FinM 331/Stat 339 Financial Data Analysis, Winter 2010

Floyd B. Hanson, Visiting Professor

Email: fhanson@uchicago.edu

Master of Science in Financial Mathematics Program
University of Chicago

Lecture 8

6:30-9:30 pm, 22 February 2010, Ryerson 251 in Chicago

7:30-10:30 pm, 22 February 2010 2010 at UBS in Stamford

7:30-10:30 am, 23 February 2010 at Spring in Singapore

8. Bayesian Distribution and Parameter Estimation, and Simulation, Methods:

• 8.1. Bayesian Inference Introduction:[a]

The Bayesian approach is by probabilistic inference through exploring the relationships between joint, conditional and marginal distributions, different from previous approaches focused on the distribution or density and called the frequentist approach. For the two random variable case, X and Y, the approach involves various forms of the **definition of the conditional probability**, i.e.,

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \tag{8.1}$$

and many generalizations[b] of this, often called **Bayesian formulas**.

[a] See Rice (2007), pp. 94-95 & 285-312.

[b] Another form follows from a pair of **multiplicative decompositions of the joint density**, i.e., $f_{Y|X}(y|x) f_X(x) = f_{X,Y}(x,y) = f_{X|Y}(x|y) f_Y(y)$, then

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) f_Y(y)}{f_X(x)}, \tag{8.2}$$

Bayes theorem for the Bayesian posterior formula used later.

○ *8.1.1. Simple Bayesian Example:*

Starting with a motivational example, suppose the market is in near equilibrium, such that the probability of an up or down move is 1/2 either way. Let the asset price be S_i and the observed IID daily changes $\Delta S_i = S_{i+1} - S_i$ for $i = 1:n$, such that the up-count is

$$X = \sum_{i=1}^{n} \mathcal{I}_{\{\Delta S_i > 0\}}. \tag{8.3}$$

Let Θ be the probability of an up or positive change, $\Delta S_i > 0$.

The **Bayesian prior distribution**, in absence of other knowledge, is an educated guess about a problem parameter, e.g, suppose that the distribution is uniform on $[0, 1]$ and the density is

$$f_\Theta(\theta) \simeq f_\Theta^{(\mathrm{prior})}(\theta) = 1, \;\; \theta \in [0, 1]. \tag{8.4}$$

Observation of the changes X_i can alter our knowledge of Θ yielding a **Bayesian posterior distribution**, $f_{\Theta|X}^{(\mathrm{post})}(\theta|x)$.

The path to the posterior distribution involves several steps of Bayesian analysis. First, the up-count $\boldsymbol{X}$ with given probability $\boldsymbol{\Theta}$ must have a binomial distribution for $\boldsymbol{n}$ trials and $\boldsymbol{X}$ successes (ups)

$$f_{X|\Theta}(x|\theta) = \binom{n}{x}\theta^{x}\left(1-\theta^{n-x}\right), \;\; for \;\; x = 0:n. \tag{8.5}$$

Next, the continuous $\boldsymbol{\Theta}$ and discrete count $\boldsymbol{X}$ have a joint distribution according the Bayesian chain rule inverting the conditional definition (8.1) and approximated with the prior density,

$$f_{\Theta,X}(\theta, x) \simeq f_{X|\Theta}(x|\theta) f_{\Theta}^{(\mathrm{prior})}(\theta) = \binom{n}{x}\theta^{x}\left(1-\theta^{n-x}\right)\times 1, \tag{8.6}$$

for $\theta \in [0, 1]$ and for $x = 0:n$. Then, the **marginal X-density** with integration is

$$f_X(x) = \int_0^1 f_{\Theta,X}(\theta, x)d\theta \simeq \binom{n}{x}\int_0^1 \theta^{x}\left(1-\theta^{n-x}\right)d\theta. \tag{8.7}$$

A quick way to integrate in (8.7), is to note that as a function of $\boldsymbol{\theta}$ is essentially a **beta density** on $[\mathbf{0}, \mathbf{1}]$,

$$\begin{aligned} f_U^{(\beta)}(u; a, b) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1}(1-u^{b-1}) \\ &= \frac{(a+b-1)!}{(a-1)!(b-1)!} u^{a-1}(1-u^{b-1}), \end{aligned} \tag{8.8}$$

by the gamma and factorial function properties on L7p10. Comparing powers of the variables, setting $u = \theta$, $a - 1 = x$ and $b - 1 = n - x$, so using the fact that $f_U^{(\beta)}(u; a, b)$ is a proper density and the factorial form of the binomial coefficient,

$$f_X(x) \simeq \frac{(\mathrm{n})!}{x!(n-x)!} \frac{x!(n-x)!}{(n+1)!} = \frac{1}{n+1}, \;\; for \;\; x = 0 \!:\! n. \tag{8.9}$$

So, a uniform prior distribution leads to the discrete form of a uniform distribution.

Finally, the Bayesian paradigm leads to the **parameter Bayesian posterior distribution**,

$$\begin{aligned} f_{\Theta|X}(\theta|x) &= \frac{f_{\Theta,X}(\theta,x)\times 1}{f_X(x)} \simeq (n+1)\binom{n}{x}\theta^x(1-\theta^{n-x}) \\ &= \frac{(n+2)!}{x!(n-x)!}\theta^x(1-\theta^{n-x}) \\ &= f_{\Theta}^{(\beta)}(\theta; x+1, n-x+1) \equiv f_{\Theta|X}^{(\text{post})}(\theta|x) \end{aligned} \tag{8.10}$$

the **posterior density is a beta density** in Θ given $X=x$ with $x=0:n$. The basic statistics[a] are a mean of $\frac{a}{a+b}=\frac{x+1}{n+2}$, a mode of $\frac{a-1}{a+b-2}=\frac{x}{n}$, and a variance of $\frac{ab}{(a+b)^2(a+b+1)}=\frac{(x+1)(n-x+1)}{(n+2)^2(n+3)}$. An approximate 95% confidence interval can be calculated from

```
CI95=[betainv(0.025,a,b),betainv(0.975,a,b)];
```
(8.11)

[a] M. Evans, N. Hastings and B. Peacock (2000), Statistical Distributions. A very useful handbook in handy pocketbook size.

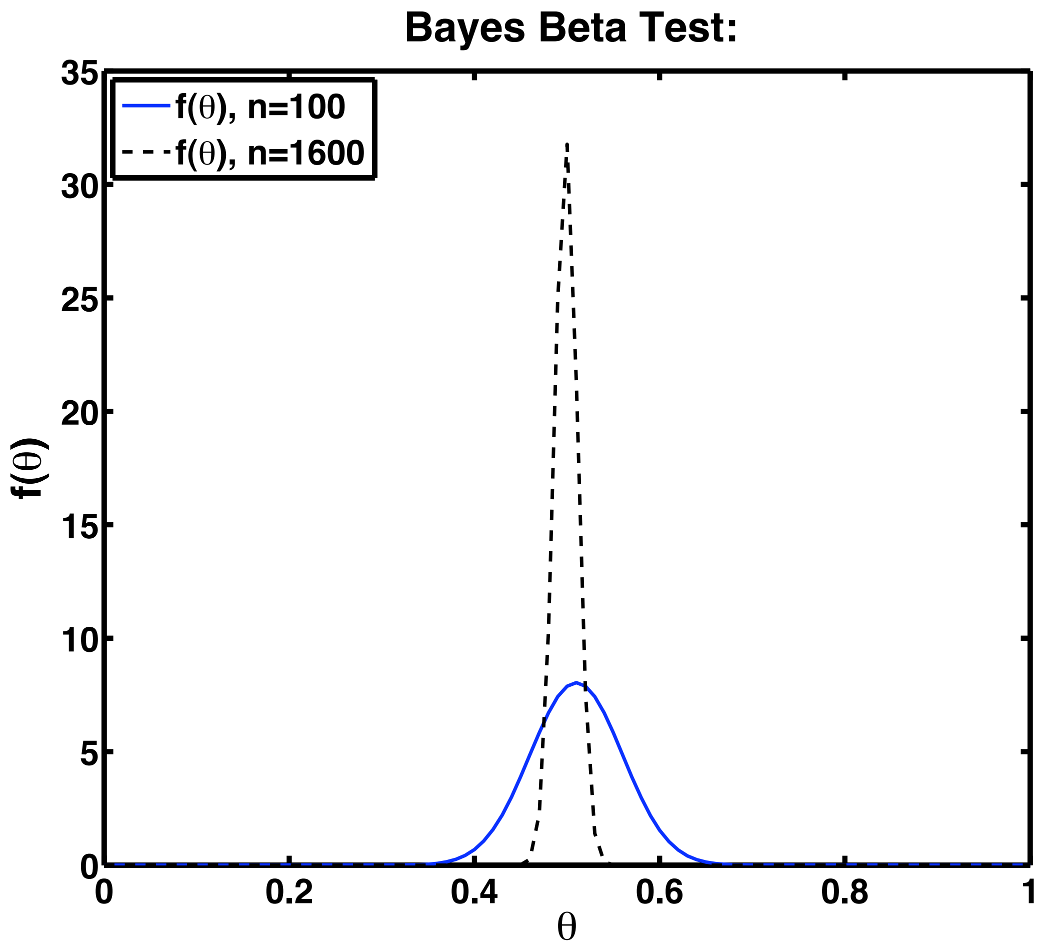

Figure 8.1: **Bayes estimation example showing the Bayes posterior density** for the probability $\boldsymbol{\theta}$ of a positive change from mean zero normal distribution sample with $\boldsymbol{\sigma = 0.02}$ and `n=[100,1600]`. For $f_{\Theta|X}(\theta|x)$, `mean=[0.5098,0.4988]`, `mode=[0.5100,0.4988]`, the `variance=[0.0493,0.0125]`.

In Figure 8.1, it is seen that the distribution is much sharper the larger the sample size, here $\mathbf{n = 1600 > 100}$ and the 95% confidence interval is $\mathbf{[0.4133, 0.6060]}$ with 100 observations, but $\mathbf{[0.4743, 0.5232]}$ for 1600. In general, the larger the sample size the closer the parameter estimate, either using the mean or the mode. Using the mode corresponds, in principle, to the maximum likelihood estimate.

∘ *MATLAB Code for Bayes Posterior Distribution for Probability of Positivity Change:*

```
function BayesBetaTest2
clc
%
fprintf('\nBayesBetaTest Output (%s):',datestr(now));
mu=0; sigma=0.02;
dth=0.01; theta=0:dth:1;
ni=0;
ftheta=zeros(2,length(theta));
for n=[100,1600]
    ni=ni+1;
    ds=normrnd(mu,sigma,1,n);
    x=sum((ds>0));
    fprintf('\nn=%i; mu=%5.2f; sigma=%5.2f; x=%i;',n,mu,sigma,x);
    a=x+1; b=n-x+1; mutheta=a/(a+b); modetheta=(a-1)/(a+b-2);
    vartheta=a*b/((a+b)^2*(a+b+1)); sigtheta=sqrt(vartheta);
    fprintf('\na=%i;b=%i;mutheta=%6.4f;modetheta=%6.4f;sigtheta=%7.5f;'...
        ,a,b,mutheta,modetheta,sigtheta);
    th1=betainv(0.025,a,b); th2=betainv(0.975,a,b); % 95%CI
    fprintf('\n95%%CI: theta in [%6.4f,%6.4f];',th1,th2);
    ftheta(ni,:)=betapdf(theta,a,b);
end
%
```

```
figure(1); nfig = 1;
scrsize = get(0,'ScreenSize'); % figure spacing for target screen
ss = [5.0,4.5,4.0,3.5]; % figure spacing factors
plot(theta,ftheta(1,:),'-b',theta,ftheta(2,:),'--k','LineWidth',2);
title('Bayes Beta Test:'...
    ,'Fontsize',24,'FontWeight','Bold');
xlabel('\theta','Fontsize',24,'FontWeight','Bold');
ylabel('f(\theta)','Fontsize',24,'FontWeight','Bold');
legend('f(\theta), n=100','f(\theta), n=1600','Location','NorthWest');
set(gca,'Fontsize',20,'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]
fprintf('\n ');
================== OUTPUT ========================
BayesBetaTest Output (17-Feb-2010 13:46:40):
n=100; mu= 0.00; sigma= 0.02; x=51;
a=52; b=50; mutheta=0.5098; modetheta=0.5100; sigtheta=0.04926;
95%CI: theta in [0.4133,0.6060];
n=1600; mu= 0.00; sigma= 0.02; x=798;
a=799; b=803; mutheta=0.4988; modetheta=0.4988; sigtheta=0.01249;
95%CI: theta in [0.4743,0.5232];
>>
```

∘ *8.1.2. Bayesian Estimation Approach Summary:*

1. The given data $\vec{X}$, providing the relationship of $\vec{X}$ to the given parameter $\boldsymbol{\Theta} = \theta$ and the **likelihood function**, has probability density $f_{\vec{X}|\Theta}(\vec{x}|\theta)$.

2. The unknown parameter θ is treated as a RV $\boldsymbol{\Theta}$ with **prior distribution** density, **often inferred from the likelihood function**,

$$f_{\Theta}(\theta) \simeq f_{\Theta}^{(\text{prior})}(\theta). \qquad (8.12)$$

3. The joint density of the data $\vec{X}$ and parameter $\boldsymbol{\Theta}$ satisfy the law of total probability reduced from the defiinition of the conditional probability,

$$f_{\vec{X},\Theta}(\vec{x},\theta) \simeq f_{\vec{X}|\Theta}(x|\theta) f_{\Theta}^{(\text{prior})}(\theta). \qquad (8.13)$$

4. The **marginal distribution of** the data $\vec{X}$ is

$$f_{\vec{X}}(\vec{x}) \simeq \int_{\mathcal{R}} f_{\vec{X}|\Theta}(\vec{x}|\theta) f_{\Theta}^{(\text{prior})}(\theta) d\theta. \qquad (8.14)$$

5. Thus, the **Bayes Theorem** for the parameter $\boldsymbol{\Theta}$ is, given the data $\vec{X} = \vec{x}$ and recalling the simpler Bayesian posterior formula (8.2),

$$f_{\Theta|\vec{X}}(\theta|\vec{x}) = \frac{f_{\Theta,\vec{X}}(\theta,\vec{x})}{f_{\vec{X}}(\vec{x})} \simeq \frac{f_{\vec{X}|\Theta}(\vec{x}|\theta) f_{\Theta}^{(\text{prior})}(\theta)}{\int_{\mathcal{R}} f_{\vec{X}|\Theta}(\vec{x}|t) f_{\Theta}^{(\text{prior})}(t) dt}. \qquad (8.15)$$

Since the integral in the denominator of (8.15) can be viewed as the normalization constant for the numerator is is often written omitting the denominator, as the proportional relationship[a],

$$f_{\Theta|\vec{X}}(\theta|\vec{x}) \propto f_{\vec{X}|\Theta}(\vec{x}|\theta) \times f_{\Theta}^{(\mathrm{prior})}(\theta), \tag{8.16}$$

where $\mathrm{LH}(\theta|\vec{x}) = f_{\vec{X}|\Theta}(\vec{x}|\theta)$ is called the **likelihood** as a function of θ, with reversed arguments, so Bayes theorem can be written in shorthand as

$$\mathrm{PosteriorDensity} \propto \mathrm{Likelihood} \times \mathrm{PriorDensity}. \tag{8.17}$$

[a] Using the rule that if C is a nonzero scalar constant, then $Cf(x,y) \propto f(x,y)$.

○ *8.1.3. Bayesian Fitting a Poisson Distribution:*

In this example, there are IIPD, Poisson distributed data observations of jump counts $\vec{X} = [X_i]_{n\times 1}$, so that the unknown parameter is the Poisson parameter λ with prior distribution or density[a] is $f_\Lambda(\lambda) \simeq f_\Lambda^{(\mathrm{prior})}(\lambda)$. Given λ, each jump-count observation is distributed as

$$f_{X_i|\Lambda}(x_i|\lambda) = e^{-\lambda}\frac{\lambda^{x_i}}{x_i!},\ for\ \ x_i = 0, 1, 2, \ldots\ \&\ for\ \ i = 1\!:\!n \tag{8.18}$$

and the joint distribution of all the observations, i.e., the **total likelihood**, is

$$f_{\vec{X}|\Lambda}(\vec{x}|\lambda) \stackrel{\mathrm{iid}}{=} \prod_{i=1}^{n} f_{X_i|\Lambda}(x_i|\lambda) = e^{-n\lambda}\frac{\lambda^{n\overline{X}_n}}{\prod_{i=1}^{n} x_i!} \tag{8.19}$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$.

---

[a]Notation Alert: In this Bayesian Poisson problem, capital letters (Λ) are used for the RV with lower case for the given variable (λ), whereas in the stochastic dynamics $\Lambda = \lambda\Delta t$ is the parameter.

Upon canceling the common factorial terms, the Bayes posterior density is

$$f_{\Lambda|\vec{X}}(\lambda|\vec{x}) \simeq f^{(\mathrm{post})}_{\Lambda|\vec{X}}(\lambda|\vec{x}) = \frac{e^{-n\lambda}\lambda^{n\overline{X}_n} f^{(\mathrm{prior})}_{\Lambda}(\lambda)}{\int_{\mathcal{R}^+} e^{-nt} t^{n\overline{X}_n} f^{(\mathrm{prior})}_{\Lambda}(t)dt}. \quad (8.20)$$

In the first example with a binomial conditional likelihood for the data, it was demonstrated how to tighten up the Bayes parameter estimate by using a larger sample size, although the smaller sample size yielded a reasonable estimate. Also, a demonstration of a weak sensitivity to the starting prior distribution approximation (***"guestimate"***) is presented. Each statistics user can choose a prior from experience that complements the data conditional likelihood distribution in a way that the combine product is close to a recognizable distribution.

Since the current data likelihood distribution is like a gamma distribution, one might select a gamma distribution for the prior distribution, i.e.,

$$f_{\Lambda}^{(\mathrm{prior},1)}(\lambda) \simeq f_{\Lambda}^{(g)}(\lambda;\nu,\alpha) = \frac{\nu^{\alpha}\lambda^{\alpha-1}}{\Gamma(\alpha)}e^{-\nu\lambda}, \tag{8.21}$$

where ν is a substitute for λ. Then, the Bayes posterior density is

$$f_{\Lambda|\vec{X}}^{(\mathrm{post},1)}(\lambda|x) = \frac{\lambda^{n\overline{X}_n+\alpha-1}e^{-(n+\nu)\lambda}}{\int_0^{\infty} t^{n\overline{X}_n+\alpha-1}e^{-(n+\nu)t}dt}. \tag{8.22}$$

Given the **sufficient statistics** $\overline{X}_n$ and n, the denominator is essentially a constant normalizing the numerator and since the posterior distribution must be proper, it must be a proper gamma distribution by inspection. Thus,

$$f_{\Lambda|\vec{X}}^{(\mathrm{post},1)}(\lambda|\vec{x}) = f_{\Lambda}^{(g)}(\lambda;\tilde{\nu},\widetilde{\alpha}) \tag{8.23}$$

where the new gamma parameters, called **hyperparameters**, are $\widetilde{\alpha} \equiv n\overline{X}_n + \alpha$ and $\tilde{\nu} \equiv n + \nu$. Thus, the posterior mean estimate is $\mu^{(\mathrm{post},1)} = \widetilde{\alpha}/\tilde{\nu}$ with mode estimate $\lambda^{(\mathrm{post},1,*)} = (\widetilde{\alpha} - 1)/\tilde{\nu}$ and standard deviation estimate $\sigma^{(\mathrm{post},1)} = \sqrt{\widetilde{\alpha}}/\tilde{\nu}$.

Another choice of the prior might be the uniform distribution, realistically assuming that the jump parameter will be finite, so

$$f^{(\mathrm{post},2)}_{\Lambda|\vec{X}}(\lambda|\vec{x}) = f^{(u)}_{\Lambda}(\lambda; 0, b) = \frac{1}{b}\mathcal{I}_{\{\lambda\in[0,b]\}}. \qquad (8.24)$$

Then, the Bayes posterior density is

$$f^{(\mathrm{post},2)}_{\Lambda|\vec{X}}(\lambda|\vec{x}) = \frac{\lambda^{n\overline{X}_n} e^{-n\lambda}}{\int_0^b t^{n\overline{X}_n} e^{-nt} dt} \sim f^{(g)}_{\Lambda}(\lambda; \widehat{\nu}, \widehat{\alpha}) \qquad (8.25)$$

only for $\lambda \in [0, b]$ and $b \gg 1$, where the posterior gamma parameters are $\widehat{\alpha} \equiv n\overline{X}_n$ and $\widehat{\nu} \equiv n$, noting that the normalization in the denominator is incomplete so that only an approximation for large b. In MATLAB, the finite gamma density using the incomplete gamma function

```
fpost2=gamma(alpha)*gampdf(lambda,alpha,1/nu)...
      /gammainc(b/nu,alpha);
```
(8.26)

where (`lambda,alpha,nu`) $= (\lambda, \widehat{\alpha}, \widehat{\nu})$. So, a simpler prior here leads to a numerically complicated posterior.

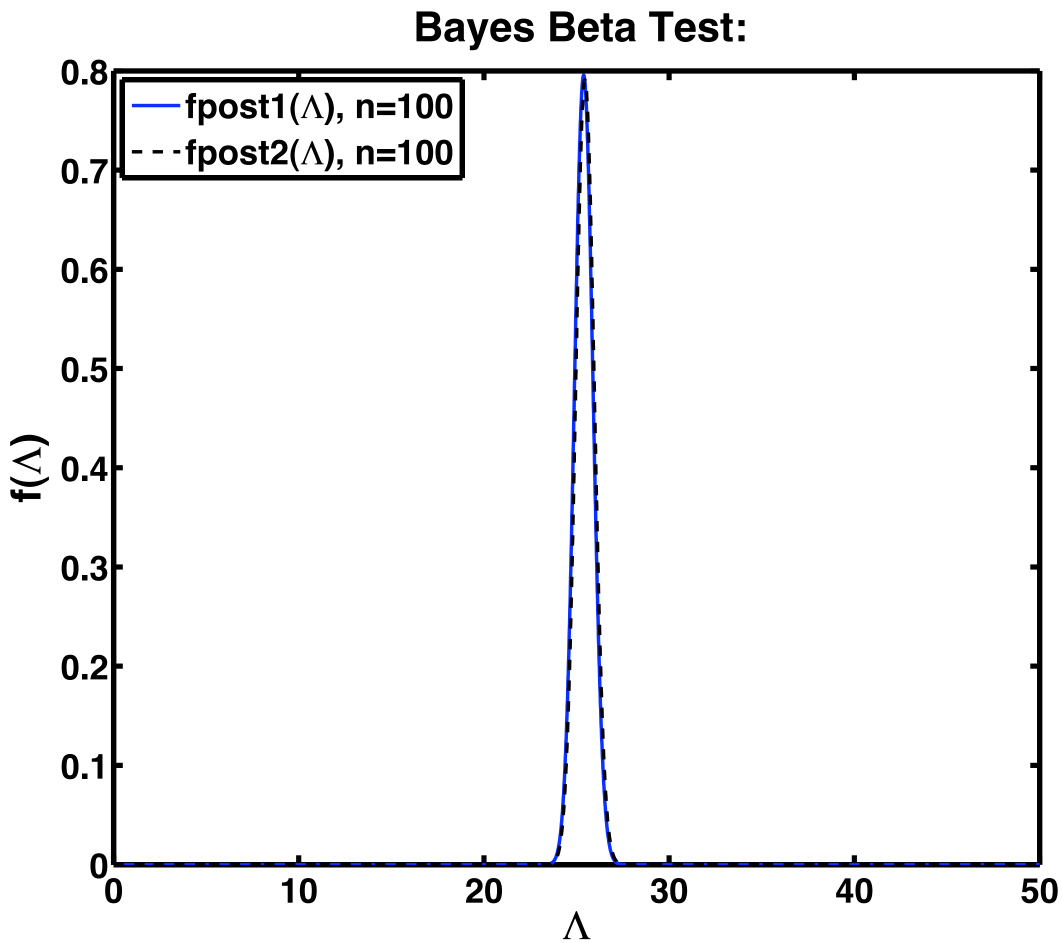

Figure 8.2: **Bayes estimation example showing the Bayes posterior density** for Poisson parameter $\mathbf{\Lambda = \lambda}$ of Poisson distribution sample with $\mathbf{\Lambda 0 = 25}$ and `n=100` simulated observations. For two different posterior distributions, $f^{(\mathrm{post})}_{\Lambda|X}(\lambda|x)$, there is only a **0.25% difference** in either estimated mean or mode.

In Figure 8.2, there is only a small difference between the two posterior distributions with $n = 100$ observations. The two means are `[25.40,25.46]`, modes are `[25.39,25.45]` and standard deviations are `[0.5014,0.5046]`. The 95% confidence intervals are $[\mathbf{24.42, 26.39}]$ and $[\mathbf{24.48, 26.46}]$, respectively for the first and second posterior distribution.

This is just a demonstration that the the Bayesian estimation is not too sensitive to a reasonable prior. Note that for the second prior, the asymptotic form in (8.25) rather than the numerically more precise (8.26) **MATLAB** formula. However, for smaller sample sizes **n** the tails will be fatter and the latter form may be needed. An advantage of the uniform prior is that it does not introduce bias toward the selection of the posterior.

◦ *MATLAB Code for Bayes Posterior Distribution for Poisson Parameter Using Gamma and Uniform Priors:*

```
function BayesGammaTest2
clc
%
fprintf('\nBayesGammaTest Output (%s):',datestr(now));
Lambda0=25; al=20; nu=1; % take meangamma=20/1=20;
fprintf('\nLambda0=%5.2f; al=%5.2f; nu=%5.2f;',Lambda0,al,nu);
dlm=0.01; lambda=0:dlm:2*Lambda0;
n=100;
X=poissrnd(Lambda0,1,n);
Xmean=mean(X);
fprintf('\nn=%i; Xmean=%5.2f;',n,Xmean);
    al1=n*Xmean+al-1; nu1=n+nu; mugam1=al1/nu1; modegam1=(al1-1)/nu1;
    vargam=al1/nu1^2; siggam=sqrt(vargam); b1=1/nu1;
    fprintf('\nal1=%i;nu1=%i;mugam=%5.2f;modegam=%5.2f;siggam=%6.4f;'...
        ,al1,nu1,mugam1,modegam1,siggam);
    lm1=gaminv(0.025,al1,b1); lm2=gaminv(0.975,al1,b1); % 95%CI
    fprintf('\n95%%CI: gamma in [%5.2f,%5.2f];',lm1,lm2);
    flm1(1,:)=gampdf(lambda,al1,b1);
%
    al2=n*Xmean; nu2=n; mugam2=al2/nu2; modegam2=(al2-1)/nu2;
    vargam=al2/nu2^2; siggam=sqrt(vargam); b2=1/nu2;
    fprintf('\nal2=%i;nu2=%i;mugam=%5.2f;modegam=%5.2f;siggam=%6.4f;'...
```

```
        ,al2,nu2,mugam2,modegam2,siggam);
    lm1=gaminv(0.025,al2,b2); lm2=gaminv(0.975,al2,b2); % 95%CI
    fprintf('\n95%%CI: Lambda in [%5.2f,%5.2f];',lm1,lm2);
    flm2(1,:)=gampdf(lambda,al2,b2); % Approx. finite by full gamma
%
fprintf('\nmugamdif=%5.2f%%; modegamdif=%5.2f%%;'...
        ,(mugam2/mugam1-1)*100,(modegam2/modegam1-1)*100);
%
figure(1); nfig = 1;
scrsize = get(0,'ScreenSize'); % figure spacing for target screen
ss = [5.0,4.5,4.0,3.5]; % figure spacing factors
plot(lambda,flm1(1,:),'-b',lambda,flm2(1,:),'--k','LineWidth',2);
title('Bayes Beta Test:'...
    ,'Fontsize',24,'FontWeight','Bold');
xlabel('\Lambda','Fontsize',24,'FontWeight','Bold');
ylabel('f(\Lambda)','Fontsize',24,'FontWeight','Bold');
legend('fpost1(\Lambda), n=100','fpost2(\Lambda), n=100'...
    ,'Location','NorthWest');
set(gca,'Fontsize',20,'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]
fprintf('\n ');
```

```
================= OUTPUT =========================
BayesGammaTest Output (17-Feb-2010 23:00:28):
Lambda0=25.00; al=20.00; nu= 1.00;
n=100; Xmean=25.46;
al1=2565; nu1=101; mugam=25.40; modegam=25.39; siggam=0.5014;
95%CI: gamma in [24.42,26.39];
al2=2546; nu2=100; mugam=25.46; modegam=25.45; siggam=0.5046;
95%CI: Lambda in [24.48,26.46];
mugamdif= 0.25%; modegamdif= 0.25%;
>>
```

○ *8.1.4. Bayesian Fitting a Normal Distribution:*

In the case of a normal distribution problem the data $\vec{X} = [X_i]_{n\times 1}$ belong to an IIND RV sample with likelihood distribution $f^{(n)}_{X_i|M,S^2}(x_i|\mu,\sigma^2)$. However, it would be better to work with the variance itself $v = \sigma^2$ as a parameter and better yet to work with the **precision** $\xi \equiv 1/\sigma^2$ to also avoid the reciprocal in the density exponent. Hence, let the better likelihood density, setting $\theta = \mu$ for notational convenience, $f^{(n)}_{X_i|M,S^2}(x_i|\mu,\sigma^2)$

$$f_{X_i|\Theta,\Xi}(x_i|\theta,\xi) = \sqrt{\frac{\xi}{2\pi}}\exp\left(-0.5\xi(x_i-\theta)^2\right). \tag{8.27}$$

Since the previous example had only a single parameter and the complexity grows rapidly with the number of parameter, we will first work on the smaller problem with θ as the unknown parameter and $\xi = \xi_0$ fixed, then ξ unknown and $\theta = \theta_0$ fixed, before tackling both (θ, ξ) unknown.

⋆ *8.1.4(a) Only Mean θ Unknown Normal Case:*

Let the normal precision be given, $\xi = \xi_0$, and used for the likelihood (8.27). The prior of the mean Θ is assumed to be normal with mean θ_0 and variance $1/\xi_1$, with $\xi_1 \ll 1$ for a low bias prior, i.e.,

$$f_{\Theta}^{(\text{prior},1)}(\theta) = \sqrt{\frac{\xi_1}{2\pi}} \exp\left(-0.5\xi_1(\theta - \theta_0)^2\right). \tag{8.28}$$

The posterior distribution in the Bayesian shorthand is

$$\begin{aligned} f_{\Theta|\vec{X}}^{(\text{post},1)}(\theta|\vec{x}) &\propto f_{\vec{X}|\Theta}(\vec{x}|\theta) \times f_{\Theta}^{(\text{prior},1)}(\theta) \\ &= \prod_{i=1}^{n}\left(\sqrt{\frac{\xi_0}{2\pi}} \exp\left(-0.5\xi_0(x_i - \theta)^2\right)\right) \\ &\quad \times \sqrt{\frac{\xi_1}{2\pi}} \exp\left(-0.5\xi_1(\theta - \theta_0)^2\right) \\ &\propto \exp\left(-0.5\left(\xi_0\sum_{i=1}^{n}(x_i - \theta)^2 + \xi_1(\theta - \theta_0)^2\right)\right) \end{aligned} \tag{8.29}$$

using the law of exponents and dropping constant coefficients.

By centering the data about the mean, $\overline{X}_n = \sum_{i=1}^n x_i/n$, more constant terms can be eliminated as follows,

$$\begin{aligned}\sum_{i=1}^n (x_i-\theta)^2 &= \sum_{i=1}^n ((x_i-\overline{X}_n) + (\overline{X}_n-\theta))^2 \\ &= n(S_n^2 + (\overline{X}_n-\theta)^2),\end{aligned} \tag{8.30}$$

where $S_n^2 = \sum_{i=1}^n (x_i-\overline{X}_n)/n$ is the **biased sample variance** which is constant respect to θ. Note there is no cross term in (8.30) since $\sum_{i=1}^n (x_i-\overline{X}_n) = 0$. Thus,

$$\begin{aligned}f_{\Theta|\vec{X}}^{(\text{post},1)}(\theta|\vec{x}) &\propto \exp\left(-0.5\left(n\xi_0\left(\overline{X}_n-\theta\right)^2 + \xi_1(\theta-\theta_0)^2\right)\right) \\ &\equiv \exp(-\phi(\theta)),\end{aligned} \tag{8.31}$$

defining the positive part of the exponent $\phi(\theta)$. Since

$$\exp(-\phi(\theta))' = -\phi'(\theta)\exp(-\phi(\theta)), \tag{8.32}$$

the critical point of the exponential will be the critical point of $\phi(\theta)$, $\phi'(\theta^*) = 0$,

$$\widehat{\theta}_n^{(\text{post},1)} = \theta^* = \frac{n\xi_0\overline{X}_n + \xi_1\theta_0}{n\xi_0+\xi_1}, \tag{8.33}$$

unfortunately depending on three guesses $\{\xi_0, \xi_1, \theta_0\}$.

However, for large samples, $n \gg 1$, the asymptotic result is that

$$\widehat{\theta}_n^{(\mathrm{post},1)} = \theta^* \sim \overline{X}_n, \tag{8.34}$$

which is also the **MLE result**. Note that

$$\exp(-\phi)''(\theta) = \Big(-\phi''(\theta) + (\phi')^2(\theta)\Big)\exp(-\phi(\theta)), \tag{8.35}$$

so that $\exp(-\phi)''(\theta^*) < 0$, ensuring a maximum, since $\phi'(\theta^*) = 0$ and $\phi''(\theta^*) = n\xi_0 + \xi_1 > 0$.

The asymptotic MLE result (8.34) is valid as long as $\xi_1 \ll n\xi_0$, assuming $\overline{X}_n = \mathrm{O}(1)$. This may be a weak assumptions if n is not large, but the result would be usable for a **flat prior**, i.e., $\xi_1 \ll 1$. Anyway, in the asymptotic limit we have $\xi^{(\mathrm{post},1)} \sim n\xi_0$ from (8.31), so

$$\sigma^{(\mathrm{post}),1} = 1/\sqrt{\xi^{(\mathrm{post},1)}} \sim 1/\sqrt{n\xi_0} = \sigma_0/\sqrt{n} \sim \mathrm{SE}\left[\widehat{\theta}_n^{(\mathrm{post},1)}\right] \tag{8.36}$$

which is a estimate of the standard error of the Bayes estimate of the mean under constant variance.

⋆ *8.1.4(b) Only Precision ξ Unknown Normal Case:*

Fix the mean $\mu = \theta = \theta_0$ and let the data likelihood be normal distributed as another variant of (8.27) with mean θ_0 and variance $1/\xi$,

$$f_{X_i|\Xi}(x_i|\xi) = \sqrt{\frac{\xi}{2\pi}}\exp\left(-0.5\xi(x_i - \theta_0)^2\right). \tag{8.37}$$

Holding off on the prior distribution of the precision until a more compatible one can be chosen, the posterior distribution in the Bayesian shorthand is

$$\begin{aligned} f^{(\text{post},2)}_{\Xi|\vec{X}}(\xi|\vec{x}) &\propto f_{\vec{X}|\Xi}(\vec{x}|\xi) \times f^{(\text{prior},2)}_{\Xi}(\xi) \\ &= \prod_{i=1}^{n}\left(\sqrt{\frac{\xi}{2\pi}}\exp\left(-0.5\xi(x_i-\theta_0)^2\right)\right)\times f^{(\text{prior},2)}_{\Xi}(\xi) \\ &\propto \xi^{n/2}\exp\left(-0.5\xi\sum_{i=1}^{n}(x_i-\theta_0)^2\right)\times f^{(\text{prior},2)}_{\Xi}(\xi). \end{aligned} \tag{8.38}$$

Note that there is an simple gamma-like dependence of the posterior on the precision ξ with the product of a power and a regular exponential, rather than normal, suggesting the a **gamma prior would be compatible**.

Consequently, let prior be a gamma distribution with parameters (α, λ),

$$f_{\Xi}^{(\text{prior},2)}(\xi) = \frac{\lambda^{\alpha}\xi^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda\xi} \tag{8.39}$$

and the posterior becomes

$$f_{\Xi|\vec{X}}^{(\text{post},2)}(\xi|\vec{x}) \propto \xi^{n/2+\alpha-1} \exp\left(-0.5\xi \sum_{i=1}^{n}(x_i - \theta_0)^2 - \lambda\xi\right), \tag{8.40}$$

so the posterior gamma parameters are

$$\widehat{\alpha}_n^{(\text{post},2)} = \frac{n}{2} + \alpha \tag{8.41}$$

and

$$\widehat{\lambda}_n^{(\text{post},2)} = \frac{1}{2}\sum_{i=1}^{n}(x_i - \theta_0)^2 + \lambda = \frac{n}{2}(S_n^2 + (\overline{X}_n - \theta_0)^2) + \lambda. \tag{8.42}$$

For a flat prior with $\alpha \ll 1$ and $\lambda \ll 1$, only the n-terms count asymptotically, so the mean and modes are,

$$\widehat{\overline{\xi}}_n^{(\mathrm{post},2)} = \frac{\widehat{\alpha}_n^{(\mathrm{post},2)}}{\widehat{\lambda}_n^{(\mathrm{post},2)}} \sim \frac{1}{S_n^2 + (\overline{X}_n - \theta_0)^2} \tag{8.43}$$

and

$$\widehat{(\xi^*)}_n^{(\mathrm{post},2)} = \frac{\widehat{\alpha}_n^{(\mathrm{post},2)} - 1}{\widehat{\lambda}_n^{(\mathrm{post},2)}} \sim \frac{1}{S_n^2 + (\overline{X}_n - \theta_0)^2}, \tag{8.44}$$

asymptotically the same if also $n \gg 1$. However, we usually are interested in the variance, $v = \sigma^2 = 1/\xi$, but the mean and mode are not quite the reciprocals (Cf. Rice, p. 292.), i.e.,

$$\widehat{\overline{v}}_n^{(\mathrm{post},2)} = \overline{\left(\frac{1}{\xi}\right)}_n^{(\mathrm{post},2)} = \frac{\widehat{\lambda}_n^{(\mathrm{post},2)}}{\widehat{\alpha}_n^{(\mathrm{post},2)} - 1} \sim S_n^2 + (\overline{X}_n - \theta_0)^2 \tag{8.45}$$

and

$$\widehat{(v^*)}_n^{(\mathrm{post},2)} = \frac{\widehat{\lambda}_n^{(\mathrm{post},2)}}{\widehat{\alpha}_n^{(\mathrm{post},2)} + 1} \sim S_n^2 + (\overline{X}_n - \theta_0)^2, \tag{8.46}$$

where (8.44) is from finding the critical point by differentiating density, $f_V(v) = v^{-2} f_{\Xi}(1/v) = \widetilde{\lambda}^{\widetilde{\alpha}} v^{-\widetilde{\alpha}-1} \exp(-\widetilde{\lambda}/v)/\Gamma(\widetilde{\alpha})$.

⋆ 8.1.4(c) Both Mean θ and Precision ξ Unknown Normal Case:

From the two previous parts, the mean and precision are taken to be independent, such that the mean is normally distributed

$$f_{\Theta}(\theta)^{(\mathrm{prior},3)} = \sqrt{\frac{\xi_3}{2\pi}}\exp\left(-0.5\xi_3(\theta-\theta_3)^2\right), \tag{8.47}$$

with the constants $\{\theta_3, \xi_3\}$and the precision is gamma distributed

$$f_{\Xi}^{(\mathrm{prior},3)}(\xi) = \frac{\lambda^{\alpha}\xi^{\alpha-1}}{\Gamma(\alpha)}e^{-\lambda\xi}. \tag{8.48}$$

Then, the Bayesian posterior is, combining the full component data likelihood (8.27) by their independence property,

$$\begin{aligned} f_{\Theta,\Xi|\vec{X}}^{(\mathrm{post},3)}(\theta,\xi|\vec{x}) &\propto f_{\vec{X}|\Theta,\Xi}(\vec{x}|\Theta,\xi)\times f_{\Theta}^{(\mathrm{prior},3)}(\theta)\times f_{\Xi}^{(\mathrm{prior},3)}(\xi) \\ &\propto \xi^{n/2}\exp\left(-0.5\xi\sum_{i=1}^{n}(x_i-\theta)^2\right) \\ &\quad\times\exp\left(-0.5\xi_3(\theta-\theta_3)^2\right)\times\xi^{\alpha-1}e^{-\lambda\xi}. \end{aligned} \tag{8.49}$$

Note that although the mean and precision were independent as priors, they are interrelated in the posterior by the data likelihood.

Due to the interdependence, normalization will be need to compute moments and other properties. There are software for treating problems like this and worse.

We can also consider the marginal distributions, one parameter at a time. However, selecting either one, $\boldsymbol{\theta}$ as normally distributed or $\boldsymbol{\xi}$ as gamma distributed, leads to more complications.

However, seem later when these problems are eliminated.

○ *8.1.5. Large Sample Maximum Likelihood Normal Approximation for the Posterior:*

It was demonstrated earlier how the posterior distribution becomes much tighter around the mode as the sample size n becomes larger. This suggests the tendency to a posterior normal approximation and toward a maximum likelihood distribution.

Let the **log-likelihood function** for scalar parameter θ and data $\vec{X} = [X_i]_{n\times 1}$, which will be supressed to focus on θ, denoted as

$$\mathbf{LLH}(\theta) = \log\left(f_{\vec{X}|\Theta}(\vec{x}|\theta)\right) \tag{8.50}$$

and let the maximum likelihood estimate be $\widehat{\theta} = \theta^*$ such that

$$\mathbf{LLH}'(\theta^*) = 0. \tag{8.51}$$

Then the posterior approximation becomes

$$\begin{aligned} f^{(\mathrm{post})}_{\Theta|\vec{X}}(\theta|\vec{x}) &\propto f_{\vec{X}|\Theta}(\vec{x}|\theta) \times f^{(\mathrm{prior})}_{\Theta}(\theta) \\ &= \exp(\mathbf{LLH}(\theta))\exp\left(\log\left(f^{(\mathrm{prior})}_{\Theta}(\theta)\right)\right). \end{aligned} \tag{8.52}$$

As the sample size n becomes large the log-likelihood $\mathrm{LLH}(\theta)$ will dominate the log-prior $\log(f_{\Theta}^{(\mathrm{prior})}(\theta))$, which becomes relatively negligible. Also at the posterior becomes tighter, i.e., the spread narrows, the posterior becomes concentrated in a small neighbor hood of the mode or $\widehat{\theta}$, so a Taylor approximation becomes increasingly valid. Hence,

$$\begin{aligned} f_{\Theta|\vec{X}}^{(\mathrm{post})}(\theta|\vec{x}) &\propto \exp(\mathrm{LLH}(\theta)) \\ &= \exp\left(\mathrm{LLH}\left(\widehat{\theta}\right) + \mathrm{LLH}'\left(\widehat{\theta}\right)\left(\theta-\widehat{\theta}\right)\right. \\ &\quad \left. +0.5\mathrm{LLH}''\left(\widehat{\theta}\right)\left(\theta-\widehat{\theta}\right)^2\right) \\ &\propto \exp\left(0.5\mathrm{LLH}''\left(\widehat{\theta}\right)\left(\theta-\widehat{\theta}\right)^2\right), \end{aligned} \tag{8.53}$$

upon using the MLE critical point condition (8.51) and losing the constant from $\mathrm{LLH}(\widehat{\theta})$, so that the posterior is asymptotically proportional to a normal distribution with mean and variance

$$\widehat{\mu}_{\theta}^{(\mathrm{post})} \sim \widehat{\theta} \;\; \& \;\; \widehat{(\sigma^2)}_{\theta}^{(\mathrm{post})} \sim -1/\mathrm{LLH}''\left(\widehat{\theta}\right). \tag{8.54}$$

respectively, inserting a minus sign for the needed negative log-likehood.

In the case when the parameter is not a scalar, but a vector $\vec{\theta} = [\theta_i]_{m\times 1}$, say, then the log-likelihood becomes a **scalar**-valued function of two vector variables,

$$\mathbf{LLH}(\vec{\theta}) = \log\left(f_{\vec{X}|\vec{\Theta}}(\vec{x}|\vec{\theta})\right) \tag{8.55}$$

and the maximum likelihood estimate be $\widehat{\theta} = \vec{\theta}^*$ satisfies the vector system of MLE equations

$$\nabla_\theta[\mathbf{LLH}](\vec{\theta}^*) = \vec{0}. \tag{8.56}$$

The posterior approximation is

$$\begin{aligned} f^{(\mathbf{post})}_{\vec{\Theta}|\vec{X}}(\vec{\theta}|\vec{x}) &\propto f_{\vec{X}|\vec{\Theta}}(\vec{x}|\vec{\theta}) \times f^{(\mathbf{prior})}_{\vec{\Theta}}(\vec{\theta}) \\ &= \exp(\mathbf{LLH}(\vec{\theta}))\exp\left(\log\left(f^{(\mathbf{prior})}_{\vec{\Theta}}(\vec{\theta})\right)\right). \end{aligned} \tag{8.57}$$

In the limit of large sample size, $n \gg 1$, the posterior is approximated by

$$f^{(\mathrm{post})}_{\vec{\Theta}|\vec{X}}(\vec{\theta}|\vec{x}) \propto \exp\left(\mathrm{LLH}\left(\widehat{\theta}\right) + \left(\vec{\theta}-\widehat{\theta}\right)^{\top} \nabla_{\theta}[\mathrm{LLH}]\left(\widehat{\theta}\right)\right.$$
$$\left. + 0.5\left(\vec{\theta}-\widehat{\theta}\right)^{\top} \nabla_{\theta}\left[\nabla_{\theta}^{\top}[\mathrm{LLH}]\right]\left(\widehat{\theta}\right)\left(\vec{\theta}-\widehat{\theta}\right)\right) \quad (8.58)$$
$$\propto \exp\left(0.5\left(\vec{\theta}-\widehat{\theta}\right)^{\top} \nabla_{\theta}\left[\nabla_{\theta}^{\top}[\mathrm{LLH}]\right]\left(\widehat{\theta}\right)\left(\vec{\theta}-\widehat{\theta}\right)\right),$$

again using the MLE critical condition and absorbing the maximum likelihood, making the **posterior asymptotically proportional to a multivariate normal**[a]**, using the negative log-likelihood, with mean vector and covariance matrix**,

$$\widehat{\mu}^{(\mathrm{post})}_{\theta} \sim \widehat{\theta} \quad \& \quad \widehat{\Sigma}^{(\mathrm{post})}_{\theta} \sim -\left(\nabla_{\theta}\left[\nabla_{\theta}^{\top}[\mathrm{LLH}]\right]\right)^{-1}\left(\widehat{\theta}\right). \quad (8.59)$$

respectively.

[a]See Hanson (2007), Online Appendix B, pp. B46-B48.

⋆ 8.1.5(a) Revisit for Large Samples: Both Mean θ and Precision ξ Unknown Normal Case, (but more about MLE than Bayes):

Returning to the data likelihood for the previous normal posterior example in (8.27) or (8.49 for large sample sizes $n \gg 1$, we have the log-likelihood, letting $v = 1/\xi$ to work directly with the variance of interest rather than the precision and also $\mu = \theta$,

$$\begin{aligned} \mathbf{LLH}(\mu, v) &= \log\left(f^{(\mathrm{post},3)}_{\Theta,\Xi|\vec{X}}(\mu, 1/v|\vec{x})\right) \\ &\propto \log\left(v^{-n/2}\exp\left(-0.5v^{-1}\sum_{i=1}^{n}(x_i - \mu)^2\right)\right) \\ &= -0.5n\left(\log(v) + v^{-1}\left(S_n^2 + (\overline{X}_n - \mu)^2\right)\right). \end{aligned} \tag{8.60}$$

Computing the first order critical point components,

$$\begin{aligned} \frac{\partial \mathbf{LLH}}{\partial \mu}(\mu, v) &= +nv^{-1}(\overline{X}_n - \mu) \stackrel{*}{=} 0; \\ \frac{\partial \mathbf{LLH}}{\partial v}(\mu, v) &= -0.5n\left(v^{-1} - v^{-2}\left(S_n^2 + (\overline{X}_n - \mu)^2\right)\right) \stackrel{*}{=} 0. \end{aligned} \tag{8.61}$$

Solving these two equations simultaneously at critical point (μ^*, v^*), the asymptotic estimate the respective means are

$$\widehat{\mu}_{\mu}^{(\text{post})} = \mu^* \sim \overline{X}_n \;\; \& \;\; \widehat{\mu}_{v}^{(\text{post})} = v^* \sim S_n^2. \tag{8.62}$$

Next calculating the second partial derivatives evaluated at the critical parameters for the estimate of the covariance matrix,

$$\begin{aligned} \frac{\partial^2 \mathbf{LLH}}{\partial \mu^2}(\mu^*, v^*) &= -\frac{n}{v^*} = -\frac{n}{S_n^2}; \\ \frac{\partial^2 \mathbf{LLH}}{\partial v^2}(\mu^*, v^*) &= -0.5n\left(-\frac{1}{(v^*)^2} + \frac{2}{(v^*)^3}\left(S_n^2 + (\overline{X}_n - \mu^*)^2\right)\right) \\ &= -\frac{n}{2S_n^4}; \\ \frac{\partial^2 \mathbf{LLH}}{\partial v \partial \mu}(\mu^*, v^*) &= \frac{\partial^2 \mathbf{LLH}}{\partial \mu \partial v}(\mu^*, v^*) = -\frac{n}{(v^*)^2}(\overline{X}_n - \mu^*) = 0, \end{aligned} \tag{8.63}$$

so the mean and variance estimate are independent in the bivariate normal. Hence the negative inverse is easy to calculate,

$$\widehat{\Sigma}_n^{(\text{post})} = -\mathbf{LLH}^{-1}(\mu^*, v^*) = \begin{bmatrix} S_n^2/n & 0 \\ 0 & 2S_n^4/n \end{bmatrix}. \tag{8.64}$$

The negative inverse of a diagonal matrix is a diagonal matrix that has negative reciprocals for diagonal elements. Thus, the asymptotic normal variance of the parameters are

$$\widehat{(\sigma^2)}_{\mu}^{(\mathrm{post})} \sim \frac{S_n^2}{n} \;\;\&\;\; \widehat{(\sigma^2)}_{v}^{(\mathrm{post})} \sim \frac{2S_n^4}{n} \;\;\&\;\; \widehat{\rho}_{\mu,v}^{(\mathrm{post})} = 0, \qquad (8.65)$$

where $\widehat{\rho}_{\mu,v}^{(\mathrm{post})}$ is the **correlation coefficient** between the estimates of μ and $v = \sigma^2$. The first two relationships can also be expressed as standard errors,

$$\mathrm{SE}_n[\widehat{\mu}_{\mu}^{(\mathrm{post})}] \sim \frac{S_n}{\sqrt{n}} \;\;\&\;\; \mathrm{SE}_n[\widehat{\mu}_{v}^{(\mathrm{post})}] = \mathrm{SE}_n[\widehat{\mu}_{\sigma^2}^{(\mathrm{post})}] \sim \sqrt{\frac{2}{n}} S_n^2. \qquad (8.66)$$

Note that the $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X}_n)^2$ is the **biased sample variance**, but that is what naturally appears.

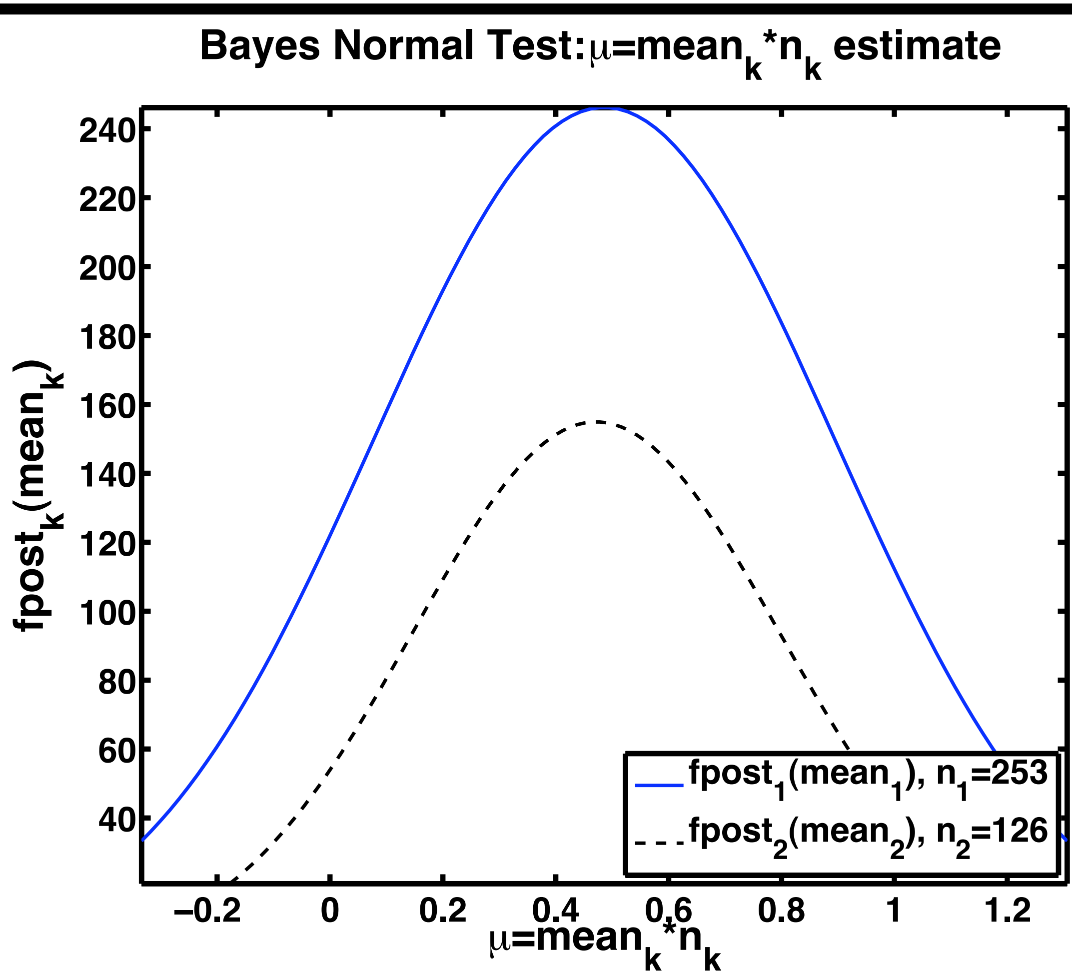

Figure 8.3: **Large sample MLE normal approximation for Bayesian estimation** example using a normal likelihood for the data. Two data samples are used from the 2008 S&P 500 log-returns, one daily and another every other day. The mean asymptotic estimate density has been plotted against the mean coefficient $\mu = n_k * \mathbf{mean}$, $\Delta t = 1/n_k$ and there is good agreement on the mode.

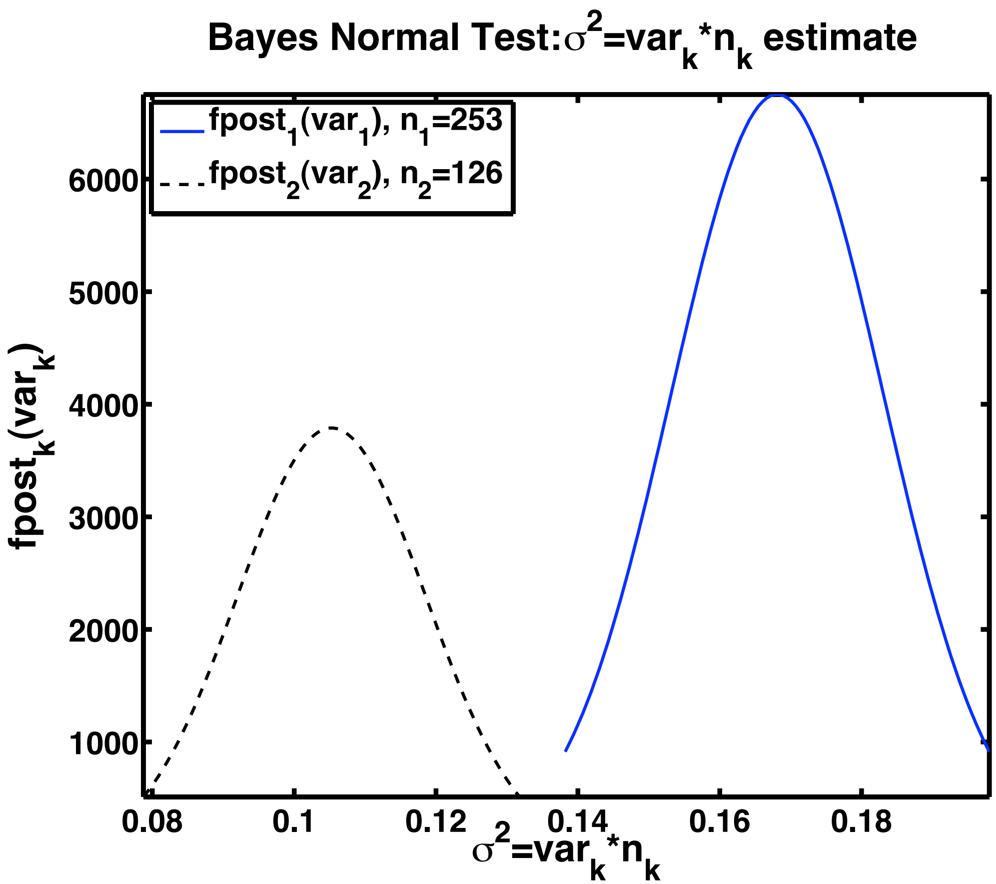

Figure 8.4: **Large sample MLE normal approximation for Bayesian estimation** example using a normal likelihood for the data. Two data samples are used from the 2008 S&P 500 log-returns, one daily and another every other day. The variance asymptotic estimate density has been plotted against the variance coefficient $\boldsymbol{\sigma^2 = n_k * \mathrm{variance}}$, $\boldsymbol{\Delta t = 1/n_k}$ and there is **NO** mode agreement.

The objective of Figs. 8.3-8.4 is to demonstrate the large sample normal approximation by means of the relationship to the maximum likelihood estimation and also the sensitivity to the sample size n. Two sample sizes were chosen from the 2008 S&P500 log-returns, one the complete daily data and the other for every other day, so $n_1 = 235$ and $n_2 = 126$. These are now very large samples. However, the results the two sample, after adjusting the x-axis for the difference in time step using Δt in the first and converting the approximately $2\Delta t$ in the second.

The estimate around the mode for μ is not very sensitive to the sample size, **but those of the variance are not even close and this reflects the some 38% difference in the σ^2 from the data**. The same would be true for the precision $\xi = 1/v$. This leads to a question about whether the σ is a constant and whether the variance $v \propto \Delta t$ is a good assumption. This merits further investigation, since one sample case is not sufficient.

Another observation is that in using the log-returns there is a bias due to the five day trading week. Since if one week has the 3 days MWF than the next week has 2 day TT, missing 60% of the week and that was a week of big 2008 changes that could have an effect.

The variance estimate in the normal approximation is strange, since no positivity constraint was used, although the could be embedded in the prior and posterior distributions. Perhaps appearance of the terms like $\log(v)$ and reciprocals like $1/v$ behave like penalties to negative behavior in the log-likelihood. The real problem with the $1/v$ dependence is that it does not lead to as easy a conjugate prior as does the precision ξ.

Note that prior parameters are sometimes arbitrary or educated guesses to initiate the Bayesian estimation, but **with a sufficiently large sample size the posterior estimates will be less dependent on those initial prior parameter guesses**.

○ *MATLAB Code for Bayes Posterior Distribution for Large Sample Normal Distribution:*

```
function BayesNormalTest2
clc
%
fprintf('\nBayesNormalTest Output (%s):',datestr(now));
load -ASCII S.mat; % Note: Change GSPC2008adjC.txt name for load function.
n1=length(S)-1; fprintf('\nn1=NLR=%3i;',n1);
X1=log(S(2:n1+1))-log(S(1:n1)); %Log-return,dt=1/n1
X1mean=mean(X1); X1var=var(X1,1); X1std=std(X1,1);%LR2008
fprintf('\nX1mean=%8.6f; X1std=%7.5f; X1var=%9.7f;',X1mean,X1std,X1var);
X1mu=n1*X1mean; X1sig2=n1*X1var; X1sig=sqrt(X1sig2);%LR X1mu=mu-sig^2/2;
X1mucor=X1mu+X1sig2/2;
fprintf('\nX1mu=%8.6f; X1sig=%7.5f; X1sig2=%9.7f;',X1mu,X1sig,X1sig2);
fprintf('\nX1mucor=X1mu+X1sig2/2=%8.6f;;',X1mucor);
% LR every 2 days:
X2=log(S(3:2:n1))-log(S(1:2:n1-2));%Roughly half the sample size
n2=length(X2); fprintf('\nn2=%3i;',n2);% Every 2 days; dt=1/n2;
X2mean=mean(X2); X2var=var(X2,1); X2std=std(X2,1);%LR2008
X2mu=n2*X2mean; X2sig2=n2*X2var; X2sig=sqrt(X2sig2);%LR X2mu=mu-sig^2/2;
X2mucor=X2mu+X2sig2/2;
fprintf('\nX2mu=%8.6f; X2sig=%7.5f; X2sig2=%9.7f;',X2mu,X2sig,X2sig2);
fprintf('\nX2mucor=X2mu+X2sig2/2=%8.6f;;',X2mucor);
% Bayesian Asymptotic estimates:
```

```
meanhat1=X1mean; meanhat2=X2mean;
muhat1=n1*meanhat1; muhat2=n2*meanhat2;%Correct for time-steps
muhatdif=100*(muhat2/muhat1-1);
fprintf('\nmuhat1=%8.6f; muhat2=%8.6f; muhatdif=%5.2f%%;'...
    ,muhat1,muhat2,muhatdif);
varhat1=X1var; varhat2=X2var;
sig2hat1=n1*varhat1; sig2hat2=n2*varhat2;%Correct for time-steps
sig2hatdif=100*(sig2hat2/sig2hat1-1);
fprintf('\nsig2hat1=%9.7f; sig2hat2=%9.7f; sig2hatdif=%5.2f%%;'...
    ,sig2hat1,sig2hat2,sig2hatdif);
SEmean1=X1std/sqrt(n1); SEvar1=X1var*sqrt(2/n1);
SEmean2=X2std/sqrt(n2); SEvar2=X2var*sqrt(2/n2);
SEmeandif=100*(SEmean2/SEmean1-1); SEvardif=100*(SEvar2/SEvar1-1);
fprintf('\nSEmean1=%8.6f; SEmean2/2=%8.6f; SEmeandif=%5.2f%%;'...
    ,SEmean1,SEmean2,SEmeandif);
fprintf('\nSEvar1=%10.8f; SEvar2/2=%10.8f; SEvardif=%5.2f%%;'...
    ,SEvar1,SEvar2,SEvardif);
%
scrsize = get(0,'ScreenSize'); % figure spacing for target screen
ss = [5.0,4.5,4.0,3.5]; % figure spacing factors
figure(1); nfig = 1;
mean10=meanhat1-2*SEmean1;
mean1f=meanhat1+2*SEmean1;
dmean1=(mean1f-mean10)/(100-1);
```

```
mean1plot=mean10:dmean1:mean1f;
mean20=meanhat2-2*SEmean2;
mean2f=meanhat2+2*SEmean2;
dmean2=(mean2f-mean20)/(100-1);
mean2plot=mean20:dmean2:mean2f;
mu1plot=mean1plot*n1; mu2plot=mean2plot*n2;
fmean1=normpdf(mean1plot,meanhat1,SEmean1); %normal approx. mean;
fmean2=normpdf(mean2plot,meanhat2,SEmean2); %but plot vs LR mu=mean/dt
plot(mu1plot,fmean1,'-b',mu2plot,fmean2,'--k','LineWidth',2);
axis tight;
title('Bayes Normal Test: \mu=mean_k*n_k estimate'...
    ,'Fontsize',24,'FontWeight','Bold');
xlabel('\mu=mean_k*n_k','Fontsize',24,'FontWeight','Bold');
ylabel('fpost_k(mean_k)','Fontsize',24,'FontWeight','Bold');
legend('fpost_1(mean_1), n_1=253','fpost_2(mean_2), n_2=126'...
    ,'Location','SouthEast');
set(gca,'Fontsize',20,'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]
%
figure(2); nfig = 2;
var10=varhat1-2*SEvar1;
var1f=varhat1+2*SEvar1;
dvar1=(var1f-var10)/(100-1);
```

```
var1plot=var10:dvar1:var1f;
var20=varhat2-2*SEvar2;
var2f=varhat2+2*SEvar2;
dvar2=(var2f-var20)/(100-1);
var2plot=var20:dvar2:var2f;
sig21plot=var1plot*n1; sig22plot=var2plot*n2;
fvar1=normpdf(var1plot,varhat1,SEvar1); %normal approx. var;
fvar2=normpdf(var2plot,varhat2,SEvar2); %but plot vs LR sig2=var/dt
plot(sig21plot,fvar1,'-b',sig22plot,fvar2,'--k','LineWidth',2);
axis tight;
title('Bayes Normal Test: \sigma^2=var_k*n_k estimate'...
    ,'Fontsize',24,'FontWeight','Bold');
xlabel('\sigma^2=var_k*n_k','Fontsize',24,'FontWeight','Bold');
ylabel('fpost_k(var_k)','Fontsize',24,'FontWeight','Bold');
legend('fpost_1(var_1), n_1=253','fpost_2(var_2), n_2=126'...
    ,'Location','NorthWest');
set(gca,'Fontsize',20,'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]
fprintf('\n ');
```

```
================= OUTPUT ========================
BayesNormalTest Output (20-Feb-2010 17:04:49):
n1=NLR=253;
X1mean=0.001921; X1std=0.02578; X1var=0.0006645;
X1mu=0.485902; X1sig=0.41001; X1sig2=0.1681075;
X1mucor=X1mu+X1sig2/2=0.569956;;
n2=126;
X2mu=0.471359; X2sig=0.32445; X2sig2=0.1052685;
X2mucor=X2mu+X2sig2/2=0.523993;;
muhat1=0.485902; muhat2=0.471359; muhatdif=-2.99%;
sig2hat1=0.1681075; sig2hat2=0.1052685; sig2hatdif=-37.38%;
SEmean1=0.001621; SEmean2/2=0.002575; SEmeandif=58.89%;
SEvar1=0.00005908; SEvar2/2=0.00010526; SEvardif=78.17%;
>>
```

○ *8.1.6. Likelihood and Prior Conjugate Pairs:*

When given a likelihood distribution and a prior distribution such that the **posterior distribution is of the same type as the prior distribution**, then the likelihood and prior are considered **conjugate pairs** or the prior is said to be the **conjugate distribution to the likelihood distribution**. The big advantage of finding a conjugate pair then the Bayes theorem calculations are usually much simpler and very convenient than that of nonconjugate pairs.

In the binomial likelihood example in subsection **8.1.1**, pp. 3-6, a beta prior produced a beta posterior, i.e., a Binomial-Beta conjugate pair or symbolically,

$$\mathbf{PosteriorBeta} \propto \mathbf{LikelihoodBinomial} \times \mathbf{PriorBeta}. \quad (8.67)$$

In the Poisson likelihood example in subsection **8.1.3**, pp. 13-16, we used a Poisson-Gamma conjugate pair,

$$\mathbf{PosteriorGamma} \propto \mathbf{LikelihoodPoisson} \times \mathbf{PriorGamma}. \quad (8.68)$$

In the normal likelihood we considered several cases, depending if the precision was known or the mean was known or both were the unknown. Each was a different conjugate problem. In **8.14(a)**, a conjugate Normal-Normal pair was used for fixed precision $\xi_1 = 1/\sigma_1^2$,

$$\mathbf{PosteriorNormal} \propto \mathbf{LikelihoodNormal}|_{\xi_1} \times \mathbf{PriorNormal}. \quad (8.69)$$

In **8.14(b)**, a conjugate Normal-Gamma pair was used for fixed mean θ_0,

$$\mathbf{PosteriorGamma} \propto \mathbf{LikelihoodNormal}|_{\theta_0} \times \mathbf{PriorGamma}. \quad (8.70)$$

In **8.14(c)**, with both mean and precision unknown, the situation became too complicated since we did not have a conjugate pair. However, in a Wikipedia table of conjugate pairs, ***Wikipedia: Conjugate Prior,***[a] a normal_gamma hybrid distribution is given as the conjugate distribution, so

$$\begin{aligned}\mathbf{PosteriorNormal\_Gamma} \propto\ & \mathbf{LikelihoodNormal} \\ & \times \mathbf{PriorNormal\_Gamma}.\end{aligned} \quad (8.71)$$

---

[a]For even for information, see the background paper by D. Fink (1997), ***A Compendium of Conjugate Priors.***

Some other conjugate pairs of interest are

$$\mathbf{PosteriorPareto} \propto \mathbf{LikelihoodUniform} \times \mathbf{PriorPareto}; \quad (8.72)$$

$$\mathbf{PosteriorGamma} \propto \mathbf{LikelihoodExponential} \times \mathbf{PriorGamma}. \quad (8.73)$$

Other advice on prior distributions:[a]

- They should be proper densities, in that they are integrable with other moments.
- They should be unbiased such that they do not appear to constrain the posterior distribution to a particular type or estimate.
- They should not produce unrealistic estimates for the unknown parameter or any parameter which is derived from it, e.g., a precision which leads to a peculiar variance.

[a] See Rice (2007) for more discussion.

∘ *8.1.7. The Bayesian Normal Likelihood with Proper Conjugate Pair:*

Applying the Normal-Gamma prior distribution, combining partial problems 8.14(a) and (b), to the Bayesian normal estimation problem for unknown mean and precision, bivariate parameters, $\vec{\theta} = (\mu, \xi)$,

$$f_{\vec{\Theta}}^{(\text{prior},4)}\left(\vec{\theta}\right) = f_M^{(\text{prior},4)}(\mu) f_{\Xi}^{(\text{prior},4)}(\xi)$$
$$= \sqrt{\frac{\xi\gamma}{2\pi}} \exp\left(-0.5\xi\gamma(\mu - \mu_0)^2\right) \frac{\lambda^{\alpha}\xi^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda\xi}, \tag{8.74}$$

with prior parameters $(\mu_0, \gamma, \lambda, \alpha)$, as conjugate to the total data likelihood

$$f_{\vec{X}|M,\Xi}(\vec{x}|\mu, \xi) = \sqrt{\frac{\xi}{2\pi}} \exp\left(-0.5\xi\left(S_n^2 + \left(\overline{X}_n - \mu\right)^2\right)\right). \tag{8.75}$$

The target template for the Normal-Gamma posterior distribution is the conjugate closure,

$$f_{\Theta|\vec{X}}^{(\text{post},4)}(\theta|\vec{x}) = \sqrt{\frac{\xi\widetilde{\gamma}}{2\pi}} e^{-0.5\xi\widetilde{\gamma}(\mu-\widetilde{\mu})^2} \frac{\widetilde{\lambda}^{\widetilde{\alpha}}\xi^{\widetilde{\alpha}-1}}{\Gamma(\widetilde{\alpha})} e^{-\widetilde{\lambda}\xi}, \tag{8.76}$$

with hyperparameters $(\widetilde{\mu}, \widetilde{\gamma}, \widetilde{\lambda}, \widetilde{\alpha})$ for the distribution of Θ and the Bayesian value estimate $\widehat{\Theta} = (\widehat{\mu}, \widehat{\xi})$.

However from the prior and likelihood, we have

$$f^{(\mathrm{post},4)}_{\Theta|\vec{X}}(\theta|\vec{x}) \propto \xi^{n/2+\alpha-1}(\xi\gamma)^{1/2} \times e^{-\frac{n\xi}{2}\left(S_n^2+\left(\overline{X}_n-\mu\right)^2\right)-\frac{\xi\gamma}{2}(\mu-\mu_0)^2-\lambda\xi}, \quad (8.77)$$

where right away we see $\widetilde{\alpha}=\widetilde{\alpha}_n=n/2+\alpha$ and $\widetilde{\gamma}=\widetilde{\gamma}_n=n+\gamma$ from the coefficient of μ^2 in the exponent. Let ϕ be the exponent of the exponential then

$$\begin{aligned}\phi &= -\frac{n\xi}{2}\left(S_n^2+\left(\overline{X}_n-\mu\right)^2\right)-\frac{\xi\gamma}{2}(\mu-\mu_0)^2-\lambda\xi \\ &= -\frac{n\xi}{2}\left(\mu^2-2\overline{X}_n\mu+\overline{X}_n^2+S_n^2\right)-\frac{\xi\gamma}{2}\left(\mu^2-2\mu_0\mu+\mu_0^2\right)-\lambda\xi \quad (8.78) \\ &\overset{\mathrm{cts}}{=} -\frac{\xi\widetilde{\gamma}}{2}(\mu-\widetilde{\mu})^2-\xi\left(\lambda+\frac{n}{2}S_n^2+\frac{n\gamma}{2(n+\gamma)}\left(\overline{X}_n-\mu_0\right)^2\right),\end{aligned}$$

where the following hyperparameters have been used:

$$\widetilde{\mu}=\widetilde{\mu}_n=\frac{n\overline{X}_n+\gamma\mu_0}{n+\gamma}\sim\overline{X}_n \quad (8.79)$$

and

$$\widetilde{\lambda}=\widetilde{\lambda}_n=\lambda+\frac{n}{2}S_n^2+\frac{n\gamma}{2(n+\gamma)}\left(\overline{X}_n-\mu_0\right)^2\sim\frac{n}{2}S_n^2, \quad (8.80)$$

which we sought to show to specify the target posterior (8.76).

Previously, presented the problem from Rice (2007), but this derivation failed to close the posterior with a proper conjugate pair to obtain a posterior of the same family of distributions and seemed have been missing a full set of proper prior parameters, namely for the precision. See Fink (1997) for background for this part of the lecture, but watch out for serious typos in the introduction.

** Reminder: Lecture 8 Homework Posted in Chalk Assignments,*
due in PDF by Lecture 9 in Chalk Assignments!

*** Summary of Lecture 8:**

1. **Bayesian Estimation Introduction**
2. **Simple Bayesian Counting Example**
3. **Bayesian Estimation Summary**
4. **Poisson Bayesian Estimation**
5. **Normal Bayesian Estimation in 3 Variations**
6. **Bayesian Large Sample Limit $\longrightarrow$ MLE**
7. **Bayesian Normal Large Sample Limit**
8. **Bayesian Normal with Proper Conjugate Priors**