§1.1 Random experiment, sample space, outcome, event, relative frequency

§1.2 Set Theory:
- Set, element, subset, empty set or null set $\emptyset$
- Union, intersection, complement
- Venn diagram, space $C$ or $D$
- DeMorgan’s laws: $(C_1 \cap C_2)^c = C_1^c \cup C_2^c$, $(C_1 \cup C_2)^c = C_1^c \cap C_2^c$
- Distributive laws: $C_1 \cap (C_2 \cup C_3) = (C_1 \cap C_2) \cup (C_1 \cap C_3)$, $C_1 \cup (C_2 \cap C_3) = (C_1 \cup C_2) \cap (C_1 \cup C_3)$
- Limit of a sequence of sets $C_1, C_2, C_3, \ldots$:
  \[ \lim_{k \to \infty} C_k = \bigcap_{k=1}^{\infty} C_k, \text{ if the sequence is non-increasing}; \]
  \[ \lim_{k \to \infty} C_k = \bigcup_{k=1}^{\infty} C_k, \text{ if the sequence is non-decreasing} \]

§1.3 The Probability Set Function
- $\sigma$-Field $\mathcal{B}$ of the sample space $\mathcal{C}$
  - Smallest $\sigma$-Field: $\{\emptyset, \mathcal{C}\}$
  - Greatest $\sigma$-Field: $2^\mathcal{C}$, the power set of $\mathcal{C}$
  - $\sigma$-Field generated by $D$: $\sigma(D) = \bigcap \{\mathcal{E} : D \subset \mathcal{E} \text{ and } \mathcal{E} \text{ is a } \sigma\text{-field} \}$
  - Borel $\sigma$-field: $\mathcal{B}_0$, the $\sigma$-field generated by the set of all open intervals
- Probability set function $P$ defined on the $\sigma$-field $\mathcal{B}$:
  - $P(C) \geq 0$, for all $C \in \mathcal{B}$
  - $P(\emptyset) = 0$
  - $P(C) = 1$
  - $P(\bigcup_{n=1}^{\infty} C_n) = \sum_{n=1}^{\infty} P(C_n)$, if $C_1, C_2, C_3, \ldots$ are mutually disjoint

  Properties of the probability set function $P$:
  - $P(C^c) = 1 - P(C)$, for all $C \in \mathcal{B}$
  - $P(\emptyset) = 0$
  - $P(C_1) \leq P(C_2)$, if $C_1 \subset C_2$
  - $0 \leq P(C) \leq 1$, for all $C \in \mathcal{B}$
  - $P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$
  - $\lim_{n \to \infty} P(C_n) = P(\lim_{n \to \infty} C_n)$, if $\{C_n\}$ is increasing or decreasing
  - $P(\bigcup_{n=1}^{\infty} C_n) \leq \sum_{n=1}^{\infty} P(C_n)$, for arbitrary sequence $\{C_n\}$

- Inclusion-Exclusion formula: $P(C_1 \cup C_2 \cup C_3) = p_1 - p_2 + p_3$
  where $p_1 = P(C_1) + P(C_2) + P(C_3)$,$$
  p_2 = P(C_1 \cap C_2) + P(C_1 \cap C_3) + P(C_2 \cap C_3),
  p_3 = P(C_1 \cap C_2 \cap C_3)$
§1.4 Conditional Probability and Independence

- Permutations and combinations: Draw \( k \) elements from \( n \) elements
  
  With order and with replacement: \( n^k \);
  
  With order and without replacement: \( P_k^n = n!/(n-k)! = \binom{n}{k} k! \);
  
  Without order and without replacement: \( \binom{n}{k} = n!/[k!(n-k)!] \)

§1.4 Conditional Probability and Independence

- Conditional probability: \( P(C_2|C_1) = P(C_1 \cap C_2)/P(C_1) \), if \( P(C_1) > 0 \)
  
  \( P(C_2|C_1) \geq 0 \);
  
  \( P(C_1 \cup C_2 \cup C_3 \cup \cdots |C) = P(C_1|C) + P(C_2|C) + P(C_3|C) + \cdots \), if
  
  \( C_1, C_2, C_3, \ldots \) are mutually disjoint;
  
  \( P(C_1 \cap C_2) = P(C_1)P(C_2|C_1) \);
  
  \( P(C_1 \cap C_2 \cap C_3) = P(C_1)P(C_2|C_1)P(C_3|C_1 \cap C_2) \)

- Law of total probability and Bayes’ theorem:
  
  \[
P(C) = \sum_{i=1}^k P(C_i)P(C|C_i),
  \]
  
  \[
P(C_j|C) = \frac{P(C_j)P(C|C_j)}{P(C)} = \frac{P(C_j)P(C|C_j)}{\sum_{i=1}^k P(C_i)P(C|C_i)},
  \]
  
  where \( \{C_1, C_2, \ldots, C_k\} \) is a partition of \( C \), and \( P(C_i) > 0, i = 1, \ldots, k \)

- The events \( C_1 \) and \( C_2 \) are independent, if and only if
  
  \( P(C_1 \cap C_2) = P(C_1)P(C_2) \).
  
  Then the following three pairs of events are independent:
  
  \( C_1 \) and \( C_1^\circ \); \( C_1^\circ \) and \( C_2 \); \( C_2^\circ \) and \( C_2^\circ \).

(updated on 10/26/2006)

- The events \( C_1, C_2, \ldots, C_n \) are independent, if and only if
  
  \( P(C_{d_1} \cap C_{d_2} \cap \cdots \cap C_{d_k}) = P(C_{d_1})P(C_{d_2}) \cdots P(C_{d_k}) \), for any \( 2 \leq k \leq n \)
  
  and any subset \( \{d_1, d_2, \ldots, d_k\} \) of \( \{1, 2, \ldots, n\} \). Therefore
  
  \( P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n) \), where \( A_i = C_i \) or \( C_i^\circ \).
  
  Those events with disjoint index sets are independent too.
  
  For example, \( C_1 \cup C_2^\circ, C_3^\circ \), and \( C_4 \cap C_5^\circ \) are independent.

§1.5 Random Variables

- Random variable \( X \): a function defined on the sample space \( \mathcal{C} \)
  
  Range (or space) of \( X \): \( \mathcal{D} = \{X(c) : c \in \mathcal{C}\} \)
  
- Cumulative distribution function (cdf) of a random variable \( X \):
  
  \[
  F(x) = P(X \leq x) = P(\{c \in \mathcal{C} : X(c) \leq x\})
  \]
  
  which always satisfies
  
  \( (a) \) \( F \) is nondecreasing, that is, \( F(a) \leq F(b) \) for all \( a < b \);
1.6 Discrete Random Variables

Other properties of the cdf \( F \) of \( X \):

- \( P(a < X \leq b) = F(b) - F(a) \);
- \( P(X = x) = F(x) - F(x-), \) where \( F(x-) = \lim_{x \uparrow x} F(z) \);
- Discrete random variable \( X \): \( D \) is finite or countable
  - Probability mass function (pmf) of \( X \):
    \[
    p(x) = P(X = x) = P\{\{c \in C : X(c) = x\}\}
    \]
    which must satisfy: [1] \( 0 \leq p(x) \leq 1 \) for all \( x \in D \); [2] \( \sum_{x \in D} p(x) = 1 \).
  - Continuous random variable \( X \): there exists a probability density function (pdf) \( f(x) \) such that the cdf
    \[
    F(x) = \int_{-\infty}^{x} f(t) dt, \text{ for all } x \in R.
    \]
    Note that \( f(x) \) must satisfy: [1] \( f(x) \geq 0 \) for all \( x \); [2] \( \int_{-\infty}^{\infty} f(x) dx = 1 \).
  - Properties of continuous random variable \( X \) with cdf \( F(x) \) and pdf \( f(x) \):
    - \( F(x) \) is continuous. Thus \( P(X = x) = F(x) - F(x-) = 0 \), for all \( x \).
    - \( \frac{d}{dx} F(x) = f(x) \), for almost all \( x \).
    - \( P(a < X \leq b) = P(a < X < b) = P(a \leq X \leq b) = \int_{a}^{b} f(x) dx \).

§1.6 Discrete Random Variables

- Uniform distribution on a finite set, for example, \( \{-2,-1,0,1,2\} \):
  - The pmf is
    \[
    \begin{array}{c|c|c|c|c|c}
    x & -2 & -1 & 0 & 1 & 2 \\
    p(x) & 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\
    \end{array}
    \]
  - Bernoulli trial:
    - \( C = \{ \text{success, failure} \} \), \( P(\{\text{success}\}) = p, P(\{\text{failure}\}) = 1 - p \),
    - where \( p \) is the parameter of the Bernoulli trial such that \( 0 < p < 1 \)
    - Bernoulli distribution:
      \( X(\text{success}) = 1, X(\text{failure}) = 0 \), pmf:
      \[
      \begin{array}{c|c|c}
      x & 0 & 1 \\
    p(x) & 1 - p & p \\
    \end{array}
    \]
  - Geometric distribution:
    - Repeat a Bernoulli trial independently until a success appears.
    - Let \( X \) be the number of trials needed.
    - The range of \( X \): \( D = \{1, 2, 3, \ldots, n, \ldots\} \)
    - The pmf of \( X \): \( p(x) = (1 - p)^{x-1} p \), for \( x = 1, 2, 3, \ldots \),
    - which has the property “lack-of-memory”:
      \[
      P(X = k + m | X > k) = P(X = m)
      \]
Hypergeometric distribution:
Suppose an urn contains $N$ balls. Exactly $M$ balls are red. Draw $n$ ($n < N - M$) balls from the urn at random and without replacement. Let $X$ be the number of red balls drawn.
The range of $X$: $D = \{0, 1, 2, \ldots, \min\{n, M\}\}$
The pmf of $X$:
$$p(x) = \binom{M}{x} \binom{N-M}{n-x} / \binom{N}{n}$$

Transformation of a discrete random variable $X$: $Y = g(X)$
$Y$ is a discrete random variable too. The pmf of $Y$:
$$p_Y(y) = \sum_{x: g(x) = y} p_X(x)$$

§1.7 Continuous Random Variables

– Uniform distribution on a finite interval $(a, b)$: The pdf is
$$f(x) = \begin{cases} 1/(b-a), & \text{if } x \in (a, b); \\ 0, & \text{elsewhere} \end{cases}$$

– Cauchy distribution: $f(x) = 1/[\pi(1+x^2)]$, $-\infty < x < \infty$

– Transformation of a continuous random variable $X$: $Y = g(X)$
$Y$ is a continuous random variable too. The pdf of $Y$:
$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|, \text{ for } y \in S_Y = \{g(x): x \in S_X\},$$
if $g(x)$ is a one-to-one differentiable function on $S_X$, the support of $X$.
Alternative approach: (1) Calculate the cdf of $Y$ first
$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

(2) The pdf of $Y$: $f_Y(y) = \frac{dF_Y(y)}{dy}$, for $y \in S_Y = \{g(x): x \in S_X\}$

– Mode of a distribution: a value of $x$ that maximizes the pdf or pmf

– Median of a distribution: a value of $x$ such that
$$P(X \leq x) \geq \frac{1}{2}, \ P(X \geq x) \geq \frac{1}{2}$$

§1.8 Expectation of a Random Variable
– Expectation of a discrete random variable X, if $\sum_x |x|p(x) < \infty$:

$$E(X) = \sum_x xp(x)$$

Expectation of $g(X)$, if $\sum_x |g(x)|p(x) < \infty$:

$$E[g(X)] = \sum_x g(x)p(x)$$

– Expectation of a continuous random variable X, if $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Expectation of $g(X)$, if $\int_{-\infty}^{\infty} |g(x)|f(x)dx < \infty$:

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

– Properties of expectations:

1. $E(c) = c$, if $c$ is a constant;
2. $E(cX) = cE(X)$, if $c$ is a constant;
3. $E(aX + bY) = aE(X) + bE(Y)$, if $a, b$ are constants

§1.9 Some Special Expectations

– Mean: $\mu = E(X)$
– Variance: $\sigma^2 = Var(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$
  
  Standard deviation: $\sigma = \sqrt{Var(X)}$
– Skewness: $\gamma_1 = E[(X - \mu)^3]/\sigma^3 = E(X^3)/\sigma^3 - 3\mu/\sigma - (\mu/\sigma)^3$
  
  $\gamma_1 < 0$ (skewed to the left); $\gamma_1 > 0$ (skewed to the right);
  
  $\gamma_1 = 0$ (not skewed)
– Moments: $E(X^m)$, $m$th moment; $E[(X - \mu)^m]$, $m$th central moment
– Moment generating function (mgf): If $E(e^{tX})$ exists for $t \in (-h, h)$,

$$M(t) = E(e^{tX}) = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \cdots$$

$M'(0) = \mu$, $M''(0) = E(X^2)$, \ldots In general, $M^{(m)}(0) = E(X^m)$ .

– If $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$, then $F_X(z) = F_Y(z)$ for all $z \in R$. That is, $X$ and $Y$ have the same distribution, denoted by $X \overset{D}{=} Y$ .
– Characteristic function (cf): $\varphi(t) = E(e^{itX}) = M(it)$, where $i = \sqrt{-1}$ is the imaginary unit. Note that cf exists for all real $t$. 

---

5
Cumulant generating function: \( \psi(t) = \log M(t) \)
\( \psi'(0) = \mu, \; \psi''(0) = \sigma^2 \)

§1.10 Important Inequalities

- Markov’s inequality: Let \( u(X) \) be a nonnegative function, \( c > 0 \),
  \[ P[u(X) \geq c] \leq \frac{E[u(X)]}{c} \]

- Chebyshev’s inequality: For \( k > 0 \),
  \[ P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \]
  or \( P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \) for all \( \epsilon > 0 \)

- Convex function \( \phi \): For all \( x, y \) and all \( 0 < \gamma < 1 \),
  \[ \phi[\gamma x + (1 - \gamma)y] \leq \gamma \phi(x) + (1 - \gamma)\phi(y) \]
  \( \phi \) is convex if \( \phi' \) is nondecreasing or \( \phi'' \) is nonnegative.

- Jensen’s inequality: If \( \phi \) is convex on the support of \( X \), then
  \[ \phi[E(X)] \leq E[\phi(X)] \]