ is the simple triangular approximation,

$$d_\epsilon(x) \equiv \frac{1}{\epsilon}\left\{ \begin{array}{ll} (1-|x|/\epsilon), & 0 \le |x| \le \epsilon \\ 0, & \epsilon \le |x| \end{array} \right\} \tag{B.163}$$

Now consider an arbitrary test function $f(x)$ that is continuous and continuously differentiable, then using the definition (B.163),

$$\begin{aligned} \int_{-\infty}^{+\infty} d_\epsilon(x)f(x)dx &= \frac{1}{\epsilon}\int_{-\epsilon}^{+\epsilon}(1-|x|/\epsilon)f(x)dx \\ &= \int_{-1}^{+1}(1-|y|)f(\epsilon y)dy \\ &= \int_{-1}^{+1}(1-|y|)[f(0)+\mathrm{O}(\epsilon)]dy \\ &= f(0)+\mathrm{O}(\epsilon) \to f(0)\,, \end{aligned}$$

as $\epsilon \to 0^+$. Since $d_\epsilon(x)$ has the same effect at $\delta(x)$ in the limit, it can be said that

$$\delta_{+0}(x) = \lim_{\epsilon\to 0^+} \overset{\text{gen}}{=} \delta(x)\,.$$

where the symbol of generalized equality is $\overset{\text{gen}}{=}$ defined below:

Definition B.70. ***Generalized Equality****:*
Let

$$g(x) \stackrel{\text{gen}}{=} h(x)$$

if for a sufficient class of test functions, $f(x)$, (sufficiently smooth, bounded with exponential decay as $x \to \infty$, depending on the application) both $g(x)$ and $h(x)$ have the same effect in integration,

$$\int_{-\infty}^{+\infty} f(x)g(x)dx = \int_{-\infty}^{+\infty} f(x)h(x)dx.$$

Using the Wiener process density $\phi_{W(t)}(w)$ (B.23), it can also be shown that in the generalized sense,

$$\phi_{W(0^+)}(w) \stackrel{\text{gen}}{=} \delta(w). \tag{B.164}$$

The generalized result (B.164) is obtained by examining the asymptotic limit as $t \to 0^+$,

$$\mathrm{E}[f(W(t))] = \int_{-\infty}^{+\infty} f(w)\phi_n(w;0,t)dw \to f(0)\,,$$

for a continuous, exponentially bounded test function $|f(w)| < K\exp(aw)$ for some $K > 0$ and $a < a_0$ for some a_0 is sufficient, since the negative quadratic exponent of the density dominates any simple exponential at infinity. One need only consider the finite interval $[-R, R]$ for some sufficiently large R, $R/\sqrt{t} \gg 1$ when $t \ll 1$ will suffice, so that the tail portion of the integral on $(-\infty, +\infty)$ is negligible

Remarks B.71. *:*

- *The technique suggested is Laplace's Method for integrals given in Eq. (B.155); see also references [60, 28], for instance, or Exercise 23.*
- *Since we are interested here in limits of the normal distribution and its density, the density has a delta function limit such that $\phi_{W(0^+)}(w) \stackrel{\text{gen}}{=} \delta(w)$ according (B.164), then the use of the $H(x)$ step function form (B.158) in the relation $\Phi_X(\xi) = \mathrm{E}_X[H(\xi - X)]$ (B.161) is inconsistent. This is because $\Phi_{\Delta W(t)}(0) = 1/2$ for all positive values of Δt, so*

  $$\Phi_{W(0^+)}(w) = \int_{-\infty}^{w} \delta(v)dv = \left\{ \begin{array}{ll} 0, & w < 0 \\ 1/2, & w = 0 \\ 1, w > 0 & \end{array} \right\} = H_a(w)$$

  or (B.159), since the averaged value at zero is needed. However, using the expectation form of the distribution (B.161) (normally, products of delta functions cannot be made), then

  $$\mathrm{E}[H(w - W(0^+))] = \int_{-\infty}^{+\infty} H(w - v)\delta(v)dv = H(w)$$

  which is incorrect when $w = 0$ when using the generalized limits for the normal density.

Examples B.72. ***Generalized Function:***

- $\delta(ax+b) \overset{\text{gen}}{=} (1/a)\delta(x+b/a)$, *for constant* $a>0$ *and* b, *by changing variables* $\xi = ax$ *in the integral definition (B.162).*

- $\delta(-x) \overset{\text{gen}}{=} \delta(x)$, *i.e.,* $\delta(x)$ *behaves as an even function, since* $f(0^-) = f(0)$ *if the function* f *is continuous.*

- $x\delta(x) \overset{\text{gen}}{=} 0$, *since by (B.162) with any* $f(x) = xF(x)$, $F(x)$ *continuous and* $x_0 = 0$,
$$\int_{-\infty}^{+\infty} F(x)x\delta(x)dx = 0 \cdot F(0) = 0\,.$$

- *Let* $f(x)$ *be any continuously differentiable function on* $\mathbb{R}$, *then the derivative of the Dirac delta function* $\delta'(x)$ *is defined by*
$$\int_{-\infty}^{+\infty} f(x)\delta'(x)dx = -f'(0)\,. \tag{B.165}$$
The motivation for this definition is the integration by parts calculus tool that
$$\int_{-\infty}^{+\infty} f(x)\delta'(x)dx = \left[f(x)\delta(x) - \int f'(x)\delta(x)dx\right]\Bigg|_{-\infty}^{+\infty} = -f'(0)\,,$$
where the fact that $\delta(x)$ *is concentrated at* $x=0$ *means the* $f(x)\delta(x)$ *vanishes at infinity since* $\delta(x)$ *dominates by vanishing faster than any* $f(x)$ *can grow.*

An alternate motivation is to use the original definition of $\delta(x-x_0)$ *in (B.162), assume that* $\delta(x-x_0)$ *is differentiable under the integral, i.e., in has been generated by a continuously differential approximation satisfying uniformity conditions, then*
$$\begin{aligned}\frac{d}{dx_0}\int_{-\infty}^{+\infty} f(x)\delta(x-x_0)dx &= -\int_{-\infty}^{+\infty} f(x)\delta'(x-x_0)dx \\ &= f'(x_0)\,,\end{aligned} \tag{B.166}$$
the minus sign arising from differentiating $(x-x_0)$ *with respect to* x_0 *as a simple application of the chain rule.*

- *Similarly,* $\delta''(x)$, *for twice continuously differentiable functions* f, *is defined in the generalized sense by*
$$\int_{-\infty}^{+\infty} f(x)\delta''(x)dx = +f''(0)\,, \tag{B.167}$$
derivable by two integrations by parts and using the concentration at $x = 0$. *The same result also follows by differentiating the integral definition of* $\delta(x-x_0)$ *in (B.162) twice.*

- *$H'(x) \stackrel{\text{gen}}{=} \delta(x)$ with respect to continuous function $f(x)$ for which $f(x)$ and its derivative vanish as $|x| \to \infty$, since by integration by parts,*

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x)H'(x)dx &= \left[f(x)H(x) - \int f'(x)H(x)dx\right]\Big|_{-\infty}^{+\infty} \\ &= -\int_0^{+\infty} f'(x)dx = f(0)\,.\end{aligned}$$

An alternate motivation for this result, is to start with the original definition of the Heaviside step function,

$$\begin{aligned}\frac{d}{dx_0}\int_{-\infty}^{+\infty} f(x)H(x-x_0)dx &= -\int_{-\infty}^{+\infty} f(x)H'(x-x_0)dx \\ &= -f(x_0)dx\,, \end{aligned} \tag{B.168}$$

so ignoring the two minus signs, we have $H'(x-x_0) \stackrel{\text{gen}}{=} \delta(x-x_0)$.

- *A discrete distribution can be transformed into a continuous distribution by using a sequence of delta functions such that the density for the discrete random variable X with $(m+1)$ possible discrete values $\{x_k|k=0:m\}$ each with probability π_k, such that the generalized density is given by*

$$\phi_X^{(\text{gen})}(x) \stackrel{\text{gen}}{=} \sum_{k=0}^{m} \pi_k \delta(x-x_k)\,.$$

Hence, the expectation of some function $f(x)$ is

$$\begin{aligned}\mathrm{E}_X^{(\text{gen})}[f(X)] &= \int_{-\infty}^{+\infty} f(x)\phi_X(x)dx = \sum_{k=0}^{m}\pi_k \int_{-\infty}^{+\infty} f(x)\delta(x-x_k)dx \\ &= \sum_{k=0}^{m}\pi_k f(x_k)\end{aligned}$$

which is the same formula as given in (B.48) previously. Also, conservation of probability is confirmed by

$$\mathrm{E}_x^{(\text{gen})}[1] = 1$$

using the discrete probability property (B.46). However, the implied probability distribution $\Phi_X^{(\text{gen})}(x)$ is problematic since neither definition, $H(x-x_k)$ or $H_a(x-x_k)$, of the step function is suitable at $x=x_k$, but see the appropriate right-continuous step function $H_R(x)$ ahead in (B.171).

Since it is an aim of the text to treat continuous and discrete distributions together, a unified applied treatment is needed. For this treatment, generalized

functions [185, 88], primarily step and delta functions, will be used for discrete distributions in a manner similar to the way they are used in differential equations, but more suited to stochastic processes. Thus, the *continued* discrete distribution will be illustrated and defined for the Poisson process since the probabilities are already ordered by integer values:

Lemma B.73.

- *The* ***Poisson distribution made right-continuous*** *is*

$$\Phi_{P(t)}(X) = \text{Prob}[X \le x] = \left\{ \begin{array}{ll} \sum_{j=0}^{\lfloor x \rfloor} p_j(\lambda t)\,, & x \ge 0 \\ 0\,, & x < 0 \end{array} \right\}\,, \quad \text{(B.169)}$$

  which readily follows and where $\lfloor x \rfloor$ *is the integer* ***floor function*** *such that* $x - 1 < \lfloor x \rfloor \le x$.

- *However, in terms of the* ***generalized right-continuous (RC) step-function*** $H_R(x)$ *this Poisson distribution can be generalized to*

$$\Phi_{P(t)}(X) = \sum_{k=0}^{\infty} p_k(\lambda t) H_R(x-k)\,, \quad \text{(B.170)}$$

  such that

$$H_R(x) = \left\{ \begin{array}{ll} 0\,, & x < 0 \\ 1\,, & x \ge 0 \end{array} \right\}\,, \quad \text{(B.171)}$$

  where the property $H_R(0) = H_R(0^+)$ *and* $H_R(0^-) = 0$ *embodies the required right-continuity property. Clearly,* $\Phi_{P(t)}(X)$ *is right-continuous, rather than purely continuous.*

Proof. The distribution form (B.169) follows directly from the definition of the continuous distribution using the discrete Poisson distribution $\text{Prob}[P(t) = k] = p_k(\lambda t)$ for $k = 0 : \infty$. Thus,

$$\text{Prob}[P(t) \le x] = \sum_{j=0}^{k} p_j(\lambda t),\ k \le x < k+1\,,$$

for $k = 0 : \infty$, since it takes k jumps for x to exceed k, i.e., $k = \lfloor x \rfloor$, so $k \le x < k+1$ is equivalent to $x - 1 < \lfloor x \rfloor \le x$, and any more will require the $(k+1)$st jump. Thus, the kth probability $p_k(\lambda t)$ is included in the sums if $x \ge k$, i.e., $p_k(\lambda t)$ is included in the form

$$p_k(\lambda t) H_R(x-k)$$

leading to B.170). □

Definition B.74. *The **Poisson process density** corresponding to this continuous distribution is denoted by*

$$\phi_{P(t)}(X) = \sum_{k=0}^{\infty} p_k(\lambda t)\delta_R(x-k)\,, \tag{B.172}$$

*where $\delta_R(x)$ is the **right-continuous (RC) delta function** such that*

$$H_R(x) = \int_{-\infty}^{x} \delta_R(y)dy \tag{B.173}$$

having the desired property that $H_R(0) = 1$ and the integral property

$$\int_{-\infty}^{\infty} f(y)\delta_R(y)dy = f(0^-). \tag{B.174}$$

These generalized functions and their properties will be encountered in more detail later in this text. The generalized $H_R(x)$ function is somewhat different from the concretely defined $H(x)$ in (B.158). Also, if the function f is continuous at $x = 0$ in B.174, then $f(0^-)$ can be replaced by $f(0)$.

The relationship between the exponential distribution and the Poisson distribution follows from the time of the arrival of the first jump T_1 under the standard assumption that the Poisson processes $P(t)$ starts at $t = T_0 \equiv 0$ and that the distribution for the first jump is the same as the probability that the Poisson jump counter exceeded one, i.e.,

$$\begin{aligned}\Phi_{T_1}(t;\lambda) \equiv \text{Prob}[T_1 \le t] = \text{Prob}[P(t) \ge 1] &= \sum_{k=1}^{\infty} p_k(\lambda t) \\ &= \sum_{k=1}^{\infty} e^{-\lambda t}\frac{(\lambda t)^k}{k!} = e^{-\lambda t}\left(e^{\lambda t} - 1\right) = 1 - e^{-\lambda t}\,,\end{aligned} \tag{B.175}$$

which is the same result as (B.40). The same result holds for the inter-arrival time, $T_{k+1} - T_k$, between successive Poisson jumps, except that the more general result depends on the property of stationarity of the Poisson process that is introduced in Chapter 1.

Summarizing distributions properties for combinations of continuous random variables and right-continuous jump processes, we have

Properties B.75. ***Right-Continuous Distribution Functions*** $\Phi(x)$*:*

- Φ *is **non-decreasing**, since probabilities must be non-negative.*
- Φ *is **right-continuous**, by properties of integrals with non-negative integrands including integrands with right-continuous delta functions or probability masses.*
- $\Phi(-\infty) = +0$*, by properties of integrals and* $X > -\infty$*.*
- $\Phi(+\infty) = 1$ *if* Φ *is a **proper distribution**.*

B.13 Fundamental Properties of Stochastic and Markov Processes

B.13.1 Basic Classification of Stochastic Processes

The classification of stochastic processes is important since the classification leads to the appropriate method of treatment of the stochastic process applications.

A **stochastic process** or **random process** is a random function of time $\xi = X(t;\omega)$ where $X(t;\omega)$ is a random variable depending on time t and some underlying random variable ω on the sample space Ω (again the ω dependence will often be suppressed unless it is needed to describe some stochastic process attribute).

If the time domain is continuous on some interval $[0,T]$ then it is said to be a **stochastic processes** in **continuous time** whether the domain is bounded or unbounded. However, if the time domain is discrete, $\xi = X_i$ in discrete time units $i = 1:\infty$ called **stages**, then it is a **stochastic process** in **discrete time** or random sequence. If $\xi = X(t)$ is not an random variable then $X(t)$ would be called a **deterministic process**.

Stochastic processes are also generally classified according to the properties of the range of the random variable $\xi = X(t)$, called the **state space** of the process. This state space can be continuous, in which case it is still referred to as a stochastic process, but if the state space is discrete with a finite or infinite number of states then the stochastic process is called a chain. The Gaussian process is an example of a process with a continuous state space, while the simple Poisson process with unit jumps is an example of a process with a discrete state space. A mixture of Gaussian and Poisson processes, called a jump-diffusion, is an example of a **hybrid stochastic system**.

B.13.2 Markov Processes and Markov Chains

An important class of stochastic process is the Markov process $X(t)$ in which the future state depends on only some current state but not on a past state. This Markov property offers many advantages in the analysis of the behavior of these processes.

Definitions B.76.

- *A stochastic process $X(t)$ for $t \geq 0$ in continuous time is a **Markov process** on a continuous state space $\mathcal{S}_{\text{csct}}$ if for any $t \geq 0$, $\Delta t \geq 0$ and $x \in \mathcal{S}_{\text{csct}}$,*

$$\text{Prob}[X(t+\Delta t) \leq x | X(s), s \leq t] = \text{Prob}[X(t+\Delta t) \leq x | X(t)]. \quad \text{(B.176)}$$

- *A stochastic process X_i for $i = 0:\infty$ in discrete time is a **Markov process** on continuous state space $\mathcal{S}_{\text{csdt}}$ if for any $n = 0:\infty$, $i = 0:\infty$, and $x_n \in \mathcal{S}_{\text{csdt}}$,*

$$\begin{aligned} &\text{Prob}[X_{n+1} \leq x_{n+1} | X_i = x_i,\ i = 0:n] \\ &\qquad = \text{Prob}[X_{n+1} \leq x_{n+1} | X_n = x_n]. \end{aligned} \quad \text{(B.177)}$$

- *A stochastic process $X(t)$ for $t \geq 0$ in continuous time is a* ***Markov chain*** *on discrete state space $\mathcal{S}_{\rm dsct} = \{0, 1, 2, \dots\}$ if for any $t \geq 0$, $\Delta t \geq 0$ and $j(t) \in \mathcal{S}_{\rm dsct}$,*

$$\begin{aligned} &\text{Prob}[X(t+\Delta t) = j(t+\Delta t) | X(s) = j(s), s \leq t] \\ &\qquad = \text{Prob}[X(t+\Delta t) = j(t+\Delta t) | X(t) = j(t)]. \end{aligned} \tag{B.178}$$

- *A stochastic process X_i for $i = 0 : \infty$ in discrete time is a* ***Markov chain*** *on discrete state space $\mathcal{S}_{\rm dsdt} = \{0, 1, 2, \dots\}$ if for any $n = 0 : \infty$, $i = 0 : \infty$ and $j_i \in \mathcal{S}_{\rm dsdt}$,*

$$\begin{aligned} &\text{Prob}[X_{n+1} = j_{n+1} | X_i = j_i,\ i = 0 : n] \\ &\qquad = \text{Prob}[X_{n+1} = j_{n+1} | X_n = j_n]. \end{aligned} \tag{B.179}$$

  The conditional probability $\text{Prob}[X_{n+1} = j_{n+1} | X_n = j_n] = P_{n,n+1}(j_n, j_{n+1})$ is called the ***transition probability*** *for the step from stage n to stage $n+1$.*

Thus, the Markov process can be called **memory-less** or **without after-effects** since, for example in the continuous time case, the **future state** $X(t+\Delta t)$ depends only on the **current state** $X(t)$, but not on the **past states** $\{x(s), s < t\}$. This memory-less property of Markov processes leads immediately to the **independent increments** property of Markov processes:

Lemma B.77. *If $X(t)$ is a Markov process in continuous time then the state* ***increment*** *$\Delta X(t) \equiv X(t+\Delta t) - X(t)$ is* ***independent*** *of $\Delta X(s) \equiv X(s+\Delta s) - X(s)$, $s, t, \Delta s, \Delta t \geq 0$, if the time intervals are disjoint except for trivial over-lap, i.e., either $s + \Delta s \leq t$ or $t + \Delta t \leq s$, such that*

$$\begin{aligned} \Phi_{\Delta X(t), \Delta X(s)}(\Delta x, \Delta y) &\equiv \text{Prob}[\Delta X(t) \leq \Delta x, \Delta X(s) \leq \Delta y] \\ &= \text{Prob}[\Delta X(t) \leq \Delta x] \text{Prob}[\Delta X(s) \leq \Delta y]. \end{aligned}$$

Note that the Markov property definition can be reformulated as

$$\text{Prob}[X(t+\Delta t) \leq x + \Delta x | X(s), s < t; X(t) = x] = \text{Prob}[\Delta X(t) \leq \Delta x | X(t) = x]$$

and thus independent of any increments in the past.

B.13.3 Stationary Markov Processes and Markov Chains

Definition B.78. *A Markov process is called* ***stationary*** *or* ***time-homogeneous*** *if the probability distribution depends only on the time difference, i.e.,*

- *if $\text{Prob}[X(t+\Delta t) - X(t) \leq y] = \text{Prob}[\Delta X(t) \leq y]$ depends on $\Delta t \geq 0$ and is independent of $t \geq 0$ in the continuous time case given y in the state space, continuous or discrete or*

- *if* $\text{Prob}[X_{i+k} - X_i \leq y]$ *depends on* $k \geq 0$ *and is independent of* $i \geq 0$ *in the discrete time case given* y *in the state space, continuous or discrete (it is also said that the* ***transition probabilities*** *are stationary).*

The stationary Markov chain in discrete time is fully characterized by the **transition probability matrix** $[P_{i-1,j-1}]_{N\times N}$ where $P_{i,j} = \text{Prob}[X_{n+1} = j | X_n = i]$ for all stages $n = 0 : N-1$ where N may be finite or infinite [265]. Although the main focus here is on Markov processes in continuous time, Markov chains serve as numerical approximation for Markov processes, such as in the Markov chain approximation methods of Kushner and co-workers [174, 175, 179].

B.14 Continuity, Jump Discontinuity and Non-Smoothness Approximations

In the standard calculus much of the emphasis is on functions that are continuous, differentiable, continuously differentiable or have similar nice properties. However, many of the models for Markov processes do not always have such nice analytical properties, since Poisson processes are discontinuous and Gaussian processes are not smooth. Thus, the standard calculus will be reviewed and revised to include the not so nice but essential properties.

B.14.1 Beyond Continuity Properties

If $X(t)$ is a process, i.e., function of time, whether stochastic or deterministic, the basic differences are here summarized:

Definitions B.79.

- *Let the* ***increment*** *for the process* $X(t)$ *be* $\Delta X(t) \equiv X(t+\Delta t) - X(t)$*, where* Δt *is the time increment.*
- *Let the* ***differential*** *for the process* $X(t)$ *be* $dX(t) \equiv X(t+dt) - X(t)$ *with respect to the time* t*, where* dt *is the infinitesimal time differential.*
- *The increment and differential are precisely related by the integral*

$$\Delta X(t) = \int_t^{t+\Delta t} dX(s).$$

While much of the regular calculus is usually cast in a more abstract form, much of applied stochastic calculus is based on differentials and increments, so much of the following will be formulated with increments or differentials, ready to use.

Definitions B.80.

- *The process $X(t)$ is a **continuous process** at the point t_0 if*

$$\lim_{\Delta t \to 0} X(t_0 + \Delta t) = X(t_0),$$

provided the limit exists;

- *Else the process $X(t)$ is **discontinuous** at t_0.*

- *The process $X(t)$ is **continuous** on the interval (t_1, t_2) if it is continuous at each point of the interval.*

- *The process $X(t)$ has a **jump discontinuity** at t_0 if*

$$\lim_{\substack{\Delta t \to 0 \\ |\Delta t| > 0}} X(t_0 + \Delta t) \neq X(t_0),$$

provided both the limit exists, i.e., the limit from the left

$$X(t_0^-) = \lim_{\Delta t \to 0+} X(t_0 - \Delta t)$$

and does not agree with the limit from the right

$$X(t_0^+) = \lim_{\Delta t \to 0+} X(t_0 + \Delta t),$$

where $\Delta t \to 0+$ means $\{\Delta t \to 0, \Delta t > 0\}$. In other words, if

$$X(t_0^+) \neq X(t_0^-)\,.$$

*then $X(t)$ has a **jump** at $t = t_0$ ([168]). The corresponding **jump** at the **jump discontinuity** (discontinuity of the first kind) is defined as*

$$[X](t_0) \equiv X(t_0^+) - X(t_0^-) = \lim_{\Delta t \to 0^+} X(t_0 + \Delta t) - \lim_{\Delta t' \to 0^+} X(t_0 - \Delta t'), \tag{B.180}$$

- *The process $X(t)$ is **right-continuous** at t_0 if*

$$\lim_{\substack{\Delta t \to 0 \\ \Delta t > 0}} X(t_0 + \Delta t) = X(t_0),$$

*such that the **jump of** X **at** t defined as*

$$[X](t_0) \equiv X(t_0) - X(t_0^-)\,, \tag{B.181}$$

*since $X(t_0^+) = X(t_0)$. **Left-continuous processes** are similarly defined.*

Remark B.81. *The jump definition is consistent with definition of the increment and consequently the differential, since if there is a jump at time t_1 then $dX(t_1^-) = X(t_1^- + dt) - X(t_1^-) = X(t_1^+) - X(t_1^-) = [X](t_1)$, accepting the convention that dt is both positive and infinitesimal so that $X(t_1^- + dt) = X(t_1^+)$. Similarly, for the increment $\Delta X(t_1^-) \to [X](t_1)$ as $\Delta t \to 0^+$.*

Definitions B.82.

- *The process $X(t)$ is **smooth** at t_0 if*

$$\lim_{\Delta t \to 0} \Delta X(t_0)/\Delta t$$

  exists, i.e., $X(t)$ is differentiable at t_0.

- *Else the process $X(t)$ is **non-smooth**.*

Remark B.83. *For example, if $\Delta X(t_1) \sim C\sqrt{\Delta t}$ for some non-trivial constant C, then $\Delta X(t_1) \to 0$ and $\Delta X(t_1)/\Delta t \sim C/\sqrt{\Delta t} \to \infty$ as $\Delta t \to 0^+$, so $X(t)$ is continuous but not smooth at t_1.*

B.14.2 Taylor Approximations of Composite Functions

Construction of application models often relies on **Taylor's formula** with remainder (Lagrange form) for small perturbations about some given point, given here in the form:

Theorem B.84. ***Taylor Approximation for a Scalar-Valued Function of a Scalar Argument,*** *$f(x)$:*
Let the function $f(x)$ be defined, continuous and be $(n+1)$ times continuously differentiable for $|\Delta x| < R$, then

$$f(x+\Delta x) = \sum_{m=0}^{n} \frac{f^{(m)}(x)}{m!}(\Delta x)^m + \frac{f^{(n+1)}(x+\theta\Delta x)}{(n+1)!}(\Delta x)^{n+1}\,, \quad \text{(B.182)}$$

*where $f^{(m)}(x)$ is the mth order derivative of f at x, $\theta \in (0,1)$ is the relative location of the mean value point $x+\theta\Delta x$ in the **remainder term** and R is the **convergence radius**.*

Further, if the highest derivative $f^{(n+1)}$ is bounded on the interval of convergence, $|\Delta x| < R$, then the remainder

$$S_n(x;\Delta x) - f(x+\Delta x) = O((\Delta x)^{n+1}),$$

as $\Delta x \to 0$, where

$$S_n(x;\Delta x) \equiv \sum_{m=0}^{n} \frac{f^{(m)}(x)}{m!}(\Delta x)^m\,,$$

is the **partial sum** of the first $(n+1)$ terms for $m = 0:n$.

For most applications, only a few terms are needed, while for stochastic applications in continuous time this form will be applied when the variable x is a process like $X(t)$. More generally, the interest is in functions that depend explicitly on time t and implicitly on time through the process $X(t)$, like $F(X(t),t)$. This is illustrated for a deterministic process increment in function $F(X(t),t)$, three times continuously differentiable in both t and x. First, the increment is split up to partially separate out the first argument $X(t)$-process and second t-argument explicit

time changes so that the one-dimensional Taylor approximation (B.182) can be separately applied to the component parts. Using partial derivatives,

Theorem B.85. ***Taylor Approximation for a Scalar-Valued Function of a Scalar Argument*** $X(t)$ ***and Time*** t, $f(X(t),t)$*:*
Let $f(x,t)$ *be three times differentiable in both* x *and* t*, let the process* $X(t)$ *be continuous and let* $\Delta X(t) = X(t+\Delta t) - X(t)$ *so* $X(t+\Delta t) = X(t) + \Delta X(t)$*, then*

$$\begin{aligned}
\Delta f(X(t),t) &\equiv f(X(t)+\Delta X(t), t+\Delta t) - f(X(t),t) \\
&= (f(X(t)+\Delta X(t), t+\Delta t) - f(X(t)+\Delta X(t),t)) \\
&\quad + (f(X(t)+\Delta X(t),t) - f(X(t),t)) \\
&= \frac{\partial f}{\partial t}(X(t),t)\Delta t + \frac{\partial f}{\partial x}(X(t),t)\Delta X(t) \qquad \text{(B.183)} \\
&\quad + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(X(t),t)(\Delta t)^2 + \frac{\partial^2 f}{\partial t \partial x}(X(t),t)\Delta t \Delta X(t) \\
&\quad + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(X(t),t)(\Delta X)^2(t) \\
&\quad + \mathrm{O}((\Delta t)^3) + \mathrm{O}((\Delta t)^2 \Delta X) + \mathrm{O}(\Delta t (\Delta X)^2) + \mathrm{O}((\Delta X)^3),
\end{aligned}$$

as $\Delta t \to 0$ *and* $\Delta X(t) \to 0$*.*

Remarks B.86.

- *Keeping the second order partial derivative terms written out explicitly is in anticipation that, although the process may be continuous, the process may not be smooth as in the case of the Gaussian process.*
- *The about expansion can be extended to vector processes* $\mathbf{X}(t) = [X_i(t)]_{nx\times 1}$ *and is best expanded by components.*
- *Another difference with the stochastic cases is that* X *will also be a function of the underlying probability space variable* ω*, so* $X = X(t;\omega)$ *and* $\Delta X = \Delta X(t;\omega) \to 0$ *in probability (only) as* $\Delta t \to 0^+$*. Since* $\Delta X(t;\omega)$ *may have an unbounded range, e.g., in the case that* $\Delta X(t;\omega)$ *is normally distributed as* $\Delta t \to 0^+$*, but* $\Delta t > 0$*, the boundedness part of the order symbol definition* O *would be invalid if, for instance, the* ΔX *in* $\mathrm{O}^3(\Delta X)$ *were replaced by* $\Delta X(t;\omega)$*. However, something like* $\mathrm{O}(\mathrm{E}[\Delta X^3(t;\omega)])$ *would be valid. Nevertheless, formula (B.183) will be useful as a preliminary or formal expansion calculation, prior to applying an expectation and neglecting very small terms.*

In the case where the space process is a vector function of time, then performing the Taylor expansion by components facilitates the calculation of the Taylor approximation:

Theorem B.87. ***Taylor Approximation for a Scalar-Valued Function of a Vector Argument*** $\mathbf{X}(t)$ ***and Time*** t, $f(\mathbf{X}(t),t)$*:*

Let $f(\mathbf{x},t)$ be three times differentiable in both $\mathbf{x}$ and t, let the column vector process $\mathbf{X}(t) = [X_i]_{nx\times 1}$ be continuous, i.e., by component, and let $\Delta\mathbf{X}(t) = \mathbf{X}(t+\Delta t) - \mathbf{X}(t)$ so $\mathbf{X}(t+\Delta t) = \mathbf{X}(t+\Delta t) + \Delta\mathbf{X}(t)$, then

$$\begin{aligned}
\Delta f(\mathbf{X}(t),t) &\equiv f(\mathbf{X}(t)+\Delta\mathbf{X}(t),t+\Delta t) - f(\mathbf{X}(t),t)\\
&= (f(\mathbf{X}(t)+\Delta\mathbf{X}(t),t+\Delta t) - f(\mathbf{X}(t)+\Delta\mathbf{X}(t),t))\\
&\quad + (f(\mathbf{X}(t)+\Delta\mathbf{X}(t),t) - f(\mathbf{X}(t),t))\\
&= \frac{\partial f}{\partial t}(\mathbf{X}(t)+\Delta\mathbf{X}(t),t)\Delta t\\
&\quad + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(\mathbf{X}(t)+\Delta\mathbf{X}(t),t)(\Delta t)^2 + \mathrm{O}((\Delta t)^3)\\
&\quad + \sum_{i=1}^{nx}\frac{\partial f}{\partial x_i}(\mathbf{X}(t),t)\Delta X_i(t)\\
&\quad + \sum_{i=1}^{nx}\sum_{j=1}^{nx}\frac{1}{2}\frac{\partial^2 f}{\partial x_i\partial x_j}(\mathbf{X}(t),t)\Delta X_i(t)\Delta X_j(t) + \mathrm{O}(||\Delta X||^3)\\
&= \frac{\partial f}{\partial t}(\mathbf{X}(t),t)\Delta t + \nabla_{\mathbf{x}}^{\top}[f](\mathbf{X}(t),t)\Delta\mathbf{X}(t) \qquad \text{(B.184)}\\
&\quad + \frac{1}{2}\frac{\partial^2 f}{\partial t^2}(\mathbf{X}(t),t)(\Delta t)^2 + \frac{1}{2}\Delta\mathbf{X}^{\top}(t)\nabla_{\mathbf{x}}\left[\nabla_{\mathbf{x}}^{\top}[f]\right](\mathbf{X}(t),t)\Delta\mathbf{X}(t)\\
&\quad + \nabla_{\mathbf{x}}\left[\frac{\partial f}{\partial t}\right](\mathbf{X}(t),t)\Delta\mathbf{X}(t)\Delta t\\
&\quad + \mathrm{O}((\Delta t)^3) + \mathrm{O}((\Delta t)^2||\Delta X||) + \mathrm{O}(\Delta t||\Delta X||^2) + \mathrm{O}(||\Delta X||^3),
\end{aligned}$$

as $\Delta t \to 0$ and $\Delta\mathbf{X}(t) \to \mathbf{0}$, where the gradient of f is the vector

$$\nabla_{\mathbf{x}}[f](\mathbf{X}(t),t) \equiv \left[\frac{\partial f}{\partial x_i}(\mathbf{X}(t),t)\right]_{nx\times 1},$$

the transpose vector is the row vector $\Delta\mathbf{x}^{\top} = [\Delta x_j]_{1\times nx}$, and $||\Delta\mathbf{x}||$ is some norm, e.g., the infinite norm $||\Delta\mathbf{x}||_{\infty} = \max_i[|Dx_i|]$.

In the case where there is a vector-valued function $\mathbf{f}$ depending on time t and a space process $\mathbf{X}(t)$ that is a vector function of time, then systematically performing the Taylor expansion by both $\mathbf{f}$ and $\mathbf{X}$ components as well at by the t argument of $\mathbf{f}$ and finally reasembling the results into matrix-vector form facilitates the calculation of the Taylor approximation:

Theorem B.88. ***Taylor Approximation for a Vector-Valued Function of a Vector Argument $\mathbf{X}(t)$ and Time t, $\mathbf{f}(\mathbf{X}(t),t)$:***
Let $\mathbf{f}(\mathbf{x},t) = [f_i(\mathbf{x},t)]_{nx\times 1}$ be three times differentiable in both $\mathbf{x}$ and t, let the column vector process $\mathbf{X}(t) = [X_i(t)]_{nx\times 1}$ be continuous, i.e., by component, and let

$\Delta\mathbf{X}(t) = \mathbf{X}(t+\Delta t) - \mathbf{X}(t)$ *so* $\mathbf{X}(t+\Delta t) = \mathbf{X}(t+\Delta t) + \Delta\mathbf{X}(t)$, *then*

$$
\begin{aligned}
\Delta\mathbf{f}(\mathbf{X}(t),t) &\equiv \mathbf{f}(\mathbf{X}(t)+\Delta\mathbf{X}(t), t+\Delta t) - \mathbf{f}(\mathbf{X}(t),t) \\
&= \mathbf{f}(\mathbf{X}(t)+\Delta\mathbf{X}(t), t+\Delta t) - \mathbf{f}(\mathbf{X}(t),t) \\
&= [f_i(\mathbf{X}(t)+\Delta\mathbf{X}(t), t+\Delta t) - f_i(\mathbf{X}(t),t)]_{nx\times 1} \\
&= \Big[\frac{\partial f_i}{\partial t}(\mathbf{X}(t),t)\Delta t + \sum_{j=1}^{nx}\frac{\partial f_i}{\partial x_j}(\mathbf{X}(t),t)\Delta X_j(t) \\
&\quad + \frac{1}{2}\frac{\partial^2 f_i}{\partial t^2}(\mathbf{X}(t)+\Delta\mathbf{X}(t),t)(\Delta t)^2 + \sum_{j=1}^{nx}\frac{\partial^2 f_i}{\Delta t\partial x_j}(\mathbf{X}(t),t)\Delta X_j(t)\Delta t \\
&\quad + \frac{1}{2}\sum_{k=1}^{nx}\sum_{j=1}^{nx}\frac{\partial^2 f_i}{\partial x_k\partial x_j}(\mathbf{X}(t),t)\Delta X_j(t)\Delta X_k(t) \\
&\quad + \mathrm{O}((\Delta t)^3) + \mathrm{O}((\Delta t)^2||\Delta X||) + \mathrm{O}(\Delta t||\Delta X||^2) + \mathrm{O}(||\Delta X||^3)\Big]_{nx\times 1} \\
&= \frac{\partial\mathbf{f}}{\partial t}(\mathbf{X}(t),t)\Delta t + \left(\Delta\mathbf{X}^\top(t)\nabla_{\mathbf{x}}\right)[\mathbf{f}](\mathbf{X}(t),t) \qquad \text{(B.185)} \\
&\quad + \frac{1}{2}\frac{\partial^2\mathbf{f}}{\partial t^2}(\mathbf{X}(t),t)(\Delta t)^2 + \left(\Delta\mathbf{X}^\top(t)\nabla_{\mathbf{x}}\right)\left[\frac{\partial\mathbf{f}}{\partial t}\right](\mathbf{X}(t),t)\Delta t \\
&\quad + \frac{1}{2}\left(\Delta\mathbf{X}(t)\Delta\mathbf{X}^\top(t)\right) : \left(\nabla_{\mathbf{x}}\nabla_{\mathbf{x}}^\top\right)[\mathbf{f}](\mathbf{X}(t),t) \\
&\quad + \mathrm{O}((\Delta t)^3) + \mathrm{O}((\Delta t)^2||\Delta X||) + \mathrm{O}(\Delta t||\Delta X||^2) + \mathrm{O}(||\Delta X||^3),
\end{aligned}
$$

as $\Delta t \to 0$ *and* $\Delta\mathbf{X}(t) \to \mathbf{0}$, *where the gradient of* $\mathbf{f}$ *is pre-multiplied by the transpose of* $\Delta\mathbf{X}(t)$ *so that dimension of* $\mathbf{f}$ *is obtained*

$$
(\Delta\mathbf{X}^\top(t)\nabla_{\mathbf{x}})[f](\mathbf{X}(t),t) \equiv \left[\sum_{j=1}^{nx}\Delta X_j(t)\frac{\partial f_i}{\partial x_j}(\mathbf{X}(t),t)\right]_{nx\times 1},
$$

the second order derivative Hessian is similarly arranged as scalar-valued operator double dot product

$$
\begin{aligned}
(\Delta\mathbf{X}(t)\Delta\mathbf{X}^\top(t) : \nabla_{\mathbf{x}})\nabla_{\mathbf{x}})^\top[\mathbf{f}](\mathbf{X}(t),t) \equiv \Bigg[&\sum_{j=1}^{nx}\sum_{k=1}^{nx}(\Delta X_j(t)\Delta X_k(t) \\
&\cdot\frac{\partial^2 f_i}{\partial x_k\partial x_j}(\mathbf{X}(t),t)\Bigg]_{nx\times 1},
\end{aligned}
$$

the transpose vector is the row vector $\Delta\mathbf{x}^\top = [\Delta x_j]_{1\times nx}$, *and* $||\Delta\mathbf{x}||$ *is some norm, e.g., the infinite norm* $||\Delta\mathbf{x}||_\infty = \max_i[|Dx_i|]$.

In general the double dot product is related to the trace of a matrix (B.122).

Definition B.89. ***Double Dot Product of Two Square Matrices:*** *where*

$$A : B \equiv \text{Trace}[AB] = \sum_{j=1}^{n} \sum_{k=1}^{n} A_{j,k} Bk, j \tag{B.186}$$

for square matrices A and B.

However, if the **process is discontinuous**, as it will be for the jumps of the Poisson process, then (B.183) is no longer valid since the assumption on $X(t)$ is not valid at the jump. Thus, if $X(t)$ has a jump discontinuity at $t = t_1$, then the most basic form for change in f, the jump, must be used,

Theorem B.90. ***Zero order Taylor Approximation or Jump Function Limit for a Scalar-Valued Function of a Discontinuous Vector Process Argument*** $\mathbf{X}(t)$ ***and Time*** t***,*** $f(\mathbf{X}(t), t)$*:*

$$\Delta f(\mathbf{X}(t_1^-), t_1^-) \to [f](\mathbf{X}(t_1), t_1) \equiv f(\mathbf{X}(t_1^+), t_1^+) - f(\mathbf{X}(t_1^-), t_1^-), \tag{B.187}$$

as $\Delta t \to 0^+$.

This result extends the jump function definition (B.180). For right continuous jumps t_1^+ can be replaced by t_1 (B.187) as in (B.181). The most fundamental changes in processes are the large jumps, such as crashes or rallies in financial markets or disasters and bonanzas in nature or machine failure and repair in manufacturing production. It is important to be able to handle jumps, even though the analysis may be much more complicated than for continuous processes.

B.15 Extremal Principles

Finding extremal properties, maxima and minima, through optimization is another area where nice function properties may be over-emphasized, but for many optimal control applications results are needed for more general functions, whether deterministic or random functions.

Definitions B.91. ***Extrema:***
Let $f(\mathbf{x})$ be defined on some connected domain $\mathcal{D}$ in $\mathbb{R}^m$.

- *Then $f(\mathbf{x})$ has an* ***global maximum*** *at $\mathbf{x}^*$ in $\mathcal{D}$ if $f(\mathbf{x}) \leq f(\mathbf{x}^*)$ for all $\mathbf{x}$ on $\mathcal{D}$.*
- *Similarly, $f(\mathbf{x})$ has an* ***global minimum*** *at some point $\mathbf{x}^*$ on $\mathcal{D}$ if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x}$ on $\mathcal{D}$.*
- *Often, such* ***global extrema*** *are called* ***absolute extrema****.*

- *Then $f(\mathbf{x})$ has a **local maximum** or **relative maximum** at $\mathbf{x}^*$ on $\mathcal{D}$ if there is a neighborhood, $\mathcal{N}(\mathbf{x}^*)$ of $\mathbf{x}^*$ on $\mathcal{D}$ such that $f(\mathbf{x}^* + \Delta\mathbf{x}) \leq f(\mathbf{x}^*)$ for sufficiently small $|\Delta\mathbf{x}|$.*

- *Similarly, $f(\mathbf{x})$ has a **local minimum** or **relative minimum** at $\mathbf{x}^*$ on $\mathcal{D}$ if there is a neighborhood, $\mathcal{N}(\mathbf{x}^*)$ of $\mathbf{x}^*$ on $\mathcal{D}$ such that $f(\mathbf{x}^* + \Delta\mathbf{x}) \geq f(\mathbf{x}^*)$ for sufficiently small $|\Delta\mathbf{x}|$.*

- *Often, such **local extrema** are called **relative extrema**.*

Remarks B.92.

- *The standard definition of **global extrema**, i.e., global maxima and global minima, covers all of the most extreme values, the biggest and the smallest, regardless of the analytic properties of the target function. The definition of global extrema is the most basic definition, the one we need to turn to when derivative methods fail. On the other hand, the finding global extrema is very difficult in general and is by no means a closed problem.*

- *However, the standard definition of **local extrema** are strictly interior extrema, due to restriction that the neighbor be in the domain of interest, so would exclude **boundary extrema** which may include the extreme value being sought.*

- *The **general recipe for global extrema** is often given by*

  1. *Find **local extrema**, usually restricted to where the target function is well-behaved.*
  2. *Find **boundary extrema**, perhaps also restricted to points where the function is well-behaved.*
  3. *Find the **function values at all points where the function is not well-behaved**, i.e., discontinuous, non-smooth, etc.*
  4. *Find the **most extreme values of all of the above** for the global extreme values.*

Theorem B.93. ***First Order Necessary Conditions for a Local Minimum (Maximum):***
Let $f(\mathbf{x})$ be continuously differentiable in an open neighborhood $\mathcal{N}(\mathbf{x}^*)$ of $\mathbf{x}^*$. If $\mathbf{x}^*$ is a local minimum (maximum), then $\nabla[f](\mathbf{x}^*) = \mathbf{0}$.

If $\nabla[f](\mathbf{x}^*) = \mathbf{0}$ then $\mathbf{x}^*$ is also called a **stationary point** or **interior critical point** of f. For proof see any good calculus or analysis text, else see Nocedal and Wright [221] for a proof using Taylor's approximation and for the following theorem.

Theorem B.94. ***Second Order Necessary and Sufficient Conditions for a Local Minimum (Maximum):***
Let $\nabla^2[f](\mathbf{x})$ *be continuous in an open neighborhood* $\mathcal{N}(\mathbf{x}^*)$ *of* $\mathbf{x}^*$.

- *If* $\mathbf{x}^*$ *is a local minimum (maximum) of* f, *then* $\nabla[f](\mathbf{x}^*) = \mathbf{0}$ *and* $\nabla^2[f](\mathbf{x})$ *is positive (negative) definite.*
- *If* $\nabla[f](\mathbf{x}^*) = \mathbf{0}$ *and* $\nabla^2[f](\mathbf{x})$ *is positive (negative) definite, then* $\mathbf{x}^*$ *is a minimum (maximum) of* f .

B.16 Exercises

Many of these exercises, depending on the instructor, can be done by MATLAB, Maple or *Mathematica*, but if theoretical, the Symbolic Toolbox of MATLAB will be needed.

1. Prove the **variance-expectation identity** for any random variable X:

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}^2[X] \,. \tag{B.188}$$

{*Note that* $\mathrm{E}^2[X] = (\mathrm{E}[X])^2$ *here, since squaring the operator also squares the value.*}

2. Prove the following identity for the **variance of the sum of two random variables** X and Y:

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + 2\mathrm{Cov}[X,Y] + \mathrm{Var}[Y] \,. \tag{B.189}$$

3. Prove the following identity for the **variance of the product of two random variables** X and Y,

$$\begin{aligned}\mathrm{Var}[XY] = \overline{X}^2\mathrm{Var}[Y] + 2\overline{XY}\mathrm{Cov}[X,Y] + \overline{Y}^2\mathrm{Var}[X] - \mathrm{Cov}^2[X,Y] \\ + 2\overline{X}\mathrm{E}[\delta X(\delta Y)^2] + 2\overline{X}\mathrm{E}[(\delta X)^2\delta X] + \mathrm{E}[(\delta X)^2(\delta Y)^2] \,,\end{aligned}$$

where $\overline{X} = \mathrm{E}[X]$ and $\overline{Y} = \mathrm{E}[Y]$ are means, while $\delta X = X - \overline{X}$ and $\delta Y = Y - \overline{Y}$ are deviations from the mean. Further, in the case that X and Y are independent random variables, show that

$$\mathrm{Var}[XY] = \overline{X}^2\mathrm{Var}[Y] + \overline{Y}^2\mathrm{Var}[X] + \mathrm{Var}[X]\mathrm{Var}[Y] \,. \tag{B.190}$$

4. Prove the **Chebyshev inequality**,

$$\mathrm{Prob}[|X| \geq \epsilon] \leq \mathrm{E}[|X|^2]/\epsilon^2 \,, \tag{B.191}$$

where $\epsilon > 0$.
{*Hint: It is sufficient to assume that a probability density* $\phi(x)$ *exists, convert* $\epsilon^2\mathrm{Prob}[|X| \geq \epsilon]$ *to integral form and use estimate of* ϵ^2 *to absorb it into the integrals as functions of* x.}

5. Prove the **Schwarz inequality** (**Cauchy-Schwarz inequality**) in terms of expectations,

$$\mathrm{E}[|XY|] \leq \sqrt{\mathrm{E}[X^2] \cdot \mathrm{E}[Y^2]}\,. \tag{B.192}$$

{*Hint (big): Use the fact that* $(u-v)^2 \geq 0$ *and let* $u = X/\sqrt{E[X^2]}$ *and* $v = Y/\sqrt{E[Y^2]}$*, assuming that* X *and* Y *have finite, positive variances. Alternatively, explore the characteristic roots of* $E[(\lambda X + Y)^2] \geq 0$ *and consider that if there are only real roots* λ_i *at the minimum, then the discriminant (square root argument) must be positive in the quadratic formula.*}

6. Prove **Jensen's inequality**: If f is a **convex function**, i.e., f is real and

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y) \tag{B.193}$$

for all x, y and $0 < \theta < 1$, then

$$\mathrm{E}[f(X)] \leq f(E[X])\,. \tag{B.194}$$

7. (a) Derive this simple form of **Bayes' Rule** for two related random variables X and Y:

$$\mathrm{Prob}[X = x | Y = y] = \frac{\mathrm{Prob}[Y = y,\ X = x]}{\mathrm{Prob}[Y = y]}, \tag{B.195}$$

provided $\mathrm{Prob}[Y = y] > 0$. {*Hint: you need only to use the conditional probability definition (B.83).*}

(b) Derive, using an expansion of (B.195) and also the law of total probability (B.92), the multiple random variables or events form of **Bayes' Rule** for the case of the random event Y that occurs in conjunction with a member of the exhaustive (complete) and countable set of disjoint (mutually exclusive) events, $\{X_i, i = 1:n\}$, i.e., the total law of probability if applicable,

$$\mathrm{Prob}[X_i = x_i | Y = y] = \frac{\mathrm{Prob}[Y = y,\ X_i = x_i]}{\sum_{j=1} \mathrm{Prob}[Y = y,\ X_j = x_j] \cdot \mathrm{Prob}[X_j = x_j]}\,.$$

8. For the **uniform distribution**, confirm the formulas for the mean, variance, coefficient of skewness and coefficient of kurtosis.

9. Derive the following identity between the standard normal and the general **normal distributions**,

$$\Phi_n((\ln(x) - \mu)/\sigma; 0, 1) = \Phi_n(\ln(x); \mu, \sigma^2)\,.$$

10. Show, for the **lognormal density** with random variable $X_{ln}(t)$, that the **maximum location**, the **mode** of the distribution or the **most likely value** is given by

$$x^* = \mathrm{Mode}\,[X_{ln}(t)] = \exp\left(\mu - \sigma^2\right)\,.$$

Also, compare the **mean** or **expected** value to the **mode** for the lognormal distribution by calculating the ratio

$$\mathrm{E}\left[X_{ln}(t)\right]/\mathrm{Mode}\left[X_{ln}(t)\right] \ ,$$

then compare this lognormal ratio to that for the normal variates,

$$\mathrm{E}\left[X_{n}(t)\right]/\mathrm{Mode}\left[X_{n}(t)\right] \ .$$

11. For the **exponential distribution**, confirm the formulas for the mean, variance, coefficient of skewness and coefficient of kurtosis.

12. Show the following equivalence between the **exponential distribution** expectation and the **uniform distribution** expectation,

$$\mathrm{E}_e[f(X_e)] = \mathrm{E}_u[f(-\mu \ln(X_u))]$$

for any integrable function f.

13. Show the sample moment formulas for a set of IID random variables X_k with $\mathrm{E}[X_k] = \mu$ and $\mathrm{Var}[X_k] = \sigma^2$ for $k = 1 : n$ of Subsection B.6 are correct, i.e.,

    (a) $\mathrm{E}[m_n] = \mu$ for sample mean m_n (B.109);
    (b) $\mathrm{E}[s_n^2] = (n-1)\sigma^2/n$ for sample variance s_n^2 (B.110);
    (c) $\mathrm{E}[\widehat{s}_n^2] = \sigma^2$ for sample variance unbiased estimate $\widehat{s}_n^2$ (B.111);
    (d) $\mathrm{Var}[m_n] = \sigma^2/n$ for sample mean m_n .

    Hint: See the remarks on page B38.

14. Show that for a set of IID random variables, that the covariance of the sample mean m_n and the sample variance s_n^2 satisfy

$$\mathrm{Cov}[m_n, s_n^2] = \mu_3/n \ ,$$

where the third central moment is $\mu_3 = \mathrm{E}[(X_k - \mu)^3]$. Discuss what probability property relating m_n and s_n^2 is implied by the result if the I.I.D. distribution is even like the normal distribution and what property is implied asymptotically as $n \to +\infty$. See Subsection B.6.

15. Let $\mathcal{S} = \sum_{k=1}^{n} X_k$ be the **partial sum** of n IID random variables $\{X_k\}$ each with mean $\mathrm{E}[X_k] = \mu$ and variance $\mathrm{Var}[X_k] = \sigma^2$. Further, let the mth central moment be defined as $\mu^{(m)} = \mathrm{E}[(X_k - \mu)^m]$, so that $\mu^{(1)} = 0$ and $\mu^{(2)} = \sigma^2$. Show that

    (a) $\mathrm{E}[\mathcal{S}] = n\mu$.
    (b) $\mathrm{Var}[\mathcal{S}] = n\sigma^2$.
    (c) $\mathrm{E}[(\mathcal{S} - \mathrm{E}[\mathcal{S}])^3] = n\mu^{(3)}$, so is zero if the distribution of X_k has no **skew** (B.11).

(d) $\mathrm{E}[(\mathcal{S}-\mathrm{E}[\mathcal{S}])^4] = n\mu^{(4)} + 3n(n-1)\sigma^2$, where the first term is related to the coefficient of **kurtosis** (B.12).
Hint: Use the binomial theorem, $\mathcal{S}-\mathrm{E}[\mathcal{S}] = \sum_{k=1}^{n}(X_k-\mu)$ and the fact $\mu^{(1)}=0$.

16. Show that the **product of two normal densities** is a proportional to a normal density, i.e.,

$$\phi_n(x;\mu_1,\sigma_1^2)\cdot\phi_n(x;\mu_2,\sigma_2^2) = \phi_n\left(x;\frac{\mu_1\sigma_2^2+\mu_2\sigma_1^2}{\sigma_1^2+\sigma_2^2},\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\right) \cdot\frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\left(-\frac{(\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right). \tag{B.196}$$

Hint: Apply the completing the square technique to combine the two densities.

17. Let X_i be independent normal random variables with density $\phi_{X_i}(x)$, mean μ_i and variance σ_i^2 for $i=1$ to K:

(a) Show that the product of two normal densities is a normal density whose mean is the sum of the means and whose variances is the sum of the variances, using (B.196),

$$\begin{aligned}\mathcal{I}_2(x) \equiv (\phi_{X_1} * \phi_{X_2})(x) &= \int_{-\infty}^{+\infty}\phi_{X_1}(x-y)\phi_{X_2}(y)dy \\ &= \phi_n(x;\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)\,.\end{aligned} \tag{B.197}$$

(b) Using (B.197) for $K=2$ as the induction initial condition, show the general result by induction that

$$\mathcal{I}_K(x) \equiv \left(\left(\prod_{i=1}^{K-1}\phi_{X_i}*\right)\phi_{X_K}\right)(x) = \phi_n\left(x;\sum_{i=1}^{K}\mu_i,\sum_{i=1}^{K}\sigma_i^2\right). \tag{B.198}$$

18. Show that the distribution of the sum of two (2) IID random variables, U_1 and U_2 unifomly distributed on $[a,b]$, is a **triangular distribution** on $[2a,2b]$, i.e. show in terms of densities that

$$\begin{aligned}\phi_{U_1+U_2}(x) &= \int_{-\infty}^{+\infty}\phi_{U_1}(x-y)\phi_{U_2}(y)dy \\ &= \frac{1}{(b-a)^2}\left\{\begin{array}{cc}(x-2a), & 2a\le x<b+a\\ (2b-x), & b+a\le x\le 2b\\ 0, & \text{otherwise}\end{array}\right\},\end{aligned} \tag{B.199}$$

Confirm that the resulting density conserves probability on $(-\infty,+\infty)$.
Hint: It may be helpful to sketch the paths for non-zero integration in y on the xy-plane, paying attention to the limits of integration are for each fixed x.

Remark B.95. *Different from the normal distribution results in (17) of the previous exercise, the convolution of two uniform random variables does not conserve the uniformity of the distribution.*

19. Show the that the distribution of the sum of three (3) IID random variables, U_i, for $i = 1:3$ unifomly distributed on $[a, b]$, is a piecewise **quadratic distribution** on $[3a, 3b]$, i.e. show in terms of densities that

$$\phi_{\sum_{i=1}^{3} U_i}(x) = \int_{-\infty}^{+\infty} \phi_{U_1+U_2}(x-y)\phi_{U_3}(y)dy \tag{B.200}$$
$$= \frac{1}{2(b-a)^3} \left\{ \begin{array}{cc} +(x-3a)^2, & 3a \le x < 2a+b \\ \left\{ \begin{array}{c} -(x-(b+2a))^2 \\ +2(b-a)^2 \\ -(2b+a-x)^2 \end{array} \right\}, & 2a+b \le x < a+2b \\ +(3b-x)^2, & a+2b \le x \le 3b \\ 0, & \text{otherwise} \end{array} \right\},$$

using the result of the previous exercise for $\phi_{U_1+U_2}(x)$.
Hint: With this and the previous exercise, symbolic computation may be more desirable, e.g,, Maple™ or Mathematica™ .

20. For the **bivariate normal distribution**, verify the inverse of $\mathbf{\Sigma}$ in (B.145) and the explicit form for the density (B.146). Also, confirm by iterated integration that $\mathrm{E}[X_1] = \mu_1$, $\mathrm{Var}[X_1] = \sigma_1^2$ and $\mathrm{Cov}[X_1, X_2] = \rho\sigma_1\sigma_2$. {*Hint: Only techniques such as* ***completing the square*** *and transformations to the generic integral*

$$\int_{-\infty}^{+\infty} \exp(-x^2/2)[c_0 + c_1 x + c_2 x^2]dx = \sqrt{2\pi}[c_0 \cdot 1 + c_2 \cdot 1]$$

for any constants $\{c_0, c_1, c_2\}$.}

21. For the **binomial distribution** in (B.150) verify that the given basic moments are correct, i.e., $\mathrm{E}[F_k] = N\pi_k$ and $\mathrm{Var}[F_k] = N\pi_k(1-\pi_k)$ for $k = 1:2$.

22. Show that $W(0^+) = 0$ **with probability one** by showing that $\phi_{W(0^+)}(w) \stackrel{\text{gen}}{=} \delta(w)$, i.e., in the generalized sense, which means that

$$\mathrm{E}[f(W(t))] = \int_{\infty}^{+\infty} \phi_{W(t)}(w) f(w) dw \to f(0^+)$$

as $t \to 0^+$ for continuous, continuously differentiable and sufficiently bounded functions $f(w)$ which vanishing at infinity.
{*Hint: For formal justification, scale* t *out of the density by a change of variables in the integral and expand* f *for small* t*, assuming that the exponential convergence property of the normal density allows term-wise integration of the expansion. Note that if* $X(t)$ *is in the set* $\mathcal{S}$ *with probability one simply means that* $\mathrm{Prob}[X(t) \in \mathcal{S}] = 1$.
If more rigor is desired, use the asymptotic techniques, such as Laplace's method for integrals (B.155, Page B53), from the text and Exercise 23.}

23. **Asymptotic Analysis, Generalized Function Problem**:
Show that the following sequences for approximate the right-continuous step-function $H_R(x)$ in (B.171) and the right-continuous delta function $\delta_R(x)$ in (B.173),

$$H_{R,n}(x) = \int_{-\infty}^{x} \delta_{R,n}(y)dy\,;$$
$$\delta_{R,n}(x) \equiv e^{-(y+\mu_n)^2/(2\epsilon_n)}/\sqrt{2\pi\epsilon_n},$$

where $\epsilon_n > 0$, $\mu_n > 0$, $\sqrt{\epsilon_n} \ll \mu_n \ll 1$ when $n \gg 1$. That is, show for $n \gg 1$ that $H_{R,n}(0) = H_{R,n}(0^+) \sim 1$, $H_{R,n}(0^-) \to 0^+$ and

$$\int_{-\infty}^{+\infty} f(y)\delta_{R,n}(y-x)dy \sim f(x^-)\,,$$

for any continuous function $f(x)$ that is exponentially bounded, $|f(x)| \leq Ke^{-a|x|}$ on $(-\infty, +\infty)$ with $a > 0$ and $K > 0$, justifying the use of $H_{R,n}(x) \to H_R(x)$ and $\delta_{R,n}(x) \to \delta_R(x)$ as $n \to \infty$ for the generalized representation of Poisson processes.
{*Hint: When using the Laplace asymptotic approximation of integrals technique [60, 28], changing variables to* $\xi = y - x + \mu_n$*, selecting the integral tail-cutoff* $(-\rho_n, \rho_n)$ *in* ξ *about the argument of the maximum of* $\delta_{R,n}(\xi - \mu_n)$ *at* $\xi = 0$ *with* $\epsilon_n \ll \rho_n^2 \ll \mu_n \ll 1$ *so that the tails are exponentially negligible being dominated by the factor* $\exp(-\rho_n^2/(2\epsilon_n))$*, approximate* $f(x - \mu_n + \xi) \sim f(x - \mu_n)$ *using continuity and then change variables to* $\eta = \xi/\sqrt{\epsilon_n}$ *so that the limits of integration can be expanded to* $\pm\infty$*. The order in which these approximations are performed is critical.*}

Suggested References for Further Reading

- Bartlett, 1978 [19].
- Bender and Orszag, 1978 [28].
- Çinlar, 1975 [55].
- Copson, 1965 [60].
- Cox and Miller, 1968 [62].
- Doob, 1953 [69].
- Feller, 1968 [83].
- Feller, 1971 [84].
- Friedman, 1956 [88].
- Glasserman, 2003 [96].

- Higham and Higham, 2000 [142].
- Karlin and Taylor, 1975 [161].
- Karlin and Taylor, 1981 [162].
- Lighthill, 1964 [185].
- Moler et al., 2000 [210].
- Neftci, 2000 [217].
- Nocedal and Wright, 1999 [221].
- Pliska, 1997 [225].
- Parzen, 1962 [224].
- Ross, 1983 [237].
- Ross, 2000 [238].
- Taylor and Karlin, 1998 [265].
- Taylor and Mann, 1972 [263].
- Tuckwell, 1995 [270].

Appendix C

MATLAB Programs (Online)

This appendix contains a selection of basic MATLAB *m*-file programs used in this text to produce figures and are listed here as sample code for readers. They may be eventually moved to on-line access only if space becomes an issue. Also, since these *m*-files were used to produce figures for this book, they have more elaborate cosmetic figure enhancements, requiring full screen height, than would normally would be used for purely testing purposes.

C.1 Program: Uniform Distribution Simulation Histograms

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function uniform03fig1
% Book Illustration for Uniform RNG Simulation
clc % clear variables, but must come before globals,
%     else clears globals too.
clf % clear figures
fprintf('\nfunction uniform03fig1 OutPut:')
kfig = 0;
for m = 3:2:5
    kfig = kfig+1; figure(kfig);
    N=10^m;
    x=rand(N,1);
    xmean=mean(x);
    xstd=std(x);
    xmin = min(x);
    xmax = max(x);
    remean=(xmean*2-1)*100;
    restd=(xstd*sqrt(12)-1)*100;
    fprintf( ...
        '\n fig=%i; m=%2i; N=%i; xmean=%f; xstd=%f; min(x)=%f; max(x)=%f' ...
```

```
        ,kfig,m,N,xmean,xstd,xmin,xmax);
    fprintf('\n fig=%i; relerrmean=%f; relerrstd=%f;' ...
        ,kfig,remean,restd);
    nbins = 30; % min(fix(sqrt(10^m)),101);
    xmin = 0; xmax = 1;
    xbin1 = xmin; xbin2 = xmax; dxbin = (xbin2-xbin1)/nbins;
    xbin = xbin1+dxbin/2:dxbin:xbin2-dxbin/2;
    fprintf( ...
        '\n fig=%i; #bins(x)=%4i; xbin in [%6f,%6f]; dxbin=%10f;' ...
        ,kfig,nbins,xbin1,xbin2,dxbin)
    nx = hist(x,xbin); % Need Edge Oriented histc.
    bar(xbin,nx)
    axis tight
    htitle=title('Histogram for x = rand(N,1)');
    ks = [0.1,0.8]; nxmax = max(nx);
    ytext=fix(ks(2)*nxmax); xtext=ks(1);
    textn=['N = ' int2str(N)];
    htext = text(xtext,ytext,textn);
    hylabel=ylabel('Uniform Bin Frequency');
    hxlabel=xlabel('x, Uniform rand-Deviate');
    patchobj = findobj(gca,'Type','patch');
    haxis = gca;
    set(haxis,'Fontsize',20,'FontName','Helvetica' ...
        ,'FontWeight','Bold','linewidth',2)
    set(patchobj,'FaceColor','w','EdgeColor','k','linewidth',2);
    set(htitle,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(htext,'Fontsize',20,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(hylabel,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(hxlabel,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
end
%End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.2 Program: Normal Distribution Simulation Histograms

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function normal03fig1
% Book Illustration for Normal RNG Simulation
clc % clear variables, but must come before globals,
%     else clears globals too.
```

```
clf % clear figures
fprintf('\nfunction normal03fig1 OutPut:')
kfig = 0;
for m = 3:2:5
    kfig = kfig+1; figure(kfig);
    N=10^m;
    x=randn(N,1);
    xmean=mean(x);
    xstd=std(x);
    remean=xmean*100;
    restd=(xstd-1)*100;
    fprintf('\nNormal Random Deviate (MATLAB randn) Test:');
    fprintf('\n fig=%i; m=%2i; N=%i; xmean=%f; xstd=%f;' ...
        ,kfig,m,N,xmean,xstd);
    fprintf('\n fig=%i; relerrmean=%f; relerrstd=%f;'....
        ,kfig,remean,restd);
    nbins = 50; % min(fix(sqrt(10^m)),101);
    xmin = min(x); xmax = max(x);
    xbin1 = xmin; xbin2 = xmax; dxbin = (xbin2-xbin1)/nbins;%
    fprintf('\n#bins(x)=%4i; xbin in [%6f,%6f]; dxbin=%10f;' ...
        ,nbins,xbin1,xbin2,dxbin)
    xbin = xbin1+dxbin/2:dxbin:xbin2-dxbin/2;
    nx = hist(x,xbin); % Need Center Oriented hist.
    bar(xbin,nx)
    axis tight
    htitle=title('Histogram for x = randn(N,1)');
    ks = [0.4,0.7];  nxmax = max(nx);
    xtext = xmax*(ks(1)-(kfig-1)*0.1); ytext=fix(ks(2)*nxmax);
    textn=['N = ' int2str(N)];
    haxis = gca;
    htext = text(xtext,ytext,textn);
    hylabel=ylabel('Normal Bin Frequency');
    hxlabel=xlabel('x, Normal randn-Deviate');
    set(haxis,'Fontsize',20,'FontName','Helvetica' ...
        ,'FontWeight','Bold','linewidth',2)
    patchobj = findobj(gca,'Type','patch');
    set(patchobj,'FaceColor','w','EdgeColor','k','linewidth',2);
    set(htitle,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(htext,'Fontsize',20,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(hylabel,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(hxlabel,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
end
```

```
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.3 Program: Lognormal Distribution Simulation Histograms

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function lognormal03fig1
% Book Illustration for LogNormal RNG Simulation
clc % clear variables, but must come before globals,
%     else clears globals too.
clf % clear figures
fprintf('\nfunction lognormal03fig1 OutPut:')
kfig = 0; mu = 0.0; sig = 0.5;
muln = exp(mu+sig^2/2);
sigln = muln*sqrt(exp(sig^2) -1);
nbins = 150;
fprintf('\n mu=%f; sig=%f; muln=%f; sigln=%f; nbins=%i' ...
    ,mu,sig,muln,sigln,nbins);
for m = 3:2:5
    kfig = kfig+1; figure(kfig);
    N = 10^m;
    y = mu*ones(N,1) + sig*randn(N,1);
    x = exp(y);
    xmean=mean(x);
    xstd=std(x);
    remean=(xmean/muln - 1)*100;
    restd=(xstd/sigln - 1)*100;
    fprintf('\nLognormal Random Deviate (exp(mu+sig*randn)) Test:');
    fprintf('\n fig=%i; m=%2i; N=%i; xmean=%f; xstd=%f;' ...
        ,kfig,m,N,xmean,xstd);
    fprintf('\n fig=%i; relerrmean=%f; relerrstd=%f;' ...
        ,kfig,remean,restd);
    xmin = min(x); xmax = max(x);
    xbin1 = xmin; xbin2 = xmax; dxbin = (xbin2-xbin1)/nbins;%
    fprintf('\n#bins(x)=%4i; xbin in [%6f,%6f]; dxbin=%10f;' ...
        ,nbins,xbin1,xbin2,dxbin)
    xbin = xbin1+dxbin/2:dxbin:xbin2-dxbin/2;
    nx = hist(x,xbin); % Need Center Oriented hist.
    bar(xbin,nx)
    axis tight
    htitle=title('Histogram for Lognormal x');
    ks = [0.4,0.7];  nxmax = max(nx);
    xtext = xmax*(ks(1)-(kfig-1)*0.1); ytext=fix(ks(2)*nxmax);
    textn=['N = ' int2str(N)];
```

```
    haxis = gca;
    htext = text(xtext,ytext,textn);
    hylabel=ylabel('Lognormal Bin Frequency');
    hxlabel=xlabel('x, Lognormal Deviate');
    set(haxis,'Fontsize',20,'FontName','Helvetica' ...
        ,'FontWeight','Bold','linewidth',2)
    patchobj = findobj(gca,'Type','patch');
    set(patchobj,'FaceColor','w','EdgeColor','k','linewidth',2);
    set(htitle,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(htext,'Fontsize',20,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(hylabel,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(hxlabel,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
end
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.4 Program: Exponential Distribution Simulation Histograms

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function exponential03fig1
% Book Illustration for Exponential RNG Simulation with mean one.
clc % clear variables, but must come before globals,
%     else clears globals too.
clf % clear figures
fprintf('\nfunction exponential03fig1 OutPut:')
kfig = 0; mu = 1.0;
for m = 3:2:5
    kfig = kfig+1; figure(kfig);
    N=10^m;
    x=-mu*log(rand(N,1));
    xmean=mean(x);
    xstd=std(x);
    remean=(xmean/mu-1)*100;
    restd=(xstd/mu-1)*100;
    fprintf('\nExponential Random Deviate (MATLAB randn) Test:');
    fprintf('\n fig=%i; m=%2i; N=%i; xmean=%f; xstd=%f;' ...
        ,kfig,m,N,xmean,xstd);
    fprintf('\n fig=%i; relerrmean=%f; relerrstd=%f;' ...
        ,kfig,remean,restd);
    nbins = 50; % min(fix(sqrt(10^m)),101);
```

```
    xmin = 0; xmax = max(x);
    xbin1 = xmin; xbin2 = xmax; dxbin = (xbin2-xbin1)/nbins;%
    fprintf('\n#bins(x)=%4i; xbin in [%6f,%6f]; dxbin=%10f;' ...
        ,nbins,xbin1,xbin2,dxbin)
    xbin = xbin1+dxbin/2:dxbin:xbin2-dxbin/2;
    nx = hist(x,xbin); % using centered defined bins,
    %                    rather than edge bins
    bar(xbin,nx)
    axis tight
    htitle=title('Histogram for x = -ln(rand(N,1))');
    ks = [0.6,0.6];  nxmax = max(nx);
    xtext = xmax*ks(1); ytext=fix(ks(2)*nxmax);
    textn=['N = ' int2str(N)];
    htext = text(xtext,ytext,textn);
    hylabel=ylabel('Exponential Bin Frequency');
    hxlabel=xlabel('x, Exponential random-Deviate');
    haxis = gca;
    set(haxis,'Fontsize',20,'FontName','Helvetica' ...
        ,'FontWeight','Bold','linewidth',2)
    patchobj = findobj(gca,'Type','patch');
    set(patchobj,'FaceColor','w','EdgeColor','k','linewidth',2);
    set(htitle,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(htext,'Fontsize',20,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(hylabel,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
    set(hxlabel,'Fontsize',24,'FontName','Helvetica' ...
        ,'FontWeight','Bold')
end
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.5 Program: Poisson Distribution versus Jump Counter k

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function poisson03fig1
% Book Illustration for Poisson distribution with 3 parameter values.
clc % clear variables, but must come before globals,
%     else clears globals too.
fprintf('\nfunction poisson03fig1 OutPut:');
lv =[0.2,1.0,2.0,5.0]; nlam = 4;
nk = 10; kv = 0:nk;
for ilam = 1:nlam
```

```
    pv(1,ilam) = exp(-lv(ilam)), kv(1) = 0;
    for k = 1:nk
        kv(k+1) = k;
        pv(k+1,ilam) = pv(1,ilam)*(lv(ilam))^k/factorial(k);
    end
end
plot(kv,pv(:,1),'ko--',kv,pv(:,2),'k^:',kv,pv(:,3),'ks-.' ...
    ,kv,pv(:,4),'kd-.' ...
    ,'MarkerSize',10,'MarkerFaceColor','k','LineWidth',2)
htitle=title('Poisson Distributions: p_k(\Lambda)');
hylabel=ylabel('p_k(\Lambda)');
hxlabel=xlabel('k, Poisson Counter');
hlegend=legend('\Lambda = 0.2','\Lambda = 1.0','\Lambda = 2.' ...
    ,'\Lambda = 5.',0);
haxis = gca;
set(haxis,'Fontsize',20,'FontName','Helvetica','FontWeight' ...
    ,'Bold','linewidth',2)
set(htitle,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hylabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hxlabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hlegend,'Fontsize',20,'FontName','Helvetica','FontWeight','Bold')
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.6 Program: Binomial Distribution versus Binomial Frequency f_1

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function binomial03fig1
% Book Illustration for Binomial distribution
%     with 3 \pi_1 parameter values.
% pv(f_1) = p(f1,N-f1;\pi_1,1-\pi_1)
%         = Bi(N,f_1)*\pi_1^{f_1}*(1-\pi_1)^{N-f_1}
clc % clear variables, but must come before globals,
%     else clears globals too.
clf % clear figures
fprintf('\nfunction binomialfig03 OutPut:');
pi1v =[0.25,0.5,0.75]; npi1 = 3;
N = 10; f1v = 0:N; nfact = factorial(N);
for ipi = 1:npi1
    pi1 = pi1v(ipi);
    pv(1,ipi) = (1-pi1)^N;
    for f1 = 1:N
        pv(f1+1,ipi) = nfact/(factorial(f1)*factorial(N-f1)) ...
            *pi1^f1*(1-pi1)^(N-f1);
```

```
    end
end
plot(f1v,pv(:,1),'ko--',f1v,pv(:,2),'k^:',f1v,pv(:,3),'ks-.' ...
    ,'MarkerSize',10,'MarkerFaceColor','k','LineWidth',2)
htitle=title('Binomial Distributions: p_1(f_1) = p(f_1,N-f_1;\pi_1,1-\pi_1)');
hylabel=ylabel('p_1(f_1)');
hxlabel=xlabel('f_1, Binomial Frequency');
hlegend=legend('\pi_1 = 0.25','\pi_1 = 0.50','\pi_1 = 0.75',1);
haxis = gca;
set(haxis,'Fontsize',20,'FontName','Helvetica','FontWeight' ...
    ,'Bold','linewidth',2)
set(htitle,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hylabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hxlabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hlegend,'Fontsize',14,'FontName','Helvetica','FontWeight','Bold');
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.7 Program: Simulated Diffusion $W(t)$ Sample Paths

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function wiener06fig1
% Book Illustration for Wiener/Diffusion Process RNG Simulation ...
%   for t in [0,1] with sample variation.
%   Generation is by summing Wiener increments DW of even spacing Dt.
clc % clear workspace of prior output.
clear % clear variables, but must come before globals,
% else clears them.
fprintf('\nfunction wiener06fig1 OutPut:'); % print code figure name
nfig = 0;
N = 1000; TF = 1.0; Dt = TF/N; % Set initial time grid: Fixed Delta{t}.
NP = N + 1; % Number of points.
mudt = 0; sqrtdt = sqrt(Dt);  % Set standard Wiener increment moments'
% for dX(t) = mu*dt + sigma*dW(t); here mu = 0, sigma = 1
%     and scaled dW(t) = sqrt(dt)*randn
% Begin Calculation:
tv = 0:Dt:TF; % time row-vector
nstate = 4; % number of states
jv = [1,2,3,4]; % selection of states; change when needed
DWv = zeros(nstate,N); Wv = zeros(nstate,NP); % DW & W vectors/arrays;
% Also sets initial Wv(j,1) = 0;
for j = 1:nstate
    randn('state',jv(j)); % Set initial state for repeatability;
    DWv(j,1:N) = sqrtdt*randn(1,N);  %Generate N sample random row-vector;
```

```
    for i=1:N % Simulated Sample paths by Increment Accumulation:
        Wv(j,i+1) = sum(DWv(j,1:i)); % Note Wv(j,1) = 0.0; sum is effic.
    end
end
%%%%% Begin Plot:
nfig = nfig + 1;
scrsize = get(0,'ScreenSize'); % figure spacing for target screen
ss = [5.0,4.0,3.5]; % figure spacing factors
fprintf('\n\nFigure(%i):  Diffusion Simulated Sample Paths(4)\n' ...
    ,nfig)
figure(nfig)
marks = {'k-','k-o','k-^','k-x'}; % easier to change marks with nstate
%
for j = 1:nstate
    plot(tv,Wv(j,1:NP),marks{j},'linewidth',2); hold on;
end
hold off
%
title('Diffusion Simulated Sample Paths (4)'...
    ,'FontWeight','Bold','Fontsize',44);
ylabel('W(t), Wiener State'...
    ,'FontWeight','Bold','Fontsize',44);
xlabel('t, Time'...
    ,'FontWeight','Bold','Fontsize',44);
hlegend=legend('State 1','State 2','State 3','State 4'...
    ,'Location','Best');
set(hlegend,'Fontsize',36,'FontWeight','Bold');
set(gca,'Fontsize',36,'FontWeight','Bold','linewidth',3);
set(gcf,'Color','White','Position' ...
   ,[scrsize(3)/ss(nfig) 60 scrsize(3)*0.60 scrsize(4)*0.80]);
% [l,b,w,h]
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.8 Program: Simulated Diffusion $W(t)$ Sample Paths Showing Variation with Time Step Size

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function wiener06fig2
% Book Illustration for Wiener/Diffusion Process RNG Simulation ...
%   for t in [0,1] with sample variation.
%   Generation is by summing Wiener increments DW of even spacing Dt.
clc % clear workspace of prior output.
clf % clear figures, else accumulative.
clear % clear variables, but must come before globals, else clears them.
```

```
fprintf('\nfunction wiener06fig2 OutPut:'); % print code figure name
nfig = 1;
N = 1000; TF = 1.0; Dt = TF/N; % Set time grids: Several dt's.
NP = N+1; % Total number of Points.
% for dX(t) = mu*dt + sigma*dW(t); here mu = 0, sigma = 1
%     and scaled dW(t) = sqrt(dt)*randn
% Begin Calculation:
% nstate = 1; % number of states.
ndt = 3; % number of local dt's.
jv = [1,2,3,4]; % selection of states; change when needed
randn('state',jv(1));  % Set common initial state for repeatability
RN = randn(1,N); % common random sample of N points.
Wv = zeros(ndt,NP); % W array of local vectors;
% Also sets all Wv(kdt,1) = 0 for tv(1) = 0;
% recall MATAB is unit based.
ts = zeros(ndt,NP); % Declare maximal local time vectors;
%%%%% Begin Plot:
nfig = nfig + 1;
scrsize = get(0,'ScreenSize'); % figure spacing for target screen
ss = [5.0,4.0,3.5]; % figure spacing factors
fprintf('\n\nFigure(%i):  Diffusion Simulated Sample Paths(4)\n' ...
    ,nfig)
figure(nfig)
marks = {'k-','k-o','k-^','k-x'}; % easier to change marks with nstate
%
for kdt = 1:ndt % Test Multiple Sample Paths with different dt's:
    S = 10^(kdt-1); % dt scalar factor;
    Ns = N/S; NPs = Ns+1; % Local counts;
    Dts = S*Dt; % Local time steps;
    sigs = sqrt(Dts); % Local diffusion scaling;
    ts(kdt,1:NPs) = 0:Dts:TF; % Local times;
    for i = 1:Ns % Simulated Sample paths by Increment Accumulation:
        Wv(kdt,i+1) =  Wv(kdt,i) + sigs*RN(1,i*S);
    end
    plot(ts(kdt,1:NPs),Wv(kdt,1:NPs),marks{kdt},'linewidth',2); hold on;
end
%
hold off
%
title('Diffusion Simulations: \Delta{t} Effects'...
    ,'FontWeight','Bold','Fontsize',44);
ylabel('W(t), Wiener State'...
    ,'FontWeight','Bold','Fontsize',44);
xlabel('t, Time'...
    ,'FontWeight','Bold','Fontsize',44);
hlegend=legend('\Delta{t} = 10^{-3}, N = 1000'...
```

```
    ,'\Delta{t} = 10^{-2}, N = 100' ...
    ,'\Delta{t} = 10^{-1}, N = 10','Location','Best');
set(hlegend,'Fontsize',36,'FontWeight','Bold');
set(gca,'Fontsize',36,'FontWeight','Bold','linewidth',3);
set(gcf,'Color','White','Position' ...
   ,[scrsize(3)/ss(nfig) 60 scrsize(3)*0.60 scrsize(4)*0.80]);
% [l,b,w,h]
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.9 Program: Simulated Simple Poisson $P(t)$ Sample Paths

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function poisson03fig2
% Book Illustration for Simple Poisson/Jump Process RNG Simulation ...
%   for P(t) = 1:K jumps with sample variation.
%   Generation is by Poisson Jump Exponentially distributed
%      jump time increments T(k+1)-T(k), T(k+1) = kth jump time,
%      T(1) := 0.
%
clc % clear variables, but must come before globals,
%     else clears globals too.
fprintf('\nfunction poisson03fig2 OutPut:')
kfig = 0;
K = 10; KP = 2*K +1; % Include sample of K jumps only.
p = zeros(KP,1); kstates = 4; LT = zeros(KP,kstates);
% Begin Calculation:
for kstate = 1:kstates; % Test Multiple Simulated Sample Paths:
    LT(1,kstate) = 0; p(1) = 0; % Set initial scaled jump time
    %                             and jump count.
    rand('state',kstate);  % Set initial state for repeatability
    %                        or path change.
    DTe = -log(rand(K,1));  % Generate random vector of
    %                         K exponential variates.
    for k = 1:K % Simulated sample scaled jump times
        %         LT(k+1) = lambda*T(k+1):
        LT(2*k,kstate) = LT(2*k-1,kstate) + DTe(k);
        LT(2*k+1,kstate) = LT(2*k,kstate);
        p(2*k) = p(2*k-1);
        p(2*k+1) = p(2*k-1) + 1;
    end
end
% Begin Plot:
kfig = kfig + 1;
```

```
fprintf('\n\nFigure(%i):  Simulated Jump Sample Paths\n',kfig)
figure(kfig)
plot(LT(1:KP,1),p,'k-',LT(1:KP,2),p,'k:',LT(1:KP,3),p,'k-.' ...
    ,LT(1:KP,4),p,'k--','LineWidth',2);
htitle=title('Simulated Simple Jump Sample Paths');
hylabel=ylabel('P(t), Poisson State');
hxlabel=xlabel('\lambda\cdot{t}, Scaled Time');
hlegend=legend('Sample 1','Sample 2','Sample 3','Sample 4',0);
haxis = gca;
set(haxis,'Fontsize',20,'FontName','Helvetica','FontWeight','Bold' ...
    ,'linewidth',2)
set(htitle,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hylabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hxlabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hlegend,'Fontsize',20,'FontName','Helvetica','FontWeight','Bold')
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.10 Program: Simulated Simple Incremental Poisson $\Delta P(t)$ Sample Paths

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function poisson03fig3
% Book Illustration for Simple Incremental Poisson/Jump Process RNG
%   Simulation for Delta{P}(t) = P(t+Delta{t})-P(t) = 1:K jumps
%   with sample variation.
%   Generation is by Poisson Jump Zero-One Law:
%         Prob(Delta{P}(t)=0] = 1-lambda*dt,
%      assuming sufficiently small Delta{t}'s.
%
clc % clear variables, but must come before globals,
%     else clears globals too.
clf % clear figures, else accumulative.
fprintf('\nfunction delpois03fig3 OutPut:')
kfig = 1;
figure(kfig);
marks = {'k-','k:','k-.','k--'};
% marks = {'k-o','k:s','k-.^','k--d'};
K = 10; KS = 500;  KP = KS + 1; % Include first K jumps from total
%                                  KS sample only.
kstates = 4; DP = zeros(KP,kstates); DT = zeros(KP,kstates);
% Begin Calculation:
for kstate = 1:kstates; % Test Multiple Simulated Sample Paths:
    k = 0; DP(1,kstate) = 0.0; DT(1,kstate) = 0;  % Set initial
    %                                               jump parms.
```

```
    rand('state',kstate-1);  % Set initial state for repeatability
    %                          or path change.
    xu = rand(KS,1);  % Generate random vector of K uniform variates.
    dt = 0.05; lambda = 1.0; % Set time step and jump rate.
    ldt = lambda*dt; % one jump prob.
    xl = (1-ldt)/2;  xr = (1+ldt)/2; % Set centered jump probability
    %                                  thresholds, using centered
    %                                  part of uniform distribution
    %                                  to avoid open end point bias.
    ip = 0; % Set plot counter.
    for i = 1:KS % Simulated sample scaled jump times
        %          LT(k+1) = lambda*T(k+1):
        ip = ip + 1;
        if xu(i) <= xr && xu(i) >= xl % Get jump if prob. in [xl,xr].
            k = k + 1;
            DP(ip+1,kstate) = DP(ip,kstate);
            DT(ip+1,kstate) = DT(ip,kstate) + dt;
            ip = ip + 1;
            DP(ip+1,kstate) = DP(ip,kstate) + 1;
            DT(ip+1,kstate) = DT(ip,kstate);
        else
            DP(ip+1,kstate) = DP(ip,kstate);
            DT(ip+1,kstate) = DT(ip,kstate) + dt;
        end
        if k == K
            KP = ip + 1;
            fprintf('\n kstate = %i; i = %i points; k = %i jumps;' ...
                ,kstate-1,i,k);
            break;
        end
    end
    plot(DT(1:KP,kstate),DP(1:KP,kstate),marks{kstate} ...
        ,'LineWidth',2), hold on
end
% Begin Plot:
fprintf('\n\nFigure(%i):  Simulated Small \Delta{t} Jump Sample Paths\n' ...
    ,kfig)
htitle=title('Simulated Small \Delta{t} Simple Jump Sample Paths');
hylabel=ylabel('\Delta{P}(t), Poisson State');
hxlabel=xlabel('t,  Time');
hlegend=legend('Sample 1','Sample 2','Sample 3','Sample 4',0);
haxis = gca;
set(haxis,'Fontsize',20,'FontName','Helvetica','FontWeight','Bold','linewidth',2)
set(htitle,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hylabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hxlabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
```

```
set(hlegend,'Fontsize',20,'FontName','Helvetica','FontWeight','Bold');
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.11 Program: Simulated Diffusion Integrals $\int (dW)^2$(t) by Itô Partial Sums

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function intdwdw
% Example MATLAB code for integral of (dW)^2.
clc % clear variables;
t0 = 0.0; tf = 1.0;
n = 1.0e+4; nf = n + 1; % set time grid: (n+1) subintervals
dt = (tf-t0)/nf;  %       and (n+2) points;
% replace these particular values according the application;
t(1) = t0; % set initial time at i = 1 for MATLAB;
W(1) = 0.0; % set initial diffusion noise condition;
sqrtdt = sqrt(dt); % dW(i) noise time scale so E[dW] = 0;
sumdw2(1) = 0.0; % set initial sum variable;
kstate = 1; randn('state',kstate); % Set randn state
%    for repeatability;
dW = sqrtdt*randn(nf,1); % simulate (n+1)-dW(i)'s sample;
t = t0:dt:tf; % get time vector t;
for i = 1:nf % simulate integral sample path.
    W(i+1) = W(i) + dW(i); % sum diffusion noise;
    sumdw2(i+1) = sumdw2(i) + (dW(i))^2; % sum whole integrand;
end
fprintf('\n\nFigure 1: int[(dW)^2](t) versus t\n');
figure(1)
plot(t,sumdw2,'k-',t,t,'k--','LineWidth',2); % plot sum;
htitle=title('\int(dW)^2(t) Simulations versus t');
hylabel=ylabel('\int(dW)^2(t) and t, States');
hxlabel=xlabel('t, Time');
hlegend=legend('\int(dW)^2(t)','t',0);
haxis = gca;
set(haxis,'Fontsize',20,'FontName','Helvetica' ...
    ,'FontWeight','Bold','linewidth',2);
set(htitle,'Fontsize',24,'FontName','Helvetica' ...
    ,'FontWeight','Bold');
set(hylabel,'Fontsize',24,'FontName','Helvetica' ...
    ,'FontWeight','Bold');
set(hxlabel,'Fontsize',24,'FontName','Helvetica' ...
    ,'FontWeight','Bold');
set(hlegend,'Fontsize',20,'FontName','Helvetica' ...
    ,'FontWeight','Bold');
```

```
scrsize = get(0,'ScreenSize');
set(gcf,'Color','White','Position' ...
        ,[scrsize(3)/3.8 35 scrsize(3)*0.70 scrsize(4)*0.86]);
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.12 Program: Simulated Diffusion Integrals $\int g(W,t)dW$: Direct Case by Itô Partial Sums

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function intgwtdw
% Book code example for int[g(w,t)dw] on [t0,t] by RNG Simulation:
%   Generation is by summing g(W(i),t(i))dW(i) of even spacing dt
%      for i=0:n, but converted to from index base 0 to base 1:
%      matlab[G(W(i),T(i))DW(i);i=1:N+1]  ...
%             = math[g(W(i),t(i))dW(i);i=0:n].
%   Sample g(w,t) = exp(w-t/2) with exact integral g(w,t) - 1
%      on [0,t].
clc % clear variables, but must come before globals,
%         else clears globals too.
clf % clear figures
fprintf('\nfunction intgwtdw OutPut:')
kfig = 0; % figure counter.
TF= 2.0; T0 = 0; N = 20000;  NI = N+1; dt = (TF-T0)/NI; % Set initial
%                      time grid:  Fixed Delta{t}.
sqrtdt = sqrt(dt); % Set std. Wiener increment time scale.
T(1) = T0;  % set T(1) = T0 in place of t(0) = t0 for base 1 vector.
W(1) = 0.0;  % Set W(1) in place of W(0) = 0 wpo for base 1 vector.
S(1) = 0.0;  % Set integral sum initially.
gv(1) = g(W(1),T(1)); % Set integrand initially.
Err(1) = 0.0; % Set Error initially.
% Begin Sample Path Calculation:
kstate = 1;
randn('state',kstate);  % set randn state for repeatability.
DW = sqrtdt*randn(1,NI); % Generate normal random vector of N+1
%                           samples for dW(t).
for i = 1:NI % Simulated Sample paths by Increment Accumulation:
    T(i+1) = T(i) + dt;
    W(i+1) = W(i) + DW(i);
    gv(i+1) = g(W(i+1),T(i+1));
    S(i+1) = S(i) + gv(i)*DW(i);% integrand g defined in subfunction.
    Err(i+1) = S(i+1) - (gv(i+1) -gv(1)); % CAUTION: FOR KNOWN g HERE!
end
T(NI+1) = TF;  % Correct for final cumulative time rounding errors.
% Begin Plot:
```

```
kfig = kfig + 1;
fprintf('\n\nFigure(%i):  int[g](t) versus t Simulations\n',kfig)
figure(kfig)
plot(T,S,'k-',T,W,'k-.',T,Err,'k--','LineWidth',2);
htitle=title('\int g(W,t)dW(t) for g = exp(W(t)-t/2)');
hylabel=ylabel('\int g(W,t)dW(t), W(t), g(W(t),t) - g(0,0)');
hxlabel=xlabel('t, Time');
hlegend=legend('\int g(W,t)dW(t)','W(t)','Error(t)',0);
haxis = gca;
set(haxis,'Fontsize',20,'FontName','Helvetica','FontWeight','Bold' ...
,'linewidth',2)
set(htitle,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hylabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hxlabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hlegend,'Fontsize',16,'FontName','Helvetica','FontWeight','Bold');
% End Main
function gv = g(W,T)
% Example g(W(t),t) = exp(W(t) - t/2); exact integral = g(W(t),t) - 1.
gv = exp(W - T/2);
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.13 Program: Simulated Diffusion Integrals $\int g(W,t)dW$: Chain Rule

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function intgxtdw
% Book code example for int[g(w,t)dw] on [t0,t] by RNG Simulation:
%   Generation is by summing g(W(i),t(i))dW(i) of even spacing dt for
%      i=0:n, but converted to from index base zero to base one:
%      matlab[G(X(i),T(i))DW(i);i=1:N+1] = math[g(X(i),t(i))dW(i);i=0:n].
%   Chain Rule Form: Int[gdW](t) = G(W,t)-G(0,0) - Int[(g_t+0.5*g_w)(w,t)dt];
%      G_w(w,t) = g(w,t), G_{ww}(w,t) = g_w(w,t).
%   Sample Test Code ofr various g's.
clc % clear variables, but must come before globals,
%     else clears globals too.
clf % clear figures
fprintf('\nfunction intgxtdw OutPut:')
kfig = 0; % figure counter.
TF= 2.0; T0 = 0; N = (TF-T0)*10000;  NI = N+1; dt = (TF-T0)/NI;
% Set initial time grid: Fixed Delta{t}, Scaled to [T0,TF] with N.
sqrtdt = sqrt(dt); % Set standard Wiener increment time scale.
t(1) = T0;  % set T(1) = T0 in place of t(0) = t0 for base 1 vector.
W(1) = 0.0;  % Set W(1) in place of W(0) = 0 wpo for base 1 vector.
sdw(1) = 0.0; sdt(1) = 0.0; % Set integral sum initially.
```

```
gv(1) = g(W(1),t(1)); % Set integrand initially.
ev(1) = 0.0; % Set error initially, if known.
% Begin Sample Path Calculation:
kstate = 1;
randn('state',kstate);  % set randn state for repeatability.
dW = sqrtdt*randn(1,NI); % Generate normal random vector of N+1
%                           samples for dW(t).
for i = 1:NI % Simulated Sample paths by Increment Accumulation:
    t(i+1) = i*dt;
    W(i+1) = W(i) + dW(i);
    X(i+1) = W(i+1); % Set State for this g Example.
    gv(i+1) = g(X(i+1),t(i+1));
    sdw(i+1) = sdw(i) + gv(i)*dW(i);% integrand g in subfunction.
    sdt(i+1) = sdt(i) - gthgw(X(i+1),t(i+1))*dt;% gthgw in subfunction.
    ev(i+1) = sdw(i+1) - exact(X(i+1),t(i+1)) - sdt(i+1);
    % CAUTION: For given g only!
end
t(NI+1) = TF;  % Correct for final cumulative time rounding errors.
% Begin Plot:
kfig = kfig + 1;
fprintf('\n\nFigure(%i):  int[g](t) versus t Simulations\n',kfig)
figure(kfig)
plot(t,sdw,'k-',t,W,'k-.',t,ev,'k--','LineWidth',2);
htitle=title('\int g(X,t)dW(t) for g = exp(X), X = W');
hylabel=ylabel('\int g(X,t)dW(t), X = W(t) and Error(t)');
hxlabel=xlabel('t, Time');
hlegend=legend('\int g(X,t)dW(t)','X = W(t)','Error(t)',0);
haxis = gca;
set(haxis,'Fontsize',20,'FontName','Helvetica','FontWeight','Bold', ...
    'linewidth',2)
set(htitle,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hylabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hxlabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hlegend,'Fontsize',16,'FontName','Helvetica','FontWeight','Bold');
% End Main
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function gv = g(x,t)
% Sample g(X(t),t) only, e.g.,
%1% gv = exp(x-t/2); % x = w.
%2% gv = exp(x); % x = w.
%3% gv = x; % x = w.
gv = exp(x);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function exactv = exact(x,t)
% Sample g(X(t),t) exact integrals only, e.g.,
%1% exactv = exp(x-t/2) - 1; % i.e., G(w,t)-G(0,0), x=w, G(w,t)=exp(w-t/2).
```

```
%2% exactv = exp(x) - 1; % i.e., G(w,t)-G(0,0), x=w, G(w,t)=exp(w).
%3% exactv = 0.5*(x^2-t);  % i.e., G(w,t)-G(0,0), x=w, G(w,t)=0.5*(w^2-t).
exactv = exp(x) - 1; % i.e., G(w,t)-G(0,0), x=w, G(w,t)=exp(w).
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function gthgwv = gthgw(x,t)
% Reg. Correction Int. of (G_t+0.5*G_{ww})(X(t),t), G_w = g.
%1% gthgwv = 0; % i.e., g=exp(x-t/2)=G, G_t=-0.5*G, G_{ww}=G.
%2% gthgwv = 0.5*exp(x); % i.e., G=g=exp(w), G_t=0, G_{ww}=g_w=exp(w).
%3% gthgw 0; % i.e., g=x=w, G=0.5*(w^2-t), G_t=-0.5, G_{ww}=g_w=1;
gthgwv = 0.5*exp(x); % i.e., G=g=exp(w), G_t=0, G_{ww}=g_w=exp(w).
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% End Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.14 Program: Simulated Linear Jump-Diffusion Sample Paths

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function linjumpdiff03fig1
% Book Illustration for Linear (Geometric)Jump Diffusion SDE RNG
%   Simulation with constant coefficients for t in [0,1]
%   with sample variation:
%        DX(t) = X(t)*(mu*Dt + sig*DW(t) + nu*DP(t),
%            X(0) = x0.
%   Or log-state:
%        DY(t) = (mu-sig^2/2)*Dt + sig*DW(t) + log(1+nu)*DP(t),
%       Y(0) = log(x0).
%   Generation is by summing Wiener increments DW of even spacing Dt
%     with Poisson jump increment added at correct time increment.
%   Sufficiently SMALL increments assumed, so zero-one jump law is
%       appropriate.
%   Allows Separate Driver Input and Special Jump
%       or Diffusion Handling.
clc % clear variables, but must come before globals,
    %    else clears globals too.
clf % clear figures
fprintf('\nfunction linjumpdiff03fig1 OutPut:');
%%% Initialize input to jdsimulator
N = 1000; T = 1.0; % Set initial time grid: Fixed Delta{t}.
mu = 0.5; sig = 0.10; nu = -0.10; lambda = 3.0;
% set constant parameters.
%
jdsimulator(mu,sig,nu,lambda,N,T);
%
% END INPUT FOR JUMP-DIFFUSION SIMULATOR.
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function jdsimulator(mu,sig,nu,lambda,N,T)
idiff = 1; ijump = 1;
if sig == 0, idiff = 0; end
if nu == 0, ijump = 0; end
kfig = 0; % figure counter.
NI = N+1;  Dt = T/NI;
iv = 2; % iv=1 for *(1+/-sqrt(Var[X])) or iv=2 for *exp(+/-sqrt(Var[Y]).
sqrtdt = sqrt(Dt); % Set standard Wiener increment moments.
muddt = (mu - sig^2/2)*Dt; % Get Ito diffusion corrected drift term.
lognu = log(1 + nu); % Get log of relative jump term amplitude.
% Begin Sample Path Calculation:
t = 0:Dt:T; kstates = 4; x0 = 1.0;
for kstate = 1:kstates % Test Multiple Simulated Sample Paths:
    if idiff == 1
      randn('state',kstate-1); % Set initial normal state
                               %    for repeatability.
      DW = sqrtdt*randn(NI,1); % Generate normal random vector
                               %    of N samples for DW(t).
      WS(1) = 0.0;  % Set W(0) = 0 wpo using MATLAB base 1 vector.
    end
    if ijump == 1
      rand('state',kstate-1); % Set initial uniform state
                              %    for repeatability.
      DU = rand(NI,1); % Generate Uniform random vector of N
                       %    DP(t) samples.
      PS(1) = 0.0;  % Set P(0) = 0 wpo using MATLAB base 1 vector.
      ldt = lambda*Dt; % one jump prob.
      ul = (1-ldt)/2;  ur = (1+ldt)/2; % Set centered jump
                                       %   probability thresholds,
    end
    YS(1) = 0.0; XS(1,kstate) = x0; % Set initial exponent and state.
    % using centered part of uniform distribution
    %     to avoid open end point bias.
    for i = 1:NI % Simulated Sample paths by Increment Accumulation:
        YS(i+1) = YS(i) + muddt;
        if idiff == 1, YS(i+1) = YS(i+1)+ sig*DW(i); end
        if ijump == 1
          if DU(i) <= ur && DU(i) >= ul % Get jump if prob. in [ul,ur]:
             YS(i+1) = YS(i+1) + lognu;
          end
        end
        XS(i+1,kstate) = x0*exp(YS(i+1));% Invert exponent to get state.
    end
end
% Compute Mean State Path and +/- One Std. Deviation:
```

```
XM(1) = x0; XT(1) = x0; XB(1) = x0;
muxexp = mu + lambda*nu;
if iv == 1, sigxexp = sig^2 + lambda*nu^2; end
if iv == 2, sigyexp = sig^2 + lambda*(log(1+nu))^2; end
for i = 1:NI
    XM(i+1) = x0*exp(muxexp*t(i+1));
    if iv == 1
        V = sqrt(exp(sigxexp*t(i+1)) - 1);
        XT(i+1) = XM(i+1)*(1 + V);
        XB(i+1) = XM(i+1)*(1 - V);
    end
    if iv == 2
        V = exp(sqrt(sigyexp*t(i+1)));
        XT(i+1) = XM(i+1)*V;
        XB(i+1) = XM(i+1)/V;
    end
end
% Begin Plot:
kfig = kfig + 1;
kjd = 4 - 2*idiff - ijump;
NP = N + 2;
stitle = {'Linear Jump-Diffusion Simulations' ...
        ,'Linear Diffusion Simulations' ...
        ,'Linear Jump Simulations'};
sylabel = {'X(t), Jump-Diffusion State','X(t), Diffusion State' ...
        ,'X(t), Jump State'};
fprintf('\n\nFigure(%i):  Linear Jump-Diffusion Simulations\n',kfig)
figure(kfig)
plot(t,XS(1:NP,1),'k-' ...
    ,t,XS(1:NP,2),'k-' ...
    ,t,XS(1:NP,3),'k-' ...
    ,t,XS(1:NP,4),'k-' ...
    ,t,XM(1:NP),'k--' ...
    ,t,XT(1:NP),'k-.' ...
    ,t,XB(1:NP),'k-.','LineWidth',2);
htitle=title(stitle(kjd));
hylabel=ylabel(sylabel(kjd));
hxlabel=xlabel('t, Time');
if iv == 1
hlegend=legend('X(t) Sample 1','X(t) Sample 2','X(t) Sample 3'...
    ,'X(t) Sample 4','E[X](t)','(E[X]*(1+V))(t)','(E[X]*(1-V))(t)',0);
end
if iv == 2
hlegend=legend('X(t) Sample 1','X(t) Sample 2','X(t) Sample 3'...
    ,'X(t) Sample 4','E[X](t)','(E[X]*V)(t)','(E[X]/V)(t)',2);
end
```

```
haxis = gca;
set(haxis,'Fontsize',20,'FontName','Helvetica','FontWeight','Bold'...
    ,'linewidth',2)
set(htitle,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hylabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hxlabel,'Fontsize',24,'FontName','Helvetica','FontWeight','Bold')
set(hlegend,'Fontsize',16,'FontName','Helvetica','FontWeight','Bold');
%
% End JDSimulator Code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.15 Program: Simulated Linear Mark-Jump-Diffusion Sample Paths

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function linmarkjumpdiff06fig1
% Book Illustration for Linear Distributed-Jump Diffusion SDE RNG
%   Simulation with variable coefficients for t in [0,1]
%   with sample variation:
%        DX(t) = X(t)*(mu(t)*Dt + sig(t)*DW(t) + nu(Q)*DP(t),
%            X(0) = x0.
%   Or log-state:
%        DY(t) = (mu(t)-sig^2(t)/2)*Dt + sig(t)*DW(t) + Q*DP(t),
%       Y(0) = log(x0) and Q = ln(1+nu(Q)).
%   Generation is by summing Wiener increments DW of even spacing Dt
%     with Poisson jump increment added at correct time increment.
%   Sufficiently SMALL increments assumed, so zero-one jump law is
%       appropriate.
%   For demonstration purposes, Q will be assumed to be
%     (qdist =1) UNIFORMLY distributed on (qparm1,qparm2)=(a,b)
%     OR
%     (qdist=2) NORMALLY distributed with (qparm1,qparm2)=(muj,sj2).
%   Allows Separate Driver Input and Special Jump
%       or Diffusion Handling.
clc % clear variables, but must come before globals,
    %    else clears globals too.
clf % clear figures
fprintf('\nfunction linjumpdiff06fig1 OutPut:');
%%% Initialize input to jdsimulator with sample parameters:
N = 1000; t0 = 0; T = 2.0; % Set initial time grid: Fixed Delta{t}.
idiff = 1; ijump = 1; x0 = 1.0;
qdist = 1; a = -2; b = +1; qparm1 = a; qparm2 = b; %e.g., Uniform
%OR E.G., Normal distribution:
%qdist = 2; muj = 0.28; sj2 = +0.15; qparm1 = muj; qparm2 = sj2;
% set constant parameters.
```

```
fprintf('\n  N=%i; x0=%6.3f; t0=%6.3f; T=%6.3f;',N,x0,t0,T);
fprintf('\n  qdist=%i*; qparm1=%6.3f; qparm2=%6.3f;'...
    ,qdist,qparm1,qparm2);
fprintf('\n  * qdist=1 for uniform Q-distribution.');
fprintf('\n  * qdist=2 for normal Q-distribution.');
%
jdsimulator(idiff,ijump,qdist,qparm1,qparm2,N,x0,t0,T);
%
% END INPUT FOR JUMP-DIFFUSION SIMULATOR.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function jdsimulator(idiff,ijump,qdist,qparm1,qparm2,N,x0,t0,T)
kfig = 0; % figure counter.
dt = (T-t0)/N; % Get number of intervals/samples and time step.
kjd = 4 - 2*idiff - ijump;
NP = N + 1; % Number of plot points = number of time steps + 1.
sqrtdt = sqrt(dt); % Set standard Wiener increment moments.
tv = t0:dt:T; % Compute time vector;
sv = zeros(size(tv)); ldtv = zeros(size(tv));
muv = mu(tv); % Get time-dependent coefficient vectors
if idiff == 1, sv = sigma(tv); end
if ijump == 1, ldtv = dt*lambda(tv); end
muddt = (muv - sv.^2/2)*dt; % Get diffusion corrected drift term.
if qdist == 1 % Average nu(Q)=exp(Q)-1 for UNIFORM Q-Dist.
    numean = (exp(qparm2)-exp(qparm1))/(qparm2-qparm1)-1;
elseif qdist == 2 % Average nu(Q)=exp(Q)-1 for NORMAL Q-Dist.
    numean = exp(qparm1-qparm2/2)-1;
end
% Compute Theoretical Mean State Path
%   E[X(t+dt)] = X(t)*exp(E[dX(t)|X(t)=x]/x), x0 > 0:
XM = zeros(1,NP); % preallocate mean state.
XM(1) = x0;
for i = 1:N
    XM(i+1) = XM(i)*exp(muv(i)*dt+numean*ldtv(i));
end
kstates = 4; kv = [1,5,9,10]; % selected random states.
XS = zeros(NP,kstates); % preallocate global state array.
% Begin Sample Path Calculation:
for k = 1:kstates % Test Multiple Simulated Sample Paths:
    if idiff == 1
        randn('state',kv(k)); % Set initial normal state
                              %    for repeatability.
        DW = sqrtdt*randn(1,N); % Generate normal random vector
                                 %    of N samples for DW(t).
    end
    if ijump == 1
        rand('state',kv(k)); % Set initial uniform state
```

```
                              %    for repeatability.
        DU = rand(1,N); % Generate Uniform random vector DP(t)
    if qdist == 1 %Generate Uniform random mark vector Q samples.
        Q = qparm1+(qparm2-qparm1)*rand(1,N);
    elseif qdist == 2 %Generate Normal random mark vector Q samples.
     sj = sqrt(qparm2); Q = qparm1+sj*randn(1,N);
    end
    ul = (1-ldtv)/2;  ur = 1-ul; % Set vector centered jump
                                        %   probability thresholds,
    end
    YS = zeros(1,N+1); % preallocate state exponent for efficiency.
    XS(1,k) = x0; % Set kth initial state.
    for i = 1:N % Simulated Sample paths by Increment Accumulation:
        YS(i+1) = YS(i) + muddt(i); % Add dY-drift:
        % Add diffusion increment:
        if idiff == 1, YS(i+1) = YS(i+1)+ sv(i)*DW(i); end
        % Using centered part of uniform distribution, with
        %     acceptance-rejection, to avoid open end point bias:
        if ijump == 1
            if DU(i) <= ur(i) && DU(i) >= ul(i) % Jump if in [ul,ur]
               YS(i+1) = YS(i+1) + Q(i); % If jump, +Y-jump amplitude.
            end % Else no jump, so do not add anything.
        end
        XS(i+1,k) = x0*exp(YS(i+1));% Invert exponent to get state.
    end % i
end % k
% Sample Mean State:
XSM = zeros(1,NP);
for i = 1:NP
    XSM(i) = mean(XS(i,:));
end
% Begin Plot:
scrsize = get(0,'ScreenSize');
ss = 5.2; dss = 0.2; ssmin = 3.0;
kfig = kfig + 1;
stitle = {'Linear Mark-Jump-Diffusion Simulations' ...
        ,'Linear Diffusion Simulations' ...
        ,'Linear Mark-Jump Simulations'};
sylabel = {'X(t), Jump-Diffusion State','X(t), Diffusion State' ...
        ,'X(t), Jump State'};
slegend = {'X(t), State 1','X(t), State 5' ...
        ,'X(t), State 9','X(t), State 10'...
        ,'XM(t), th. Mean=E[X(t)]','XSM(t), Sample Mean'};
fprintf('\n\nFigure(%i):  Linear Jump-Diffusion Simulations\n',kfig)
figure(kfig)
plot(tv,XS(1:NP,1),'k+-' ...
```

```
    ,tv,XS(1:NP,2),'k:' ...
    ,tv,XS(1:NP,3),'k--' ...
    ,tv,XS(1:NP,4),'k-.' ...
    ,tv,XM(1:NP),'k-',tv,XSM(1:NP),'b.-' ...
    ,'LineWidth',2); % Add for more States?
title(stitle(kjd),'Fontsize',24,'FontName','Helvetica'...
    ,'FontWeight','Bold');
ylabel(sylabel(kjd),'Fontsize',24,'FontName','Helvetica'...
    ,'FontWeight','Bold');
xlabel('t, Time','Fontsize',24,'FontName','Helvetica'...
    ,'FontWeight','Bold');
legend(slegend,'Location','Best','Fontsize',16,...
    'FontName','Helvetica','FontWeight','Bold');
haxis = gca;
set(haxis,'Fontsize',20,'FontName','Helvetica'...
    ,'FontWeight','Bold','linewidth',2);
    ss = max(ss - dss,ssmin);
set(gcf,'Color','White','Position' ...
   ,[scrsize(3)/ss 60 scrsize(3)*0.60 scrsize(4)*0.80]); %[l,b,w,h]
%
% End JDSimulator Code
%
% linear Time-Dependent SDE Coefficient Functions:
%   (Change with application; functions must be vectorizable,
%       using vector element dot operations or vector functions.)
%
function v = mu(t)
% drift coefficient example, change with applications:
v = 0.1*sin(t);
% end mu(t)
%
function v = sigma(t)
% drift coefficient example, change with applications:
v = 1.5*exp(-0.01*t);
% end sigma(t)
%
function v = lambda(t)
% drift coefficient example, change with applications:
v = 3.0*exp(-t.*t);
% end lambda(t)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.16 Program: Euler-Maruyama Simulations for Linear Diffusion SDE

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function sdeeulersim
% Euler-Maruyama Simulations: Linear, Time-Dep. Coeff. SDE,
%   dX(t) = X(t)(mu(t)dt+sigma(t)dW(t)), X(0) = x0, t0 < t < tf,
% Given  Initial data: x0, t0, tf, Nt; functions: f, g
clc
x0 = 1; t0 = 0; tf = 5; Nt = 2^10;
randn('state',8);
DT = tf/Nt; sqrtdt = sqrt(DT);
Xeul(1) = x0; Xexact(1) = x0; Xdiff(1) = Xeul(1) - Xexact(1);
t = [t0:DT:tf];
DW = randn(1,Nt)*sqrtdt;
W = cumsum(DW); % Note: omits initial zero value; count is off by 1;
%
for k = 1:Nt % Exact formula to fine precision for exact consistency:
    Xexact(k+1) = xexact(x0,t(k+1),W(k));
end
% Lumped coarse sample from fine sample:
L = 2^3; NL = Nt/L; KL = [0:L:Nt]; DTL = L*DT; tL = [t0:DTL:tf];
fprintf('(N_t,NL)=(%i,%i); Size(t,KL,tL)=[(%i,%i);(%i,%i);(%i,%i)];'...
    ,Nt,NL,size(t),size(KL),size(tL));
for k = 1:NL % Euler-Maruyama formula to coarse precision:
    DWL = sum(DW(1,KL(k)+1:KL(k+1)));
    Xeul(k+1) = Xeul(k) + f(Xeul(k),tL(k))*DTL+g(Xeul(k),tL(k))*DWL;
    Xdiff(k+1) = Xeul(k+1) - Xexact(KL(k+1));
end
%
scrsize = get(0,'ScreenSize');
ss = [3.0,2.8,2.6,2.4,2.2,2.0];
%
nfig = 1;
figure(nfig);
plot(tL,Xeul,'k--','linewidth',3); hold on
plot(t,Xexact,'k-','linewidth',3); hold off
axis([t0 tf 0 max(max(Xeul),max(Xexact))]);
title('Euler-Maruyama and Exact Linear SDE Simulations'...
    ,'Fontsize',36,'FontName','Helvetica','FontWeight','Bold');
xlabel('t, Time'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
ylabel('X(t), State'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
legend('Xeul(t): Euler','Xexact(t): Exact','Location','Best');
set(gca,'Fontsize',28,'FontName','Helvetica','FontWeight','Bold'...
```

```
,'linewidth',3);
set(gcf,'Color','White','Position' ...
    ,[scrsize(3)/ss(nfig) 70 scrsize(3)*0.60 scrsize(4)*0.80]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
Xdiffmax = max(abs(Xdiff));
fprintf('\nMaximal Euler-Exact Absolute Error:');
fprintf('\n     max(abs(Xeul(TL)-Xexact(TL)))=%8.2e=%8.2e*DTL;\n'...
    ,Xdiffmax,Xdiffmax/DTL);
% (N_t,NL) = (1024,128); Size(t,KL,tL) = [(1,1025);(1,129);(1,129)];
% Maximal Euler-Exact Abs. Error:
%      max(abs(Xeul(TL)-Xexact(TL))) = 1.31e+00 = 3.36e+01*DTL;
%
nfig = nfig+1;
figure(nfig);
plot(tL,Xdiff,'k-','linewidth',3);
axis tight;
title('Euler and Exact Linear SDE Simulations Error'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
xlabel('t, Time'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
ylabel('Xeul(t)-Xexact(t), Error'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
set(gca,'Fontsize',28,'FontName','Helvetica','FontWeight','Bold'...
,'linewidth',3);
set(gcf,'Color','White','Position' ...
    ,[scrsize(3)/ss(nfig) 70 scrsize(3)*0.60 scrsize(4)*0.80]);
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y = f(x,t)
mu = 1/(1+0.5*t)^2;
y = mu*x;
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y = g(x,t)
sig = 0.5;
y = sig*x;
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y = xexact(x0,t,w)
% exact solution if available for general linear SDE:
mubar = 2-2/(1+0.5*t); sig = 0.5; sig2bar = sig^2*t/2;
y = x0*exp(mubar-sig2bar + sig*w);
%
% end sdeeulersim.m
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.17 Program: Milstein Simulations for Linear Diffusion SDE

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function sdemilsteinsim
% Milstein SDE Simulations: Linear, Time-Dep. Coeff. SDE
%   dX(t) = X(t)(mu(t)dt+sigma(t)dW(t)), X(0) = x0, t0 < t < tf,
% Given  Initial data: x0, t0, tf, Nt; functions: f, g
clc
x0 = 1; t0 = 0; tf = 5; Nt = 2^12;
randn('state',8);
DT = tf/Nt; sqrtdt = sqrt(DT);
Xmil(1) = x0; Xeul(1) = x0; Xexact(1) = x0;
Xdiff(1) = Xmil(1) - Xexact(1);
Xmileul(1) = Xmil(1) - Xeul(1);
t = [t0:DT:tf];
DW = randn(1,Nt)*sqrtdt;
W = cumsum(DW); % Note: omits initial zero value; count if off by 1;
for k = 1:Nt % Exact formula to fine precision for exact consistency:
    Xexact(k+1) = xexact(x0,t(k+1),W(k));
end
% Lumped coarse sample from fine sample:
L = 2^3;
NL = Nt/L; KL = [0:L:Nt]; DTL = L*DT; tL = [t0:DTL:tf];
fprintf('(N_t,NL)=(%i,%i); Size(t,KL,tL)=[(%i,%i);(%i,%i);(%i,%i)];'...
    ,Nt,NL,size(t),size(KL),size(tL));
for k = 1:NL % Milstein and Euler formulas to coarse precision:
    DWL = sum(DW(1,KL(k)+1:KL(k+1)));
    Xmil(k+1)=Xmil(k)+f(Xmil(k),tL(k))*DTL+g(Xmil(k),tL(k))*DWL...
        +0.5*g(Xmil(k),tL(k))*gx(Xmil(k),tL(k))*(DWL^2-DTL);
    Xeul(k+1)=Xeul(k)+f(Xeul(k),tL(k))*DTL+g(Xeul(k),tL(k))*DWL;
    Xdiff(k+1) = Xmil(k+1) - Xexact(KL(k+1));
    Xmileul(k+1) = Xmil(k+1) - Xeul(k+1);
end
%
scrsize = get(0,'ScreenSize');
ss = [3.0,2.8,2.6,2.4,2.2,2.0];
%
nfig = 1;
figure(nfig);
plot(tL,Xmil,'k--','linewidth',3); hold on
% plot(tL,Xeul,'k:','linewidth',3); hold on
plot(t,Xexact,'k-','linewidth',3); hold off
axis([t0 tf 0 max(max(max(Xmil),max(Xeul)),max(Xexact))]);
title('Milstein and Exact Linear SDE Simulations'...
    ,'Fontsize',36,'FontName','Helvetica','FontWeight','Bold');
```

```
xlabel('t, Time'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
ylabel('X(t), State'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
hlegend = legend('Xmil(t): Milstein','Xexact: Exact'...
    ,'Location','Best');
set(hlegend,'Fontsize',32,'FontName','Helvetica'...
    ,'FontWeight','Bold');
set(gca,'Fontsize',28,'FontName','Helvetica'
    ,'FontWeight','Bold','linewidth',3);
set(gcf,'Color','White','Position' ...
    ,[scrsize(3)/ss(nfig) 70 scrsize(3)*0.60 scrsize(4)*0.80]);
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
Xdiffmax = max(abs(Xdiff));
fprintf('\nMaximal Milstein-Exact Absolute Error:');
fprintf('\n     max(abs(Xmil(TL)-Xexact(TL)))=%8.2e=%8.2e*DTL;\n'...
    ,Xdiffmax,Xdiffmax/DTL);
% (N_t,NL) = (1024,128); Size(t,KL,tL) = [(1,1025);(1,129);(1,129)];
% Maximal Milstein-Exact Absolute Error:
%      max(abs(Xmil(TL)-Xexact(TL))) = 1.23e+00 = 3.16e+01*DTL;
% Maximal Milstein-Euler Absolute Error:
%     max(abs(Xmil(TL)-Xeul(TL))) = 9.54e-01 = 2.44e+01*DTL;
%
nfig=nfig+1;
figure(nfig);
plot(tL,Xdiff,'k-','linewidth',3);
axis tight;
title('Milstein and Exact SDE Simulations Error'...
    ,'Fontsize',36,'FontName','Helvetica','FontWeight','Bold');
xlabel('t, Time'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
ylabel('Xmil(t)-Xexact(t), Error'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
set(gca,'Fontsize',28,'FontName','Helvetica','FontWeight','Bold'...
    ,'linewidth',3);
set(gcf,'Color','White','Position' ...
    ,[scrsize(3)/ss(nfig) 70 scrsize(3)*0.60 scrsize(4)*0.80]);
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
Xmileulmax = max(abs(Xmileul));
fprintf('\nMaximal Milstein-Euler Absolute Error:');
fprintf('\n     max(abs(Xmil(TL)-Xeul(TL))) = %8.2e = %8.2e*DTL;\n'...
    ,Xmileulmax,Xmileulmax/DTL);
```

```
%
nfig=nfig+1;
figure(nfig);
plot(tL,Xmileul,'k-','linewidth',3);
axis tight;
title('Milstein and Euler SDE Simulations Difference'...
    ,'Fontsize',36,'FontName','Helvetica','FontWeight','Bold');
xlabel('t, Time'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
ylabel('Xmil(t)-Xeul(t), Difference'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
set(gca,'Fontsize',28,'FontName','Helvetica','FontWeight','Bold'...
,'linewidth',3);
set(gcf,'Color','White','Position' ...
    ,[scrsize(3)/ss(nfig) 70 scrsize(3)*0.60 scrsize(4)*0.80]);
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y = f(x,t)
mu = 1/(1+0.5*t)^2;
y = mu*x;
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y = g(x,t)
sig = 0.5;
y = sig*x;
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y = gx(x,t)
sig = 0.5;
y = sig;
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y = xexact(x0,t,w)
% exact solution if available for general linear SDE:
mubar = 2-2/(1+0.5*t); sig = 0.5; sig2bar = sig^2*t/2;
y = x0*exp(mubar-sig2bar + sig*w);
%
% end sdemilsteinsim.m
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.18 Program: Monte Carlo Simulation Comparing Uniform and Normal Errors

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function mcm0unifnorm
```

```
% Book Chapter 10 Sims, Section Monte Carlo;
%   compare error for uniform and normal on
%   int(exp(-x^2/2)/sqrt(2*pi),x=-R..R)
clc
clear
%
global R V
%
fprintf('Compare Uniform and Normal Monte Carlos:\n');
n = 100; srtn = sqrt(n);
R=5; V = 2*R;
fprintf('\nn=%i; R=%6.4f; V=%6.4f;\n',n,R,V);
% erfc(x) = 2\sqrt(pi)*int(exp(-t^2),t=x..infty);
% normcdf(x)=0.5*erfc(-x/sqrt(2));
exact = 0.5*(erfc(-R/sqrt(2))-erfc(+R/sqrt(2)));
sig2uexact = 2.5/sqrt(pi)*(erf(R)-erf(-R))-exact^2;
sig2nexact = exact*(1-exact);
fprintf('\nexact integral = %10.8f;',exact);
fprintf('\nsig2unifexact = %9.4e; sigunifexact = %9.4e;'...
    ,sig2uexact,sqrt(sig2uexact));
fprintf('\nsig2normexact = %9.4e; signormexact = %9.4e;\n'...
    ,sig2nexact,sqrt(sig2nexact));
U = -R+V*rand(1,n);
X = randn(1,n);
for i=1:n
    fuv(i)=fu(U(i));
    fnv(i)=fn(X(i));
end
% Monte Carlo estimators:
sun = mean(fuv);
snn = mean(fnv);
fprintf('\nsunifn=%10.8f; snormn=%10.8f;',sun,snn);
fprintf('\nsunifnabserror=%9.4e%%; snormnabserror=%9.4e%%;'...
    ,sun-exact,snn-exact-1);
fprintf('\nsunifnrelerror=%9.4e%%; snormnrelerror=%9.4e%%;\n'...
    ,100*(sun/exact-1),100*(snn/exact-1));
% Monte Carlo variance estimators:
sig2un = var(fuv);  % MATLAB var(x); gives unbiased variance
sig2nn = var(fnv);
fprintf('\nsig2unifn=%9.4e; sig2normn=%9.4e;',sig2un,sig2nn);
fprintf('\nsig2unifnabserror=%9.4e%%; sig2normnabserror=%9.4e%%;'...
    ,sig2un-sig2uexact,sig2nn-sig2nexact);
fprintf('\nsig2unifnrelerror=%9.4e%%; sig2normnrelerror=%9.4e%%;\n'...
    ,100*(sig2un/sig2uexact-1),100*(sig2nn/sig2nexact-1));
% std. errors:
seunifexact = sqrt(sig2uexact)/srtn;
```

```
senormexact = sqrt(sig2nexact)/srtn;
seunifn = sqrt(sig2un)/srtn;
senormn = sqrt(sig2nn)/srtn;
fprintf('\nstderrunifexact=%9.4e; stderrnormexact=%9.4e;'...
    ,sqrt(sig2uexact)/srtn,sqrt(sig2nexact)/srtn);
fprintf('\nstderrunifn=%9.4e; stderrnormn=%9.4e;'...
    ,sqrt(sig2un)/srtn,sqrt(sig2nn)/srtn);
fprintf('\nstderrunifndiff=%9.4e; stderrnormndiff=%9.4e;\n'...
    ,seunifn-seunifexact,senormn-senormexact);
%
%%%%%
function y = fu(x)
global R V
y = V*exp(-x.*x/2)/sqrt(2*pi);
%%%
function y = fn(x)
global R V
y = 1;
if abs(y)>R, y=0; end
%%%%%
%
% end mcm0unifnorm.m
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.19 Program: Monte Carlo Simulation Comparing Uniform and Normal Errors

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function mcm1test
% mcm1test:   Monte Carlo Method, nx = 1 dim, uniform dist,
%     I = int(F(x),x=a..b), F(x) = sqrt(1-x^2), -1 <= a < b <= +1;
%     technically, f(x) = (b-a)F(x) = (b-a)*sqrt(1-x^2) to account for
%     uniform density phi(x) = 1/(b-a) on [a,b], so I = meanf.
%
clc; clear
%
fprintf('Monte Carlo Test of 1-dim Uniform Dist. on (a,b)');
fprintf('\n   with F(x)=sqrt(1-x^2) and f(x) = (b-a)F(x):\n');
a = 0; b = +1; % -1 <= a < b <= +1;
% integral of f(x) = sqrt(1-x^2); on [a,b]:
IntExact = 0.5*(asin(b)-asin(a))+0.5*(b*sqrt(1-b^2)-a*sqrt(1-a^2));
MufExact = IntExact;
Sigf = sqrt((b-a)^2*(1-(b^2+a*b+a^2)/3)-MufExact^2);
fprintf('\nk        n muhatn mufExact   sighatn     Sigf    stderrn  AbsErrorf\n');
```

```
kmax = 7;
for k = 1:kmax
   rand('state',0); % set state or seed
   n(k) = 10^k; % sample size, k = log10(n(k)) ;
   x = a+(b-a)*rand(n(k),1); % get n(k) X 1 random sample on (a,b);
   f = (b-a)*sqrt(1-x.^2); % vectorized f;
   meanf(k) = mean(f); % E[f(X)];
   sigf(k) = std(f); % sqrt(sigmaf^2), sigmaf^2 = unbiased variance of f;
   sigdrn(k) = sigf(k)/sqrt(n(k));
   error(k) = abs(meanf(k)-MufExact);
   fprintf('%1i %8i %6.4f   %6.4f %9.3e %9.3e  %9.3e %9.3e\n'...
       ,k,n(k),meanf(k),MufExact,sigf(k),Sigf,sigdrn( k),error(k))
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
scrsize = get(0,'ScreenSize');
ss = [3.0,2.8,2.6,2.4,2.2,2.0];
%
nfig = 1;
figure(nfig);
kv = [1:kmax];
plot(kv,meanf,'k-o','linewidth',3,'MarkerSize',12); hold on
plot(kv,sigf,'k-x','linewidth',3,'MarkerSize',12); hold off
axis([min(kv) max(kv) 0 1]);
title('Monte Carlo Results, Uniform Dist., F(x) = sqrt(1-x^2)'...
    ,'Fontsize',36,'FontName','Helvetica','FontWeight','Bold');
xlabel('log(n), Log_{10} Sample Size'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
ylabel('f-Moments \mu_n, \sigma_n'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
legend('\mu_n, Mean-est.','\sigma_n, StdDev-est.','Location','Best');
set(gca,'Fontsize',28,'FontName','Helvetica','FontWeight','Bold','linewidth',3);
set(gcf,'Color','White','Position' ...
    ,[scrsize(3)/ss(nfig) 70 scrsize(3)*0.60 scrsize(4)*0.80]);
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
nfig = nfig+1;
figure(nfig);
kv = [1:kmax];
plot(kv,log10(sigdrn),'k-o','linewidth',3,'MarkerSize',12); hold on
plot(kv,log10(error),'k-x','linewidth',3,'MarkerSize',12); hold off
ymin = min(min(log10(sigdrn)),min(log10(error)));
ymax = max(max(log10(sigdrn)),max(log10(error)));
axis tight; %axis([min(kv) max(kv) ymin ymax]);
title('Monte Carlo Errors, Uniform Dist., F(x) = sqrt(1-x^2)'...
```

```
    ,'Fontsize',36,'FontName','Helvetica','FontWeight','Bold');
xlabel('log(n), Log_{10} Sample Size'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
ylabel('f-Errors log(StdError_n), log(AbsError_n)'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
legend('log_{10}(StdError_n)','log_{10}(AbsError_n)','Location','Best');
set(gca,'Fontsize',28,'FontName','Helvetica','FontWeight','Bold','linewidth',3);
set(gcf,'Color','White','Position' ...
    ,[scrsize(3)/ss(nfig) 70 scrsize(3)*0.60 scrsize(4)*0.80]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% end mcm1test.m
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.20 Program: Monte Carlo Acceptance-Rejection Technique

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function mcm2acceptreject
% mcm2acceptreject: Monte Carlo Method, nx = 1 dim, normal dist.,
%     I = int(F(x),x=a..b), F(x) = exp(-x^2/2)/sqrt(2pi), -1 <= a < b <= +1;
%     technically, f(x) = I_{x in [a,b]} to account for truncated integral,
%     so I = meanf.
%
clc; clear
%
fprintf('Monte Carlo and Finite Difference Comparison:');
fprintf('\n   including Acceptance-Rejection Technique Application,');
fprintf('\n   with Normal Dist. on (a,b)');
fprintf('\n   and with int(F(x),x=a..b), F(x) = exp(-x^2/2)/sqrt(2pi);\n');
%
a = -2; b = 2; % limits of integration;
nfd = 100; % number of finite difference steps;
kmax = 7;
nmc = 10^kmax; % select Monte Carlo random sample size;
F = inline ('exp(-x.*x/2)./sqrt(2*pi)','x');  % x in [a,b]
%   Thus, relative to the normal density, f(x)={1, x in [a,b]; 0, else};
h = (b - a)/nfd; % step size;
% Trapezoid Rule (see also MATLAB trapz(x,y) built-in function):
trap = (F(a)+F(b))/2;
for i = 1:nfd-1,
    trap = trap+F(a+i*h);
end
trap = h*trap;
```

```
fprintf('\n%3i-point Trapezoidal Rule: I(-1,1) = %.6f\n',nfd+1,trap);
% Simpson's (1/3) Rule:
simp = F(a)+F(b);
for i = 1:nfd-1
    if mod(i,2)
        simp =simp+ 4*F(a + i*h);
    else
        simp=simp+2*F(a + i*h);
    end
end
simp = h*simp/3;
fprintf('\n%3i-point Simpson''s rule: I(-1,1) = %.6f\n',nfd+1,simp);
% MATLAB quad built-in function (adaptive Simpson's rule, default 1.e-6 accuracy):
tol = 1.e-9;
quadfn = quad(F,a,b,tol);
fprintf('\n%7.1e-accurate quad: = %.6f\n',tol,quadfn);
% Direct von Neumann Acceptance-Rejection Technique:
fprintf('\nMonte Carlo results by von Neumann''s Acceptance-Rejection technique:\n');
fprintf('\n k          n    muhatn    stderrn\n');
nac = 0;
x = randn(nmc,1); % MATLAB nmc X 1 normal distribution;
for n = 1:nmc
    if (x(n) >= a) & (x(n) <= b)
        nac = nac + 1; % counts accepted points;
    end
    if (n==10)|(n==100)|(n==1000)|(n==10000)|(n==100000)|(n==1000000)|(n==nmc)
        k = log10(n);
        kv(k) = k;
        muhatn(k) = nac/n;
        stderrn(k) = sqrt(muhatn(k)*(1-muhatn(k))/(n-1));
        fprintf ('%2i   %8i  %8.6f  %9.3e\n',k,n,muhatn(k),stderrn(k));
    end
end
fprintf('\n 101-pt. trap: %8.6f  %9.3e*',trap,abs(trap-quadfn));
fprintf('\n 101-pt. simp: %8.6f  %9.3e*',simp,abs(simp-quadfn));
fprintf('\n    accurate:  %8.6f  %9.3e*',quadfn,abs(quadfn-quadfn));
fprintf('\n * Absolute Errors\n');
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
scrsize = get(0,'ScreenSize');
ss = [3.0,2.8,2.6,2.4,2.2,2.0];
%
nfig = 1;
figure(nfig);
kv = [1:kmax];
plot(kv,muhatn,'k-o','linewidth',3,'MarkerSize',12); hold on
```

```
plot(kv,stderrn*10,'k-x','linewidth',3,'MarkerSize',12); hold off
axis([min(kv) max(kv) 0 1]);
title('Monte Carlo Results, Normal Dist., F(x) = \phi_n(x) on [a,b]'...
    ,'Fontsize',36,'FontName','Helvetica','FontWeight','Bold');
xlabel('log(n), Log_{10} Sample Size'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
ylabel('Moments \mu_n, 10*std-err_n'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
legend('\mu_n, Mean-est.','10*std-err_n','Location','Best');
set(gca,'Fontsize',28,'FontName','Helvetica','FontWeight','Bold','linewidth',3);
set(gcf,'Color','White','Position' ...
    ,[scrsize(3)/ss(nfig) 70 scrsize(3)*0.60 scrsize(4)*0.80]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% end mcm2acceptreject.m
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

C.21 Program: Monte Carlo Multidimensional Integration

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function mcm3multidim
% mcm3multidim: Monte Carlo Multidimensional Integration,
%     nx = 2:5 dims, normal distribution,
%     I = int(F(x),x=a..b), F(x) = exp(-sum(x.^2)/2)/sqrt(2*pi)^nx,
%     so f(x) = I_{a<= x <= b}, an indicator function using vector inequalities.
%
clc; clear
%
fprintf('Monte Carlo Multidimenstional Integration:');
fprintf('\n   including Acceptance-Rejection Technique Application,');
fprintf('\n   with Normal Dist. on (a,b)');
fprintf('\n   and with int(F(x),x=a..b), F(x) = exp(-x.^2/2)/sqrt(2pi)^nx;\n');
%
nxmax = 5; % dimension
kmax = 6; % power of 10
f = inline ('exp(-sum(x.*x)/2) / sqrt(2*pi)^length(x)','x');
for nx = 2:nxmax
    a = -2*ones(1,nx); % lower vector limit
    b = 2*ones(1,nx); % upper vector limit
    for k = 1:kmax
        nmc(k) = 10^k; % sample size
        nac = 0;
        for n = 1:nmc(k)
```

```
            x = randn(1,nx); % MATLAB 1Xnmc normal distribution;
            if (x >= a) & (x <= b) % von Neumann accept-reject technique
                nac = nac + 1; % counts accepted points;
            end
        end
        muhatn(k,nx) = nac/nmc(k);
        stderrn(k,nx) = sqrt(muhatn(k,nx)*(1-muhatn(k,nx))/(nmc(k)-1));
    end
end
%
fprintf('\nMonte Carlo results in mutlidimension,');
fprintf('\n    by von Neumann''s Acceptance-Rejection technique:\n');
fprintf('\nMonte Carlo Mean Estmate, muhatn:');
fprintf('\n k        n      nx=2      nx=3      nx=4      nx=5\n');
for k = 1:kmax
    fprintf ('%2i %8i  %8.6f  %8.6f  %8.6f  %8.6f\n'...
        ,k,nmc(k),muhatn(k,2:nxmax));
end
fprintf('\nMonte Carlo Std Error Estmate, stderrn:');
fprintf('\n k        n       nx=2       nx=3       nx=4       nx=5\n');
for k = 1:kmax
    fprintf ('%2i %8i  %9.3e  %9.3e  %9.3e  %9.3e\n'...
        ,k,nmc(k),stderrn(k,2:nxmax));
end
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
scrsize = get(0,'ScreenSize');
ss = [3.0,2.8,2.6,2.4,2.2,2.0];
%
nfig = 1;
figure(nfig);
kv = [1:kmax];
plot(kv,muhatn(:,2),'k-o'...
    ,kv,muhatn(:,3),'k-x'...
    ,kv,muhatn(:,4),'k-+'...
    ,kv,muhatn(:,5),'k-*'...
    ,'linewidth',3,'MarkerSize',14);
axis([min(kv) max(kv) 0.5 1]);
title('Monte Carlo Means, Normal Dist., F(x) = \phi_n(x) on [a,b]'...
    ,'Fontsize',36,'FontName','Helvetica','FontWeight','Bold');
xlabel('log(n), Log_{10} Sample Size'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
ylabel('Mean Estimates, \mu_n'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
legend('nx = 2','nx = 3','nx = 4','nx = 5','Location','Best');
set(gca,'Fontsize',28,'FontName','Helvetica','FontWeight','Bold','linewidth',3);
```

```
set(gcf,'Color','White','Position' ...
    ,[scrsize(3)/ss(nfig) 70 scrsize(3)*0.60 scrsize(4)*0.80]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
nfig = nfig+1;
figure(nfig);
kv = [1:kmax];
plot(kv,stderrn(:,2),'k-o'...
    ,kv,stderrn(:,3),'k-x'...
    ,kv,stderrn(:,4),'k-+'...
    ,kv,stderrn(:,5),'k-*'...
    ,'linewidth',3,'MarkerSize',14);
axis tight; % axis([min(kv) max(kv) 0 0.2]);
title('Monte Carlo Std. Errors, Normal Distribution on [a,b]'...
    ,'Fontsize',36,'FontName','Helvetica','FontWeight','Bold');
xlabel('log(n), Log_{10} Sample Size'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
ylabel('Std. Errors, stderr_n'...
    ,'Fontsize',32,'FontName','Helvetica','FontWeight','Bold');
legend('nx = 2','nx = 3','nx = 4','nx = 5','Location','Best');
set(gca,'Fontsize',28,'FontName','Helvetica','FontWeight','Bold','linewidth',3);
set(gcf,'Color','White','Position' ...
    ,[scrsize(3)/ss(nfig) 70 scrsize(3)*0.60 scrsize(4)*0.80]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% end mcm3multidim.m
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```