# Probability, Random Processes, and Statistical Analysis

**Instructor's Solution Manual**

(Revision: April 5, 2012)

HISASHI KOBAYASHI, Princeton University
BRIAN L. MARK, George Mason University
WILLIAM TURIN, AT&T Bell Laboratories

CAMBRIDGE
UNIVERSITY PRESS

# 2 Solutions for Chapter 2: Probability

## 2.2 Axioms of Probability

**2.1* Tossing a coin three times.**

(a)

$$\Omega = \{(\text{hhh}), (\text{hht}), (\text{hth}), (\text{htt}), (\text{thh}), (\text{tht}), (\text{tth}), (\text{ttt})\}.$$

(b)

$$\begin{aligned}
E_0 &= \{(\text{ttt})\}, & |E_0| &= 1; \\
E_1 &= \{(\text{htt}), (\text{tht}), (\text{tth})\}, & |E_1| &= 3; \\
\\
E_2 &= \{(\text{hht}), (\text{hth}), (\text{thh})\}, & |E_2| &= 3; \\
E_3 &= \{(\text{hhh})\}, & |E_3| &= 1.
\end{aligned}$$

(c)

$$F = \{(\text{hhh}), (\text{hht}), (\text{hth}), (\text{thh})\}.$$

**2.2* Tossing a coin until "head" or "tail" occurs twice in succession.** There are countably infinite sample points.

$$\Omega = \{(\text{hh}), (\text{tt}), (\text{thh}), (\text{htt}), (\text{hthh}), (\text{thtt}), (\text{ththh}), (\text{hthtt}), \ldots\}.$$

As is seen, there are only two outcomes (or sample points) of string length $n = 2, 3, 4, 5, \ldots$.

**2.3 Placing distinguishable particles in different cells.**

(a) Each of particle can be placed in one of three cells. That is, particle $a$ can be placed in cell 1, 2, or 3; particle $b$ can be placed in cell 1, 2, or 3; and particle $c$ can be placed in cell 1, 2, or 3. Therefore, an equivalent representation of an outcome is a 3-tuple $(n_a, n_b, n_c)$, where $n_a, n_b, n_c \in \{1, 2, 3\}$, denote, respectively, the number of the cell occupied by particles $a$, $b$, and $c$. The sample space in this case is given by

$$\hat{\Omega} = \{(n_a, n_b, n_c) : n_a, n_b, n_c \in \{1, 2, 3\}\}.$$

One can see that

$$|\hat{\Omega}| = 3 \times 3 \times 3 = 3^3 = |\Omega| = 27.$$

The sample space $\Omega$, with samples shown also in the equivalent representation of $\hat{\Omega}$, is given as follows:

$$\begin{aligned}\Omega = \{&(abc|-|-) \equiv 111, (ab|c|-) \equiv 112, (ab|-|c) \equiv 113,\\ &(ac|b|-) \equiv 121, (a|bc|-) \equiv 122, (a|b|c) \equiv 123,\\ &(ac|-|b) \equiv 131, (a|c|b) \equiv 132, (a|-|bc) \equiv 133\\ &(bc|a|-) \equiv 211, (b|ac|-) \equiv 212, (b|a|c) \equiv 213,\\ &(c|ab|-) \equiv 221, (-|abc|-) \equiv 222, (-|ab|c) \equiv 223,\\ &(c|a|b) \equiv 231, (-|ac|b) \equiv 232, (c|ab|-) \equiv 233,\\ &(bc|-|a) \equiv 311, (b|c|a) \equiv 312, (b|-|ac) \equiv 313,\\ &(-|ab|a) \equiv 321, (-|bc|a) \equiv 322, (-|b|ac) \equiv 323,\\ &(c|-|ab) \equiv 331, (-|c|ab) \equiv 332, (-|-|abc) \equiv 333\}\end{aligned}$$

(b) Consider the set $A^c$ and its equivalent counterpart

$$\hat{A}^c = \{(n_a, n_b, n_c) \in \hat{\Omega} : n_a, n_b, n_c \text{ all different}\}.$$

One can recognize $\hat{A}^c$ as being equivalent to the set of permutations on three elements; hence $|\hat{A}^c| = 3! = 6 = |A^c|$ and

$$|A| = |\Omega| - |A^c| = 27 - 6 = 21.$$

Consider the set $B^c$ and its equivalent counterpart

$$\hat{B}^c = \{(n_a, n_b, n_c) : n_a, n_b, n_c \in \{2, 3\}\}.$$

Then one can see that $|\hat{B}^c| = 2 \times 2 \times 2 = 2^3 = 8 = |B^c|$; hence,

$$|B| = |\Omega| - |B^c| = 27 - 8 = 19.$$

Note that

$$C^c = (A \cap B)^c = A^c \cup B^c,$$

and furthermore, that $A^c$ and $B^c$ are disjoint. Therefore, $|C^c| = |A^c| + |B^c| = 6 + 8 = 14$ and $|C| = |\Omega| - |C^c| = 27 - 14 = 13$.

**2.4 Placing indistinguishable particles in different cells.**

(a) Consider five items of two kinds; one is three indistinguishable particles, denoted by "star ($*$)"; the other is two indistinguishable walls between cells, denoted by "bar ($|$)". Then the different arrangements in the original problem is equivalent to the number of ways of choosing the three locations for the particles from the five locations for either bars or stars. There are $\binom{5}{2} = \frac{5!}{2!3!} = 10$ different arrangements.

$$\begin{aligned}&\omega_1 = \{***| - | -\},\ \omega_2 = \{-|***|-\},\ \omega_3 = \{-| - |***\}\\ &\omega_4 = \{** \mid * \mid -\},\ \omega_5 = \{** \mid - \mid *\},\ \omega_6 = \{*|**|-\}\\ &\omega_7 = \{*| - |**\},\ \omega_8 = \{-|**|*\},\ \omega_9 = \{-| * |**\}\\ &\omega_{10} = \{*| * |*\}.\end{aligned}$$

(b) $\tilde{A}$ contains 9 sample points; $\tilde{B}$ contains 6 sample points, and $\tilde{C} = \tilde{A} \cap \tilde{B}$ contains 5 sample points. It so happens that $\tilde{A} \cup \tilde{B} = \Omega$ in this case, as well.
See Feller [2], pp. 9-10 for a further discussion.

**2.5*** **Probability assignment to the coin tossing experiment.** Assuming the coin tossing is fair, the probability measure should assign a probability of $1/8$ to each sample point in $\Omega$. I.e., $P[\{\omega\}] = 1/8$ for each $\omega \in \Omega$.

$$P[E_0] = \frac{1}{8},\ P[E_1] = \frac{3}{8},\ P[E_2] = \frac{3}{8},\ P[E_3] = \frac{1}{8}.$$

**2.6*** Probability assignment to the coin tossing experiment in Exercise 2.2.

(a) Since the coin is fair and tosses are independent, we have

$$P[\{(\mathrm{hh})\}] = P[\{(\mathrm{tt})\}] = \frac{1}{4},\ \mathrm{P}[\{(\mathrm{thh})\}] = \mathrm{P}[\{(\mathrm{htt})\}] = \frac{1}{8},$$
$$P[\{\mathrm{hthh}\}] = P[\{(\mathrm{thtt})\}] = \frac{1}{16},\ \ \ldots.$$

In general, to each of the two possible outcomes or sample points requiring $n$ tosses we assign $\frac{1}{2^{n-1}}$.

(b)

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = \frac{15}{16}.$$

(c)

$$\frac{1}{2} + \frac{1}{8} + \frac{1}{32} + \ldots = \frac{1}{2}\frac{1}{1-\frac{1}{4}} = \frac{2}{3}.$$

**2.7** **Placing distinguishable particles in cells: Maxwell-Boltzmann statistics.**

(a) The appropriate probability measure assigns the same probability to each sample point in $\Omega$:

$$P[\{\omega\}] = \frac{1}{|\Omega|} = \frac{1}{27},\ \text{ for all } \omega \in \Omega.$$

(b) The event that each cell contains exactly one cell is equivalent to the complement of the set $B$ defined in Exercise 2.3(b). From Exercise 2.3(b), we know that $|B^c| = 6$. Therefore,

$$p = \frac{|B^c|}{27} = \frac{6}{27} = \frac{2}{9}.$$

(c) Label the particles as $1, \ldots, r$ and let $n_i$ denote the cell in which particle $i$ is places, $1 \le i \le r$. Then the sample space is equivalent to

$$\Omega = \{(n_1, \ldots, n_r) : n_i \in \{1, \ldots, n\}, 1 \le i \le r\}.$$

Clearly, $|\Omega| = n^r$. We are interested in the event

$$D = \{(n_1, \ldots, n_r) : n_1, \ldots, n_r \text{ are all different}\},$$

which is equivalent to the set of permutations on $r$ objects. Hence, $|D| = r!$ and

$$P[D] = \frac{|D|}{|\Omega|} = \frac{r!}{n^r}.$$

**2.8 Placing indistinguishable particles in cells: Bose-Einstein statistics.**

(a) As found in Exercise 2.4, there are 10 sample points, we assign probability 1/10 to each of them.

(b) Since there is only one such arrangement, $p = 1/10$.

(c) The number of sample points in the sample space can be calculated using star/wall construction used in the solution to 2.4(a), and is: $\binom{r+n-1}{r} = \frac{(r+n-1)!}{r!(n-1)!}$. Then the event that each of the chosen $r$ cells contains one particle is one of the above simple events. Thus,

$$\binom{r+n-1}{r}^{-1}.$$

**2.9** The sample space can be represented as follows:

$$\Omega = \{(m_1, \ldots, m_n) : m_1 + \cdots + m_n = r;\ m_i \in \{0, 1\}, 1 \leq i \leq n\}.$$

One can see the each sample point $\omega \in \Omega$ can be determined by choosing $r$ elements from the set $\{1, \ldots, n\}$ and then setting the corresponding values of $m_i$ equal to 1. Hence,

$$|\Omega| = \binom{n}{r},$$

and the probability of each sample point $\omega \in \Omega$ is given by

$$P[\{\omega\}] = \frac{1}{|\Omega|} = \binom{n}{r}^{-1}.$$

## 2.3 Bernoulli Trial's and Bernoulli's Theorem

**2.10*** **Distribution laws and Venn diagram.** Draw a Venn diagram in which the areas $A, B$ and $C$ intersect each other.

**2.11*** **DeMorgan's law.** Draw a Venn diagram in which the areas $A$ and $B$ intersect. Then we readily see that the areas $A \cap B$ and $A^c \cup B^c$ are the complements of each other in the diagram. A formal proof is as follows:
Suppose that $\omega \in (A \cap B)^c$. Then $\omega$ does *not* belong to *both* $A$ and $B$. This implies that either $\omega$ belongs to $A^c$ or $\omega$ belongs to $B^c$; i.e., $\omega \in A^c \cup B^c$. Hence, $(A \cap B)^c \subseteq A^c \cup B^c$.
Conversely, suppose that $\omega \in A^c \cup B^c$. Then $\omega$ belongs to either $A^c$ or $B^c$. In the former case, $\omega$ does not belong to $A$, which implies that $\omega \notin A \cap B$. In the latter case, $\omega$ does not belong to $B$, which also implies that $\omega \notin A \cap B$. Hence, $\omega \in (A \cap B)^c$, so $A^c \cup B^c \subseteq (A \cap B)^c$. This establishes that $(A \cap B)^c = A^c \cup B^c$.

**2.12** **Axiom 3.**

(a) $P[A_1 \cup A_2] = P[A_1] + P[A_2]$ (from Property 3)

(b) Assume that the equation is true for some $M = N (\geq 2)$, i.e.,

$$P[\bigcup_{m=1}^{N} A_m] = \sum_{m=1}^{N} P[A_m].$$

Letting $B = \bigcup_{M=1}^{N} A_m$, we can write $\bigcup_{m=1}^{N+1} = A_{N+1} \cup B$. Then

$$P\left[\bigcup_{m=1}^{N+1}\right] = P\left[A_{N+1} \cup B\right] = P[A_{N+1}] + P[B],$$

since $A_{N+1}$ and $B$ are mutually exclusive. Using the assumed property

$$P[B] = \sum_{m=1}^{N} P[A_m],$$

we obtain

$$P\left[\bigcup_{M=1}^{N+1} A_m\right] = \sum_{m=1}^{N+1} P[A_m].$$

Thus, we have shown that the equation holds for $M = N + 1$ as well.

**2.13 Closure property of $\sigma$-field.**

$$\begin{aligned} A_m \in \mathcal{F}, m = 1, 2, \ldots &\Longrightarrow A_m^c \in \mathcal{F}, m = 1, 2, \ldots \quad \text{(Property (b))}. \\ &\Longrightarrow \bigcup_{m=1}^{\infty} A_m^c \in \mathcal{F}. \quad \text{(Property (c))}. \end{aligned}$$

Taking the complement and using de Morgan's law,

$$\left(\bigcup_{m=1}^{\infty} A_m^c\right)^c = \bigcap_{m=1}^{\infty} A_m$$

which must be also a member of $\mathcal{F}$ because of Property (b). Hence the $\sigma$-field is closed under countable intersection.

**2.14* Derivation of (2.48).**

$$\begin{aligned} &\sum_{k=0}^{n}(k - np)^2 \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n}(k^2 - 2npk + n^2p^2) \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} - 2np \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\ &\quad + n^2p^2 \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned} \tag{1}$$

In the last equation, the second summation term can be easily shown to equal $np$, and the third summation is clearly equal to $[p + (1-p)]^n = 1$. The first summation term is evaluated below:

$$\begin{aligned}
&\sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np \sum_{j=0}^{n-1} (j+1) \binom{n-1}{j} p^j (1-p)^{n-1-j} \\
&= np \cdot [(n-1)p + 1].
\end{aligned}$$

Returning to (1), we have

$$\begin{aligned}
&\sum_{k=0}^{n} (k-np)^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= np[(n-1)p+1] - 2n^2p^2 + n^2p^2 = np[(n-1)p + 1 - np] = np(1-p).
\end{aligned}$$

**2.15 Multinomial coefficient.**

**Approach 1:** The $n$ distinguishable objects can be divided into the $r$ groups as follows: Choose $k_1$ objects out of $n$. This can be done in $\binom{n}{k_1}$ ways. Then choose $k_2$ objects out of $n - k_1$ remaining objects. This can be done in $\binom{n-k_1}{k_2}$ ways. Continue in this manner until we have chosen $k_{r-1}$ objects out of $n - k_1 - \cdots - k_{r-2}$ remaining , which can be done in $\binom{n-k_1-\cdots-k_{r-2}}{k_{r-1}}$ ways. Hence, the number of ways of determining the partition is given by

$$\begin{aligned}
&\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-\sum_{i=1}^{r-2} k_i}{k_{r-1}} \\
&= \frac{n!}{k_1!(n-k_1)!} \cdot \frac{(n-k_1)!}{k_2!(n-k_1-k_2)!} \cdots \frac{\left(n-\sum_{i=1}^{r-2} k_i\right)!}{k_{r-1}!\left(n-\sum_{i=1}^{r-1} k_i\right)!} \\
&= \frac{n!}{k_1!k_2!\cdots k_r!}
\end{aligned}$$

**Approach 2:** The $n$ distinguishable objects can be divided into the $r$ groups as follows: Obtain a permutation of the $n$ objects. Choose the first $k_1$ objects to form the first group, the second $k_2$ to for the second group, . . ., the last $k_r$ to form the $r$th group. The number of ways of doing this is $n!$. But the partitioning into $r$ groups does not distinguish the between the $k_1!$ different orderings of the objects in the first group, or the $k_2!$ orderings of the second group, ..., or the $k_r!$ orderings of the $r$th group. Therefore, the total number of partitions is

$$\frac{n!}{k_1!k_2!\cdots k_r!}$$

## 2.4 Conditional Probability, Bayes' Formula, and Statistical Independence

**2.16*** **Joint probabilities.**

$$\sum_{m=1}^{M}\sum_{n=1}^{N} f_N(A_m, B_n) = 1.$$

where $f_N(A_m, A_n)$ is given by (2.42). The above formula readily follows from the relation:

$$\sum_{m=1}^{M}\sum_{n=1}^{N} N(A_m, B_n) = N.$$

**2.17*** **Proof of Bayes' theorem.** The joint probability $P[A_j, B]$ can be written as

$$P[A_j, B] = P[A_j|B]P[B] = P[A_j]P[B|A_j],$$

from which we have

$$P[A_j|B] = \frac{P[A_j]P[B|A_j]}{P[B]}.$$

The marginal probability $P[B]$ can be expressed, in terms of probabilities $P[A_j]$ and the conditional probabilities $P[B|A_j]$ $(j = 1, 2, \ldots, n)$, as in (2.61). Then (2.63) ensues.

**2.18*** **Independent events.** Since $A$ and $B$ are independent,

$$P[A, B] = \frac{1}{12} = P[A] \cdot P[B]. \tag{2}$$

We are also given that

$$P[(A \cup B)^c] = \frac{1}{3}. \tag{3}$$

We also know that

$$P[(A \cup B)^c] = 1 - P[A \cup B] = 1 - (P[A] + P[B] - P[A \cap B]). \tag{4}$$

Using (2) and (3) in (4), we can obtain the following equation in terms of $P[A]$:

$$1 - \left(P[A] + \frac{1}{12 \cdot P[A]} - \frac{1}{12}\right) = \frac{1}{3}. \tag{5}$$

This equation can be re-arranged as follows:

$$12(P[A])^2 - 9 \cdot P[A] + 1 = 0,$$

which is a quadratic equation. Applying the quadratic formula, we obtain two possible solutions:

$$P[A] \approx 0.6143 \text{ or} \tag{6}$$

$$P[A] \approx 0.1356. \tag{7}$$

Using (2), we can solve for $P[B]$. For the solution (6), $P[B] \approx 0.09104$. For the solution (7), $P[B] \approx 0.9154$. Thus, the values of $P[A]$ and $P[B]$ are approximately $0.1356$ and $0.6143$, respectively or vice versa.

**2.19* Medical test.**

(a)

$$P[A_2|B_2] = \frac{0.001 \times 0.99}{0.999 \times 0.05 + 0.001 \times 0.99} = \frac{0.00099}{0.05094} = 0.0194 = 1.94\%.$$

Hence, $P[A_1|B_2] = 98.06\%$.

(b)

$$P[A_2|B_2] = \frac{0.001 \times 0.95}{0.999 \times 0.01 + 0.001 \times 0.95} = \frac{0.00095}{0.01094} = 0.0868 = 8.68\%.$$

Hence $P[A_1|B_2] = 91.32\%$.

**2.20 Birthday problem.** The size of the sample space in this case is $|\Omega| = 365^r$. We are interested in the event $A$ that $r$ people have different birthdays. Hence,

$$|A| = 365 \times 364 \times \cdots \times (365 - r + 1),$$

so

$$\begin{aligned} P[A] &= \frac{365 \times 364 \times \cdots \times (365 - r + 1)}{365^n} \\ &= \left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{365 - r + 1}{365}\right) \\ &= \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{r-1}{365}\right). \end{aligned}$$

Taking the natural logarithms, and using the approximation $\ln(1 - x) \approx -x$, for $|x| \ll 1$, we have

$$\begin{aligned} \ln P[A] &\approx -\frac{1}{365} - \frac{2}{365} - \cdots \frac{r-1}{365} \\ &= -\frac{1}{365} \cdot \frac{r(r-1)}{2} = \frac{-r(r-1)}{730}. \end{aligned}$$

Hence,

$$P[A] \approx \exp\left[\frac{-r(r-1)}{730}\right].$$

For $r = 23$, we find that $P[A] \approx 0.500$ and for $r = 56$, we find that $P[A] \approx 0.0147$. The true probability for $r = 56$ is closer to 0.011; so we see the approximation begins to fail for large $r$ (as $\frac{r}{365}$ approaches 1).

**2.21 Web access pattern.**

(a) Let $k$ be the number of segments that observe an arrival and let $A$ be the event that at least 2 arrivals occur in the interval $T$. Then:

$$\begin{aligned} P[A] &= 1 - P[A^c] = 1 - P[\{k = 0\}] - P[\{k = 1\}] \\ &= 1 - (1 - p)^5 - \binom{5}{1} p^1 (1 - p)^5 \\ &= 1 - (0.8)^5 - 5(0.2)(0.8)^4 = 0.2627 \end{aligned}$$

(b) Let $B$ be the event that $n$ segments are required to see the first arrival. Then:

$$P[B] = (1-p)^{n-1}p = (0.8)^{n-1}(0.2).$$

(c) **Approach 1:** Given that $k$ of the $n$ Bernoulli trials are successes, the probability that a given trial (out of the $n$) is a success has to be $\frac{k}{n}$.

**Approach 2:** More formally, we can define the following events:

$$A_k = \{k \text{ successes in } n \text{ Bernoulli trials}\}$$
$$B_1 = \{\text{trial 1 results in success}\}.$$

We want to compute the conditional probability

$$P[B_1|A_k] = \frac{P[A_k, B_1]}{P[A_k]}. \tag{8}$$

We know that

$$P[A_k] = B(n;k,p) = \binom{n}{k} p^k (1-p)^{n-k}. \tag{9}$$

Also,

$$\begin{aligned} P[A_k, B] &= P[\{\text{success in trial 1; } k-1 \text{ successes in } n-1 \text{ other trials}\}] \\ &= p \cdot \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} = \binom{n-1}{k-1} p^k (1-p)^{n-k}. \end{aligned} \tag{10}$$

Applying (9) and (10) in (8), we obtain

$$P[B|A_k] = \frac{\binom{n-1}{k-1}}{\binom{n}{k}} = \frac{k}{n}.$$

# 3 Solutions for Chapter 3: Discrete Random Variables

## 3.1 Random Variables

**3.1*** **Property 4 of (3.3).** Suppose $b > a$. Then

$$\{X \le b\} = \{X \le a\} \cup \{a < X \le b\}.$$

Since the two events on the right-hand side are disjoint, we can apply Axiom 3 to obtain

$$P[X \le b] = P[X \le a] + P[a < X \le b],$$

or

$$P[a < X \le b] = P[X \le b] - P[X \le a] = F_X(b) - F_X(a).$$

**Another Answer:**
Let $A = \{X \le a\}$ and $B = \{a < X \le b\}$. Then $A$ and $B$ are mutually exclusive events, and thus $P[A \cup B] = P[A] + P[B]$. Since $A \cup B = \{X \le b\}$, we have

$$F_X(b) = F_X(a) + P[a < X \le b],$$

which leads to (**??**) .

## 3.2 Discrete Random Variables and Probability Distributions

**3.2*** **A nonnegative discrete RV.**

(a) From (3.16)

$$F_X(\infty) = \sum_{i=0}^{\infty} p_i = \frac{k}{1-\rho} = 1.$$

Hence, we find $k = 1$.

(b)

$$F_X(x) = \begin{cases} 0, & \text{for } x < 0 \\ 1 - \rho^{n+1}, & \text{for } \ n \le x \le n+1, \ \ n = 0, 1, 2, \ldots \end{cases}$$

**3.3*** **Statistical independence.** Suppose that (3.25) holds. Then

$$
\begin{aligned}
F_{XY}(x_i, y_j) &= P[X \le x_i, Y \le y_j] \\
&= \sum_{x \le x_i, y \le y_j} P[X = x, Y = y] = \sum_{x \le x_i, y \le y_j} p_{XY}(x, y) \\
&= \sum_{x \le x_i, y \le y_j} p_X(x) p_Y(y) \\
&= \left( \sum_{x \le x_i} p_X(x) \right) \left( \sum_{y \le y_j} p_Y(y) \right) = F_X(x_i) F_Y(y_j). \qquad (1)
\end{aligned}
$$

Thus, (3.25) implies (3.26).
Now assume that (3.26) holds. Define

$$x_{i-1} = \max_{x < x_i} x, \quad y_{j-1} = \max_{y < y_j} y.$$

Then

$$
\begin{aligned}
&p_X(x_i) p_Y(y_j) \\
&\quad = [F_X(x_i) - F_X(x_{i-1})][F_Y(y_j) - F_Y(y_{j-1})] \\
&\quad = F_X(x_i) F_Y(y_j) - F_X(x_{i-1}) F_Y(y_j) - F_X(x_i) F_Y(y_{j-1}) \\
&\quad\quad + F_X(x_{i-1}) F_Y(y_{j-1}) \\
&\quad = F_{XY}(x_i, y_j) - F_{XY}(x_{i-1}, y_j) - F_{XY}(x_i, y_{j-1}) \\
&\quad\quad + F_{XY}(x_{i-1}, y_{j-1}) \\
&\quad = P[X \le x_i, Y \le y_j] - (P[X \le x_{i-1}, Y \le y_j] \\
&\quad\quad + P[X \le x_i, Y \le y_{j-1}] - P[X \le x_{i-1}, Y \le y_{j-1}]) \\
&\quad = P[X \le x_i, Y \le y_j] - P[\{X \le x_{i-1}, Y \le y_j\} \cup \{X \le x_i, Y \le y_{j-1}\}] \\
&\quad = P[X = x_i, Y = y_j] = p_{XY}(x_i, y_j). \qquad (2)
\end{aligned}
$$

Thus, (3.26) implies (3.25).
We have already shown the equivalence of (3.27) and (3.28) for two discrete RVs. Proceeding by mathematical induction, assume the equivalence of (3.27) and (3.28) holds for $k \ge 2$ discrete RVs $X_1, X_2, \ldots, X_k$. Suppose that

$$p_{X_1 X_2 \cdots X_{k+1}}(x_1, x_2, \ldots, x_{k+1}) = p_{X_1}(x_1) p_{X_2}(x_2) \cdots p_{X_{k+1}}(x_{k+1}), \qquad (3)$$

for all values of $x_1, x_2, \ldots, x_{k+1}$. Using an argument similar to that used to obtain (1), we can show that

$$F_{X_1 X_2 \cdots X_{k+1}}(x_1, x_2, \ldots, x_{k+1}) = F_{X_1 X_2 \cdots X_k}(x_1, x_2, \ldots, x_k) F_{X_{k+1}}(x_{k+1}), \qquad (4)$$

for all values of $x_1, x_2, \ldots, x_{k+1}$. By invoking the induction hypothesis, we then obtain

$$F_{X_1 X_2 \cdots X_{k+1}}(x_1, x_2, \ldots, x_{k+1}) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_{k+1}}(x_{k+1}). \qquad (5)$$

Conversely, assume that (5) holds. Using an argument similar to that used to obtain (2), we can show that

$$p_{X_1X_2\cdots X_{k+1}}(x_1, x_2, \ldots, x_{k+1}) = p_{X_1X_2\cdots X_k}(x_1, x_2, \ldots, x_k)p_{X_{k+1}}(x_{k+1}). \tag{6}$$

Then by invoking the induction hypothesis, we establish (3).

**3.4 Discrete RVs, distribution function and conditional probability.**

(a) The RV $X$ takes values in the set $\{1, 2, 3\}$. The probability distribution of $X$ is as follows:

$$\begin{aligned}
P[X = 1] &= P[\{\omega_1, \omega_2\}] = \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \\
P[X = 2] &= P[\{\omega_3\}] = \frac{1}{8} \\
P[X = 3] &= P[\{\omega_4\}] = \frac{1}{8}.
\end{aligned}$$

The CDF of $X$ is:

$$F_X(x) = \begin{cases} 0, & x < 1 \\ 3/4, & 1 \leq x < 2 \\ 7/8, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

The RV $Y$ takes values in $\{1, 3\}$. The probability distribution of $X$ is as follows:

$$\begin{aligned}
P[Y = 1] &= P[\{\omega_3, \omega_4\}] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \\
P[Y = 3] &= P[\{\omega_1, \omega_2\}] = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.
\end{aligned}$$

The distribution function of $Y$ is:

$$F_Y(y) = \begin{cases} 0, & y < 1 \\ 1/4, & 1 \leq y < 3 \\ 1, & y \geq 3 \end{cases}$$

(b)

$$P[Y = 1|X = 1] = \frac{P[\{\omega_3, \omega_4\} \cap \{\omega_1, \omega_2\}]}{P[X = 1]} = 0$$

In a similar way, the other conditional probabilities can be found:

$$\begin{aligned}
P[Y = 1|X = 2] &= 1 \\
P[Y = 1|X = 3] &= 1 \\
P[Y = 3|X = 1] &= 1 \\
P[Y = 3|X = 2] &= 0 \\
P[Y = 3|X = 3] &= 0
\end{aligned}$$

(c) No, $X$ and $Y$ are *not* independent. Note, for example, that

$$P[Y = 1|X = 1] \neq P[Y = 1]. \tag{7}$$

Since

$$P[Y = 1|X = 1] = \frac{P[Y = 1, X = 1]}{P[X = 1]},$$

(7) implies that

$$P[Y = 1, X = 1] \neq P[Y = 1]P[X = 1],$$

which shows that $X$ and $Y$ are not independent.

**3.5 Expectation of nonnegative random variable.**

$$\begin{aligned} E[X] &= \sum_{k=0}^{\infty} kP[X = k] = \sum_{k=0}^{\infty} k\left(P[X > k-1] - P[X > k]\right) \\ &= \sum_{k=1}^{\infty}[(k-1)+1]P[X > k-1] - \sum_{k=0}^{\infty} kP[X > k] \\ &= \sum_{j=0}^{\infty} jP[X > j] + \sum_{j=0}^{\infty} P[X > j] - \sum_{k=0}^{\infty} kP[X > k] = \sum_{j=0}^{\infty} P[X > j]. \end{aligned}$$

**3.6 Expectation of a function of a random variable.** Define a new RV $Z$ in terms of RV $X$ by $Z = h(X)$. When $X$ takes on a value $x_i$, $Z$ takes on its corresponding value $z_i = h(x_i)$. Then clearly the probability $z_i$ occurs is the same as the probability $x_i$ occurs, namely, $p_Z(z_i) = p_X(x_i)$. Analogous to (3.32), we have

$$E[Z] = \sum_i z_i p_Z(z_i).$$

Then it readily follows that

$$E[h(X)] = \sum_i h(x_i)p_X(x_i),$$

which is (3.33).

**3.7 Statistical independence and expectations**

If (3.25) holds, then

$$\begin{aligned} E[h(X)g(Y)] &= \sum_i \sum_j h(x_i)g(y_j)P_{XY}(x_i, y_j) \\ &= \sum_i h(x_i)P_X(x_i) \sum_j g(y_j)P_Y(y_j) \\ &= E[h(X)]E[g(Y)]. \end{aligned}$$

Hence (3.25) is a sufficient condition for (3.34) to hold.

Conversely, suppose (3.34) holds for arbitrary functions $h(X)$ and $g(Y)$. Choose $h(X) = \delta_{X,x_i}$, i.e., $h(X) = 1$ for $X = x_i$ and $X = 0$ for $X =\neq x_i$. Then clearly

$$E[h(X)] = \sum_k h(x_k)p_X(x_k) = p_X(x_i).$$

Similarly, set $g(Y) = \delta_{Y,y_j}$. Then clearly

$$E[g(Y)] = \sum_k g(y_k)p_Y(y_k) = p_Y(y_j).$$

Similarly,

$$E[h(X)g(Y)] = p_{XY}(x_i, y_j).$$

Thus, (3.34) implies (3.25). In other words (3.25) is a necessary condition for (3.34) to hold. Combining the above two results we have shown (3.34) is a necessary and sufficient condition for (3.25).

**3.8* Properties of conditional expectations**

(a)

$$E[E[X|Y]] = E[\psi(Y)],$$

where

$$\psi(Y) = \sum_i p_{X|Y}(x_i|Y).$$

Then

$$\begin{aligned} E[E[X|Y]] &= \sum_j \psi(y_j)p_Y(y_j) \\ &= \sum_j \sum_i x_i p_{X|Y}(x_i|y_j)p_Y(y_j) \\ &= \sum_i x_i \sum_j p_{X|Y}(x_i|y_j)p_Y(y_j) = \sum_i x_i \sum_j p_{XY}(x_i, y_j) \\ &= \sum_i x_i p_X(x_i) = E[X]. \end{aligned}$$

(b) The LHS of the equation in (b) is

$$\begin{aligned} \text{LHS} &= \sum_i h(Y)g(x_i)p_{X|Y}(x_i|Y) \\ &= h(Y)\sum_i g(x_i)p_{X|Y}(x_i|Y) \\ &= h(Y)E[g(X)|Y], \end{aligned}$$

which is the RHS in (b).

(c) Consider a set of random variables $X_i$'s and scalars $a_i$'s. Then

$$E\left[\sum_i a_i X_i|Y\right] = \sum_i a_i E[X_i|Y],$$

which means that $E[\cdot|Y]$ is a linear operator.

**3.9 Conditional variance**

For notational convenience, we denote

$$\mu_X = E[X], \ \ \mu_X(Y) = E_X[X|Y],$$

where $E_X$ means the expectation with respect to the probability measure of the RV $X$. Clearly $\mu_X$ is a constant, whereas the conditional mean $\mu_X(Y)$ is a random variable, a function of $Y$. It is easy to show

$$E_Y[\mu_X(Y)] = \mu_X.$$

Write

$$X - \mu_X = X - \mu_X(Y) + \mu_X(Y) - \mu_X.$$

Take the square of the above expression and consider its expectation with respect to the probability measure of the join RVs $(X, Y)$:

$$\begin{aligned} E_X[(X-\mu_X)^2] &= E_{X,Y}[(X-\mu_X(Y))^2] + E_Y[(\mu_X(Y)-\mu_X)^2] \\ &\quad + 2E_{X,Y}[(X-\mu_X(Y))(\mu_X(Y)-\mu_X)]. \end{aligned}$$

It is clear that the LHS is

$$E_X[(X-\mu_X)^2] = Var[X],$$

The fist term of the RHS is

$$E_{X,Y}[(X-\mu_X(Y))^2] = E_Y E_X[(X-\mu_X(Y))^2|Y] = E_Y \mathrm{Var}[X|Y],$$

and the second term of the RHS is

$$E_Y[(\mu_X(Y)-\mu_X)^2] = \mathrm{Var}[\mu_X(Y)],$$

where we used the identity $\mu_X = E_Y[\mu_X(Y)]$. The third term of the RHS is zero because

$$\begin{aligned} &E_{X,Y}[(X-\mu_X(Y))(\mu_X(Y)-\mu_X)] \\ &= E_Y E_X[X\mu_X(Y)|Y] + \mu_X E_Y[\mu_X(Y)] - \mu_X E_X[X] - E_Y[\mu_X(Y)^2] \\ &= E_Y[\mu_X(Y)^2] + \mu_X^2 - \mu_X^2 - E_Y[\mu_X(Y)^2] = 0. \end{aligned}$$

Thus, we have shown

$$\mathrm{Var}[X] = E_Y \mathrm{Var}[X|Y] + E_Y \mathrm{Var}[X|Y],$$

from which the expression for the conditional variance ensues. Note that the regression analysis and the analysis of variance discussed in Section 22.1.4 (pp. 651-652) are closely related to this result.

**3.10*** **Correlation coefficient and Cauchy-Schwarz inequality** For given random variables $X$ and $Y$, define new random variables

$$X^* = X - E[X], \ \ Y^* = Y - E[Y].$$

Then the Cauchy-Schwarz inequality applied to the RVs $X^*$ and $Y^*$ gives

$$(E[X^*Y^*])^2 \le E[X^{*2}]E[Y^{*2}],$$

which is equivalent to

$$(\mathrm{Cov}(X, Y))^2 \leq \mathrm{Var}[X]\mathrm{Var}[Y],$$

where the equality holds iff

$$X^* = cY^*,$$

where $c$ is a scalar constant. The above inequality is equal to

$$(\mathrm{Cov}(X, Y))^2 \leq \mathrm{Var}[X]\mathrm{Var}[Y],$$

which is equivalent to

$$(\rho_{XY})^2 \leq 1.$$

The equality holds iff $Y - E[Y] = c(X - E[X])$ with probability 1, for some constant $c$. If $c > 0$, then $\rho_{XY} = 1$. This corresponds to perfect positive correlation.
If $c < 0$, then $\rho_{XY} = -1$. This corresponds to perfect negative correlation.

## 3.3 Important Probability Distributions

**3.11* Alternative derivation of the expectation and variance of binomial distribution.**
The mean and variance of the Bernoulli random variables $B_i$'s are

$$E[B_i] = p, \ \ E[B_i^2] = p, \ \ \text{hence,} \ \ \mathrm{Var}[X] = p - p^2 = p(1-p) = pq.$$

Since $B_i$'s are mutually independent, they are pairwise independent. Thus, we can apply Theorem 3.4, yielding

$$E[X] = p + p + \ldots + p = np, \ \ \mathrm{Var}[X] = pq + pq + \ldots + pq = npq.$$

**3.12* Trinomial and multinomial distributions.**

(a)

$$P[E_1] = p, P[E_2] = q, P[E_3] = 1 - p - q.$$

Since $E_1 \cup E_2 \cup E_3 = \Omega$, and $E_i$'s are independent $E_2 \cup E_3 = E_1^c$. Thus, out of the $n$ independent trials, the probability that event $E_1$ occurs $k_1$ times and $E_1^c$ occurs $n - k_1$ times is given by the following binomial distribution:

$$P[N(E_1) = k_1] = \binom{n}{k_1} p_1^{k_1}(1 - p_1)^{n-k_1}.$$

Then we distinguish whether a given outcome that shows $E_1^c$ is whether $E_2$ or $E_3$. The conditional probability of $E_2$ given that the event belongs to $E_1^c = E_2 \cup E_3$ is

$$P[E_2|E_1^c] = \frac{P[E_2 \cap E_1^c]}{P[E_1]} = \frac{P[E_2]}{P[E_1^c]} = \frac{q}{1-p}$$

and

$$P[E_3|E_1^c] = 1 - \frac{q}{1-p} = \frac{1-p-q}{1-p}.$$

Thus,

$$P[N(E_2) = k_2|N(E_1) = k_1] = P[N(E_2) = k_2|N(E_1^c) = n - k_1]$$
$$= \binom{n-k_1}{k_2}\left(\frac{q}{1-p}\right)^{k_2}\left(\frac{1-p-q}{1-p}\right)^{n-k_-k_2}.$$

Thus, the joint probability is obtained as

$$P[N(E_1) = k_1, N(E_2) = k_2] = P[N(E_1) = k_1]P[N(E_2) = k_2|N(E_1) = k_1]$$
$$= \binom{n}{k_1}p_1^{k_1}(1-p_1)^{n-k_1}\binom{n-k_1}{k_2}\left(\frac{q}{1-p}\right)^{k_2}\left(\frac{1-p-q}{1-p}\right)^{n-k_-k_2}$$
$$= \frac{n!}{k_1!k_2!(n-k_1-k_2)!}p^{k_1}q^{k_2}(1-p-q)^{n-k_1-k_2}.$$

(b) We can prove the formula by mathematical induction. Suppose that the multinomial formula is true for some $m = M_1 \geq 2$. Consider the following composite event

$$E_2 \cup E_3 \cup \cdots \cup E_M = E_1^c.$$

Then the distribution of observing $E_1$ $k_1$ times out of $n$ independent trials, and $E_1^c$ $(n-k_1)$ times is the binomial distribution:

$$P[N(E_1) = k_1] = \binom{n}{k_1}p_1^{k_1}(1-p_1)^{n-k_1}.$$

and the conditional probability of having $N(E_i) = k_i,\ i = 2, 3, \dots, M$ given $N(E_1) = k_1$ is from the assumption (i.e., the formula is true up to $m = M_1$)

$$P[N(E_1) = k_1, N(E_3) = k_3, \cdots, N(E_M) = k_M|N(E_1) = k_1]$$
$$= \frac{(n-k_1)!}{k_2!k_3!\cdots k_M!}\left(\frac{p_2}{1-p_1}\right)^{k_2}\left(\frac{p_3}{1-p_1}\right)^{k_3}\cdots\left(\frac{p_M}{1-p_1}\right)^{k_M}.$$

Then the joint probability is obtained by multiplying the above two expression, which leads to (3.117).

**3.13 Variance of geometric distribution.**

We write

$$E[X^2] = E[X(X-1) + X]$$

By differentiating the identity equation once more, we have

$$\sum_{k=1}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}.$$

Then by setting $x = q$ again, we find

$$E[X^2] = pq\frac{2}{(1-q)^3} + \frac{p}{(1-q)^2} = \frac{q}{p^2} + \frac{1}{p^2}.$$

Thus,

$$\text{Var}[X] = \frac{q}{p^2}.$$

**3.14 Shifted geometric distribution.**

$$E[Y] = E[X] - 1 = \frac{1}{p} - 1 = \frac{q}{p}.$$

The variance should not change by shifting of the distribution. Thus,

$$\text{Var}[Y] = \text{Var}[X] = \frac{q}{p^2}.$$

**3.15 Hypergeometric distribution**

(a) There are $\binom{N_1}{k}$ ways of choosing $k$ class-1 items, and $\binom{N_2}{n-k}$ ways of choosing $n-k$ class-2 ones. Since there are $\binom{N}{n}$ possible ways of choosing a sample of size $n$, and any choice of $k$ class-1 items may be combined with any choice of $n-k$ class-2 ones, we find (3.124).

(b) From the definition of binomial coefficients (2.37), $\binom{n}{k}$ is zero whenever $k > n$. Thus, the definition (3.124) is valid for all $k \geq 0$, provided the relation $p_k = 0$ is interpreted as impossibility. The hypergeometric distribution defined by (3.124) has the expression (3.125).

(c)

$$E[X] = \sum_{k=0}^{n} \frac{k\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}} = \sum_{k=1}^{n} \frac{k\binom{N_1}{k}\binom{N_2}{n-k}}{\binom{N}{n}}. \tag{8}$$

By writing

$$k\binom{N_1}{k} = N_1\binom{N_1-1}{k-1}, \quad \text{and} \quad \binom{N}{n} = \frac{N}{n}\binom{N-1}{n-1},$$

we obtain (**??**).

(d)

$$p_k^* \triangleq \frac{\binom{N_1-1}{k-1}\binom{N_2}{n-k}}{\binom{N-1}{n-1}}, \tag{9}$$

which is also a hyper-geometric distribution, as expected.

(e) Since the sum of a probability distribution $\{p_k^*\}$ amounts to unity, we find (3.127).

(f) The $j$th moment of $X$ is

$$E[X^j] = \sum_{k=0}^{n} k^j P[X=k] = \sum_{k=1}^{n} k^j \binom{N_1}{k}\binom{N_2}{n-k} \Big/ \binom{N}{n}.$$

Use the identities

$$k\binom{N_1}{k} = N_1\binom{N_1-1}{k-1} \quad \text{and} \quad n\binom{N}{n} = N\binom{N-1}{n-1},$$

we can rewrite

$$
\begin{aligned}
E[X^j] &= \frac{nN_1}{N}\sum_{k=1}^{n} k^{j-1}\binom{N_1-1}{k-1}\binom{N_2}{n-k}\Big/\binom{N-1}{n-1} \\
&= \frac{nN_1}{N}\sum_{i=0}^{n-1}(i+1)^{j-1}\binom{N_1-1}{i}\binom{N_2}{n-1-i}\Big/\binom{N-1}{n-1} \\
&= \frac{nN_1}{N}E\left[(X^*+1)^{j-1}\right].
\end{aligned}
$$

By setting $j=1$, we see $E[X]=\frac{nN_1}{N}$ as shown in the text.

(g) For $j=2$, we have

$$E[X^2]=\frac{nN_1}{N}\left[\frac{(n-1)(N_1-1)}{N-1}+1\right].$$

Thus

$$
\begin{aligned}
\mathrm{Var}[X] &= \frac{nN_1}{N}\left[\frac{(n-1)(N_1-1)}{N-1}+1-\frac{nN_1}{N}\right] \\
&= \frac{N-n}{N-1}npq.
\end{aligned}
$$

**3.16 Relation between hyper-geometric and binomial distributions.** The hypergeometric distribution differs from the binomial distribution in that the population is a finite number $N$ and that sampling is done without replacement.

(a)

$$p_k \to \binom{n}{k}p^k(1-p)^{n-k},\ \ k=0,1,2,\ldots,n. \tag{10}$$

The above result suggests that when the population $N_1$ and $N_2$ are both much greater than the sample size $n$, the distribution $\{p_k\}$ hardly depends on whether the sampled items are replaced back into the box or not after each drawing.

Note that if sampling is done with replacement, then the resulting distribution should be equivalent to the binomial distribution with $p=\frac{N_b}{N}$, even when $N$ is finite.

(b)

$$p_{k_1,k_2}=\frac{\binom{N_1}{k_1}\binom{N_2}{k_2}\binom{N_3}{n-k_1-k_2}}{\binom{N}{n}}. \tag{11}$$

The above probabilities are defined for $(k_1,k_2)$ such that

$$0\le k_1\le N_1,\ 0\le k_2\le N_2,\ \text{and}\ 0\le n-k_1-k_2\le N_3.$$

**3.17 Bridge game.**

$$\frac{\binom{13}{5}\binom{13}{4}\binom{13}{2}\binom{13}{2}}{\binom{52}{13}}.$$

**3.18* Mean, second moment and variance of the Poisson distribution.**

(a)

$$E[X] = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda\sum_{k=1}^{\infty}\frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!}$$
$$= \lambda e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

(b)

$$E[X^2] = \sum_{k=0}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} = \lambda\sum_{k=1}^{\infty} k\frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} = \lambda\sum_{i=0}^{\infty}\frac{(i+1)e^{-\lambda}\lambda^i}{i!}$$
$$= \lambda\left[\sum_{i=0}^{\infty}\frac{ie^{-\lambda}\lambda^i}{i!} + \sum_{i=0}^{\infty}\frac{e^{-\lambda}\lambda^i}{i!}\right] = \lambda(\lambda+1).$$

(c)

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2 = \lambda.$$

**3.19 Tail of the Poisson distribution.**

(a) Since $P[S=j] = \frac{\lambda^j}{j!}e^{-\lambda}$, the left hand side inequality is obvious. The RHS inequality can be shown by

$$P[S \geq k] = \sum_{j=k}^{\infty}\frac{\lambda^j}{j!}e^{-\lambda} = \frac{\lambda^k}{k!}e^{-\lambda}\left[1 + \frac{\lambda}{k+1} + \frac{\lambda^2}{(k+2)(k+1)} + \cdots\right]$$
$$< \frac{\lambda^k}{k!}e^{-\lambda}\left[1 + \frac{\lambda}{k+1} + \left(\frac{\lambda}{k+1}\right)^2 + \cdots\right] = \frac{\lambda^k}{k!}e^{-\lambda}\frac{1}{1-\frac{\lambda}{k+1}}.$$

(b) By substituting the Stirling's approximation $k! \approx \sqrt{2\pi k}k^k e^{-k}$,

$$\text{LHS} = \frac{\lambda^k e^{-\lambda}}{\sqrt{2\pi k}k^k e^{-k}} = \frac{1}{\sqrt{2\pi k}}\left(\frac{\lambda}{k}\frac{e^{-\frac{\lambda}{k}}}{e^{-1}}\right)^k$$
$$= \frac{1}{\sqrt{2\pi k}}\left(\frac{\lambda}{k}e^{1-\frac{\lambda}{k}}\right)^k$$
$$= \frac{\Delta^k}{\sqrt{2\pi k}}.$$

The RHS can be shown exactly in the same way.

**3.20* Identities between $p(k;\lambda)$ and $Q(k;\lambda)$.**

(a) By substituting the definitions of $p(k;\lambda)$ and $Q(k;\lambda)$, the left hand side (LHS) becomes

$$\text{LHS} = \sum_{k'=0}^{k}\frac{\lambda_1^{k-k'}}{(k-k')!}e^{-\lambda_1}\sum_{i=0}^{k'}\frac{\lambda_2^i}{i!}e^{-\lambda_2}.$$

By setting $k' = i + k - j$ and changing the order of summation, we have

$$\begin{aligned}
\text{LHS} &= \sum_{j=0}^{k} \sum_{i=0}^{j} \frac{\lambda_1^{j-i}}{(j-2)!} \frac{\lambda_2^i}{i!} e^{-(\lambda_1+\lambda_2)} \\
&= \sum_{j=0}^{k} \frac{(\lambda_1+\lambda_2)^j}{j!} e^{-(\lambda_1+\lambda_2)} = Q(k; \lambda_1 + \lambda_2).
\end{aligned}$$

(b) Using the hint, we have

$$\begin{aligned}
\int_\lambda^\infty p(k; y)\, dy &= \left[-e^{-y} \frac{y^k}{k!}\right]_\lambda^\infty - \int_\lambda^\infty \left(-e^{-y}\right) \frac{y^{k-1}}{(k-1)!}\, dy \\
&= p(k; \lambda) + \int_\lambda^\infty p(k-1; y)\, dy.
\end{aligned}$$

From this recursive relation, we have

$$\begin{aligned}
\text{LHS} &= p(k; \lambda) + p(k-1; \lambda) + \ldots + p(1; \lambda) + \int_\lambda^\infty e^{-y}\, dy \\
&= \int_{i=0}^{k} p(i; \lambda) = Q(k; \lambda).
\end{aligned}$$

(c)

$$(k + \lambda + 1) Q(k; \lambda) + (k+1) Q(k; \lambda).$$

By substituting the recursive relations $Q(k; \lambda) = Q(k-1; \lambda) + p(k; \lambda)$ and $Q(k; \lambda) = Q(k+1; \lambda) - p(k+1; \lambda)$, we can rewrite the LHS of the first expression as

$$\text{LHS} = \lambda Q(k-1; \lambda) + (k+1) Q(k+1; \lambda) + \lambda p(k; \lambda) - (k+1) Q(k+1; \lambda).$$

The last two terms cancel each other, since

$$\lambda p(k; \lambda) = (k+1) p(k+1; \lambda) = \frac{\lambda^{k+1}}{k!} e^{-\lambda}.$$

Thus, the formula (c) follows.

(d) From the result (c), we have

$$Q(k-1; \lambda) = \frac{Q(k; \lambda) + k Q(k; \lambda)}{k + \lambda},$$

from which we can derive

$$\begin{aligned}
k Q(k; \lambda) - \lambda Q(k-1; \lambda) &= k Q(k-1; \lambda) - \lambda Q(k-2; \lambda) \\
&= Q(k-1; \lambda) + (k-1) Q(k-1; \lambda) - \lambda Q(k-2; \lambda)
\end{aligned}$$

By applying this recursive relation, we have

$$\begin{aligned}
\text{LHS} &= Q(k-1; \lambda) + Q(k-2; \lambda) + \ldots + Q(1; \lambda) + \ldots + Q(1; \lambda) + Q(1; \lambda) - \lambda Q(0; \lambda) \\
&= \sum_{j=0}^{k-1} Q(j; \lambda),
\end{aligned}$$

where we used the relation

$$Q(1;\lambda) - \lambda Q(0;\lambda) = \lambda e^{-\lambda} + e^{-\lambda} - \lambda e^{-\lambda} = e^{-\lambda} = Q(0;\lambda).$$

(e) From the result of (d), the right hand side is

$$\text{RHS} = \sum_{j=0}^{k-1} Q(j;\lambda).$$

Then using the result of (b), we have

$$\text{RHS} = \sum_{j=0}^{k-1} \int_{\lambda}^{\infty} p(j;\lambda)\,dy.$$

By interchanging the order of summation and integration, we have

$$\text{RHS} = \int_{\lambda}^{\infty} \sum_{j=0}^{k-1} p(j;y)\,dy = \int_{\lambda}^{\infty} Q(k-1;y)\,dy.$$

**3.21*** **Derivation of the identity (3.96).** Let $f(x) = (1-x)^{-r} = (-1)^r(x-1)^{-r}$ and expand this around $x = 0$ using the Taylor series expansion:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n,$$

where

$$f'(x) = r(1-x)^{-r-1},\ \ f''(x) = r(r+1)(1-x)^{-r-2}, \cdots,$$
$$f^{(n)} = r(r+1)\cdots(r+n-1)(1-x)^{-(r+n)}.$$

Hence,

$$\frac{f^{(n)}(0)}{n!} = \frac{r(r+1)\cdots(r+n-1)}{n!} = \binom{r+n-1}{r-1}.$$

Set $x = q$, then

$$(1-q)^{-r} = \sum_{n=0}^{\infty} \binom{r+n-1}{r-1} q^n.$$

Setting $n = k - r$, we have

$$(1-q)^{-r} = \sum_{k=r}^{\infty} \binom{k-1}{r-1} q^{k-r}.$$

**3.22*** **Equivalence of two expressions for the negative binomial distribution.** We want to show that

$$\binom{k-1}{k-r} p^r q^{k-r} = \sum_{i=r}^{k} \binom{k}{i} p^i q^{k-i} - \sum_{i=r}^{k-1} \binom{k-1}{i} p^i q^{k-1-i},\ \ k \geq r,$$

where $q = 1 - p$. By moving the second term of the right-hand side, we want to show

$$\sum_{i=r}^{k} \binom{k}{i} p^i q^{k-i} \stackrel{?}{=} \binom{k-1}{k-r} p^r q^{k-r} + \sum_{i=r}^{k-1} \binom{k-1}{i} p^i q^{k-1-i}. \tag{12}$$

The left-hand side (LHS) and right-hand side (RHS) of the above can be written as

$$\text{LHS} = (p+q)^k - \sum_{i=0}^{r-1} \binom{k}{i} p^i q^{k-i} = 1 - \sum_{i=0}^{r-1} \binom{k}{i} p^i q^{k-i}. \tag{13}$$

and

$$\begin{aligned}
\text{RHS} &= \binom{k-1}{k-r} p^r q^{k-r} + (p+q)^{k-1} - \sum_{i=0}^{r-1} \binom{k-1}{i} p^i q^{k-1-i} \\
&= \binom{k-1}{r-1} + 1 - \binom{k-1}{r-1} p^{r-1} q^{k-r} - \sum_{i=0}^{r-2} \binom{k-1}{i} p^i q^{k-1-i} \\
&= 1 + \binom{k-1}{r-1} p^{r-1}(p-1) q^{k-r} - \sum_{i=0}^{r-2} \binom{k-1}{i} p^i q^{k-1-i} \\
&= 1 - \binom{k-1}{r-1} p^{r-1} q^{k+1-r} - \sum_{i=0}^{r-2} \binom{k-1}{i} p^i q^{k-1-i}.
\end{aligned} \tag{14}$$

Thus we need to examine

$$\binom{k-1}{r-1} p^{r-1} q^{k+1-r} \stackrel{?}{=} \sum_{i=0}^{r-1} \binom{k}{i} p^i q^{k-i} - \sum_{i=0}^{r-2} \binom{k-1}{i} p^i q^{k-1-i}. \tag{15}$$

Using the well-known formula

$$\binom{k}{i} = \binom{k-1}{i-1} + \binom{k-1}{i},$$

we rearrange the RHS of (15) as follows:

$$\begin{aligned}
\text{RHS} &= \binom{k}{r-1} p^{r-1} q^{k-r+1} + \sum_{i=0}^{r-2} \left[ \binom{k-1}{i-1} q + \binom{k-1}{i} q - \binom{k-1}{i} \right] p^i q^{k-i-1} \\
&= \binom{k}{r-1} p^{r-1} q^{k+1-r} + \sum_{i=0}^{r-2} \binom{k-1}{i-1} p^i q^{k-i} - \sum_{i=0}^{r-2} \binom{k-1}{i} p^{i+1} q^{k-i-1} \\
&= \binom{k}{r-1} p^{r-1} q^{k+1-r} + \sum_{j=0}^{r-3} \binom{k-1}{j} p^{j+1} q^{k-j-1} - \sum_{i=0}^{r-2} \binom{k-1}{i} p^{i+1} q^{k-i-1} \\
&= \binom{k}{r-1} p^{r-1} q^{k+1-r} - \binom{k-1}{r-2} p^{r-1} q^{k-r+1} \\
&= \binom{k-1}{r-1} p^{r-1} q^{k+1-r},
\end{aligned}$$

which is equal to the LHS of (15).

**3.23 Another negative binomial distribution.**
Let $Y_r$ be the number of trials required to achieve $r$ successes. Then $Z_r = Y_r - r$ is the number of failures that precede the $r$th success. Then:

$$P[Z_r = k] = P[Y_r = r + k] = \binom{r+k-1}{k} p^r (1-p)^k, \quad k = 0, 1, 2, \cdots .$$

**3.24 Negative binomial and Poisson distribution.** We have:

$$P[X_r = j] = \frac{\Gamma(r+j)}{\Gamma(r)\Gamma(j+1)} p^r (1-p)^j, \quad j = 0, 1, 2, \cdots$$

In considering the limit as $r \to \infty$ and $(1-p) \to 0$ such that $r(1-p) \to \lambda$, let us assume that $r$ runs through the positive integers. Then in this limiting regime:

$$\begin{aligned}
&\lim_{r\to\infty} P[X_r = j] \\
&\quad = \lim_{r\to\infty} \frac{(r+j-1)(r+j-2)\cdots(r)}{j!} p^r (1-p)^j \\
&\quad = \lim_{r\to\infty} \frac{[r(1-p)+(j-1)(1-p)][r(1-p)+(j-2)(1-p)]\cdots r(1-p)]}{j!} p^r \\
&\quad = \lim_{r\to\infty} \frac{[r(1-p)]^j}{j!} p^r = \frac{\lambda^j}{j!} \lim_{r\to\infty} p^r = \frac{\lambda^j}{j!} \lim_{r\to\infty} e^{r \ln p}.
\end{aligned} \tag{16}$$

From the hint that was given in the problem statement, we have that:

$$\lim_{r\to\infty} r \ln p = -\lambda. \tag{17}$$

Applying (17) in (16), we obtain the desired result:

$$\lim_{r\to\infty} P[X_r = j] = \frac{\lambda^j}{j!} e^{-\lambda}.$$

To prove that (17) holds, note that the Taylor series expansion of $\ln p$ about $1$ is given by:

$$\ln p = (p-1) - \frac{1}{2}(p-1)^2 + \frac{1}{3}(p-1)^3 - \cdots$$

Therefore,

$$r \ln p = r(p-1)\left[1 - \frac{1}{2}(p-1) + o(p-1)\right],$$

where the "little O" notation $o(h)$ means that

$$\lim_{h\to 0} o(h)/h = 0.$$

Hence, in the limit as $r \to \infty$ and $p \to 1$ with $r(1-p) = \lambda$, we have

$$\lim_{r\to\infty} r \ln p = -\lambda.$$

**3.25 Zipf's law with $\alpha = 1$.**

(a) Let $z_N = \sum_{n=1}^{N} \frac{1}{n}$. By applying the "rectangle rule" for obtaining an overestimate of the integral of the function $f(x) = \frac{1}{x}$ from 1 to $N$, we see that

$$z_{N-1} > \int_1^N \frac{1}{x}\, dx = \ln N.$$

Since $\ln N \to \infty$ as $N \to \infty$, we have

$$\lim_{N\to\infty} z_{N-1} = \infty = \zeta(1)..$$

(b) Using the same argument as in part (a), we find that

$$z_N > \ln(N+1).$$

If we use the rectangle rule to underestimate the integral of the function $f(x) = \frac{1}{x}$ from 1 to $N$, we find that

$$\sum_{n=1}^{N-1} \frac{1}{n+1} < \int_1^N \frac{1}{x}\, dx = \ln N.$$

Since the right-hand side is equal to $z_N - 1$, we obtain

$$z_{N+1} - 1 < \ln N \text{ or } z_{N+1} < \ln N + 1.$$

Since $z_N < z_{N+1}$, we have

$$z_N < \ln N + 1.$$

**3.26 Generalized Zipf's law.**

(a) Let

$$z_N(\alpha) \triangleq \sum_{n=1}^{N} \frac{1}{n^\alpha}, \;\; 0 < \alpha < 1.$$

Also note that

$$\int_1^y \frac{1}{x^\alpha}\, dx = \frac{y^{1-\alpha} - 1}{1-\alpha}. \tag{18}$$

Using the rectangle rule to overestimate the LHS of the above equation, we find that

$$z_{N-1}(\alpha) > \frac{N^{1-\alpha} - 1}{1-\alpha}.$$

Since the RHS of the above inequality goes to infinity as $N \to$, so to does $z_N$. Hence,

$$\zeta(a) = \lim_{N\to\infty} z_N = \infty.$$

(b) Using the argument of part (a), we readily find that

$$z_N(\alpha) > \frac{(N+1)^{1-\alpha} - 1}{1-\alpha}. \tag{19}$$

We can also apply the rectangle rule to underestimate the integral (18) as follows:

$$\sum_{n=1}^{N-1} \frac{1}{(n+1)^\alpha} < \int_1^N \frac{1}{x^\alpha}\, dx = \frac{N^{1-\alpha}-1}{1-\alpha}.$$

Hence,

$$z_N(\alpha) - 1 < \frac{N^{1-\alpha}-1}{1-\alpha}. \tag{20}$$

Combining (19) and (20), we have

$$\frac{(N+1)^{1-\alpha}-1}{1-\alpha} < z_N(\alpha) < \frac{N^{1-\alpha}-1}{1-\alpha} + 1.$$

# 4 Solutions for Chapter 4: Continuous Random Variables

## 4.1 Continuous Random Variables

**4.1*** **Expectation of a nonnegative continuous RV.**

$$\begin{aligned}\int_0^\infty x f_X(x)\,dx &= -\int_0^\infty x(1-F_X(x))'\,dx \\ &= -[x(1-F_X(x)]_0^\infty + \int_0^\infty (1-F_X(x))\,dx \\ &= \int_0^\infty (1-F_X(x))\,dx.\end{aligned}$$

Thus, the formula for non-negative random variables is proved. If we drop the assumption of nonnegativity, we proceed as follows:

$$\begin{aligned}\int_{-\infty}^0 x f_X(x)\,dx &= \int_{-\infty}^0 x F_X^{'}(x)\,dx \\ &= [xF_X(x)]_{-\infty}^0 - \int_{-\infty}^0 F_X(x)\,dx \\ &= -\int_{-\infty}^0 F_X(x)\,dx.\end{aligned}$$

Combining the above two, we have shown (4.10).

**4.2*** **Properties of discrete RVs.** Let the discrete random variables have probability masses $p_i > 0$ at $x = x_i; \quad -\infty < i < \infty$ such that

$$\cdots < x_{-2} < x_1 < \cdots < x_0 (= 0) < x_1 < x_2 < \cdots,$$

If $x = 0$ is not a mass point, assign $p_0 = 0$.

We can write

$$F_X(x) = \sum_{i=-\infty}^{\infty} p_i u(x - x_i),$$

Let

$$F_j = F(x_j) = \sum_{i=-n}^{m} p_i u(x_j - x_i) = \sum_{i=-\infty}^{j} p_i.$$

and

$$F_j^c = 1 - F(x_j) = 1 - \sum_{i=-\infty}^{j} p_i.$$

Then

$$\begin{aligned}\int_0^\infty (1 - F_X(x))\,dx &= \sum_{i=1}^{m} \int_{x_{i-1}}^{x_i} F_X^c(x)\,dx \\ &= \sum_{i=1}^{\infty} (x_i - x_{i-1}) F_{i-1}^c = \sum_{i=1}^{\infty} F_{i-1}^c x_i - \sum i = 0^\infty F_i^c x_i \\ &= -F_0^c + \sum_{i=1}^{\infty} \left(F_{i-1}^c - F_i^c\right) x_i \\ &= 0 + \sum i = 1^\infty p_i x_i,\end{aligned}$$

where we used the property

$$F_{i-1}^c - F_i^c = p_i.$$

Similarly,

$$\begin{aligned}\int_{-\infty}^{0} F_X(x)\,dx = &\sum_{i=-\infty}^{0} (x_i - x_{i-1}) F_{i-1} \\ &\sum_{i=-\infty}^{-1} x_i (F_{i-1} - F_i) + x_0 F{-1} = \sum_{i=-\infty}^{-1} x_i(-p_i).\end{aligned}$$

Hence

$$\begin{aligned}E[X] &= \int_0^\infty (1 - F_X(x))\,dx - \int_{-\infty}^{0} F_X(x)\,dx = + \sum_{i=-\infty}^{-1} p_i x_i + \sum i = 1^\infty p_i x_i \\ &= \sum_{i=-\infty}^{\infty} p_i x_i,\end{aligned}$$

as expected.
For a nonnegative discrete RV,

$$E[X] = \int_0^\infty (1 - F_X(x))\,dx = \sum i = 1^\infty p_i x_i = \sum i = 0^\infty p_i x_i,$$

as expected.

**4.3 Mixture models and mixed RVs.**

(a) Since $Z = X_i$ with probability $P[Z = X_i] = p_i$, we have

$$P[Z \le z] = \sum_{i=1}^{n} P[Z = X_i] P[X_i \le z | Z = X_i] = \sum_{i=1}^{n} p_i P[X_i \le z | Z = X_i].$$

Then

$$F_Z(z) = \sum_{i=1}^{n} p_i F_{X_i}(z).$$

(b) The expectation is a linear operator, thus

$$E[Z] = \sum_{i=1}^{n} p_i E[X_i].$$

(c) (i) If all $X_i$ are continuous RVs, then $F_{X_i}(z)$'s are continuous everywhere. Then from (a) we see $F_Z(z)$ is everywhere continuous, hence $Z$ is a continuous RV.

(ii) Suppose $X_i$ takes only discrete values $V_i = \{x_i(1), x_i(2), \ldots, x_i(i_j), \ldots\}$, where $V_i$ is a finite or infinite set. Then, $Z$ takes on the values in the set $\cup_{i=1}^{n} V_i$, which is also discrete. Hence, $Z$ is a discrete RV. An alternative argument to prove this is as follows. The distribution function $F_{X_i}(z)$ of a discrete RV $X_i$ is a staircase function as shown in Fig. 3.3 (b). Then a weighted sum of such staircase functions $F_Z(z)$ given in (a) is also a staircase function. Hence, $Z$ is a discrete RV.

(iii) If some $F_{X_i}(z)$'s are everywhere continuous, while the others are staircase functions, then the weighted sum $F_Z(z)$ is neither everywhere continuous nor a staircase function. Thus $Z$ is neither a discrete RV nor a continuous RV.

## 4.4* Expectation and the Riemann-Stieltjes integral

(a) We can write the PDF as

$$f_X(x) = \sum_i p_X(x_i)\delta(x - x_i).$$

Then (4.9) implies

$$\begin{aligned}\mu_X &= \int_{-\infty}^{\infty} x \sum_i p_X(x_i)\delta(x - x_i)\, dx \\ &= \sum_i p_X(x_i) \int_{-\infty}^{\infty} x\delta(x - x_i)\, dx = \sum_i p_X(x_i) x_i.\end{aligned}$$

(b) (i) If $X$ is a continuous RV,

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du, \text{ and } dF_X(x) = f_X(x)dx.$$

Then (4.159) is reduced to (4.9) (not to (3.32)).

(ii) If $X$ is a discrete RV,

$$F_X(x) = \sum_i p_X(x_i)u(x - x_i), \text{ and } , dF_X(x) = \sum_i p_X(x_i)\delta(x - x_i)\, dx.$$

Then (4.159) reduces to (3.32) (not to (4.9)).

(iii)

$$\begin{aligned}\mu_X &= \int_0^\infty x dF_X(x) + \int_{-\infty}^0 x dF_X(x)\\ &= -\int_0^\infty x d(1-F_X(x)) + \int_{-\infty}^0 x dF_X(x)\\ &= -[x(1-F_X(x))]_0^\infty + \int_0^\infty (1-F_X(x))\,dx + [xF_X(x)]_{-\infty}^0 - \int_{-\infty}^0 F_X(x)\,dx\\ &= \int_0^\infty (1-F_X(x))\,dx - \int_{-\infty}^0 F_X(x)\,dx,\end{aligned}$$

which is (4.10). For a nonnegative RV, the second term disappears and we obtain (4.11).

**4.5 Expectation of functions of RVs.**

(a) Let

$$Y = h(X),$$

and the PDF of $Y$ be denoted as $f_Y(y)$. Then $f_Y(y)\,dy = f_X(x)\,dx$. Then $E[Y] = E[h(X)]$ and

$$E[Y] = \int_{-\infty}^\infty y f_Y(y)\,dy = \int_{-\infty}^\infty h(x) f_X(x)\,dx.$$

(b) If $X$ and $Y$ are independent, $f_{XY}(x,y) = f_X(x)f_Y(y)$. Then

$$\begin{aligned}E[h(X)g(Y)] &= \int\int h(x)f(y)f_{XY}(x,y)\,dx\,dy\\ &= \int h(x)f_X(x)\,dx \int g(y)f_Y(y)\,dy = E[h(X)]E[h(Y)].\end{aligned}$$

Conversely, if $E[h(X)g(Y)] = E[h(X)]E[g(Y)]$ for arbitrary functions $h(\cdot)$ and $g(\cdot)$, let

$$h(x) = \delta(x-x_0), \text{ and } g(y) = \delta(y-y_0).$$

Then,

$$E[h(X)g(Y)] = \int\int \delta(x-x_0)\delta(y-y_0)f_{XY}(x,y)\,d\,dy = f_{XY}(x_0,y_0),$$

and

$$E[h(X)] = \int \delta(x-x_0)f_X(x)\,dx = f_X(x_0) \text{ and} E[g(Y)] = \int \delta(y-y_0)f_Y(x)\,dy = f_Y(y_0).$$

Thus

$$f_{XY}(x_0,y_0) = f_X(x_0)f_Y(y_0),$$

for all $-\infty < x_0 < \infty, -\infty < y_0 < \infty$. Thus, $X$ and $Y$ are independent.

**4.6 Second moment of a continuous RV.**

$$
\begin{aligned}
E[X^2] &= \int_0^\infty x^2 f_X(x)\,dx + \int_{-\infty}^0 x^2 f_X(x)\,dx \\
&= -\int_0^\infty x^2[1-F_X(x)]'\,dx + \int_{-\infty}^0 x^2 f_X(x)\,dx \\
&= -[x^2(1-F_X(x))]_0^\infty + \int_0^\infty 2x[1-F_X(x)]\,dx \\
&+ [x^2F_X(x)]_{-\infty}^0 - \int_{-\infty}^0 2xF_X(x)\,dx.
\end{aligned}
$$

The first and third terms are zero by using the hint of Problem 4.1.

**4.7 Joint PDF of two continuous RVs.**

(a)

$$
\begin{aligned}
F_{XY}(x,y) &= \int_0^y \int_0^x f_{XY}(u,v)du\,dv = k\int_0^y \int_0^x e^{-\lambda u-\mu v}du\,dv \\
&= k\left(\int_0^y e^{-\mu v}dv\right)\left(\int_0^x e^{-\lambda u}du\right) = \frac{k}{\mu\lambda}(1-e^{-\mu y})(1-e^{-\lambda x})
\end{aligned}
$$

Since $\lim_{x,y\to\infty} F_{XY}(x,y) = 1$, we find that $k = \mu\lambda$. Therefore,

$$F_{XY}(x,y) = (1-e^{-\mu y})(1-e^{-\lambda x}).$$

(b)

$$
\begin{aligned}
F_X(x) &= \lim_{y\to\infty} F_{XY}(x,y) = 1-e^{-\lambda x} \\
F_Y(y) &= \lim_{x\to\infty} F_{XY}(x,y) = 1-e^{-\lambda y} \\
F_{Y|X}(y|x) &= \frac{\frac{\partial}{\partial x}F_{XY}(x,y)}{f_X(x)} = \frac{\lambda e^{-\lambda x}(1-e^{-\mu y})}{\lambda e^{-\lambda x}} = 1-e^{-\mu y} = F_Y(y).
\end{aligned}
$$

**4.8 Conditional expectations of two RVs.** This is the continuous analog of what we proved for discrete RVs in Problem 3.8

(a)

$$E[X|Y] = \int_x x f_{X|Y}(x)\,dx,$$

and

$$
\begin{aligned}
E[E[X|Y]] &= \int_y f_Y(y)\,dy \int_x x f_{X|Y}(x)\,dx = \int_x x\left(\int_y f_{XY}(x,y)\,dy\right)\,dx \\
&= \int_x x f_X(x)\,dx = E[X].
\end{aligned}
$$

(b)

$$\begin{aligned} E[h(Y)g(X)|Y] &= \int h(Y)g(X)f_{X|Y}(x)\,dx = h(Y)\int g(x)f_{X|Y}(x)\,dx \\ &= h(Y)E[g(X)|Y] \end{aligned}$$

(c) Let $X_i$'s be random variables. Then

$$\begin{aligned} E\left[\sum_i a_i X_i|Y\right] &= \int\int\cdots\int\left(\sum_{i=1}^n a_i x_i\right) f_{X_1X_2\cdots X_n|Y}(x_1,x_2,\cdots,x_n)\,dx_1\,dx_2\cdots dx_n \\ &= \sum_{i=1}^n a_i\int x_i f_{X_i|Y}(x_i)\,dx_i = \sum_{i=1}^n a_i E[X_i], \end{aligned}$$

for all $n$ and scalars $a_1, a_2, \ldots, a_n$. Thus $E[\cdot|Y]$ is a linear operator.

## 4.2 Important Continuous Random Variables and Their Distributions

**4.9* Expectation, second moment and variance of the uniform RV.**

$$\begin{aligned} \mu_X = E[X] &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx \\ &= \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2}, \end{aligned}$$

which is the midpoint of the interval $[a,b]$.
The 2nd moment can be found in a similar fashion:

$$\begin{aligned} E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \frac{1}{b-a}\int_a^b x^2\,dx \\ &= \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ba+a^2}{3}. \end{aligned}$$

Thus,

$$\sigma_X^2 = E[X^2] - \mu_X^2 = \frac{b^2+ba+a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

**4.10* Moments of uniform RV.**

(a)

$$\frac{1}{b-a}\int_a^b x^n\,dx = \frac{b^{n+1}-a^{n+1}}{(n+1)(b-a)}.$$

(b)

$$\frac{1}{b-a}\int_a^b \left(x - \frac{b+a}{2}\right)^n\,dx = \frac{1+(-1)^n}{(n+1)2^n}\frac{(b-a)^{n+1}}{b-a}.$$

**4.11 Erlang distribution**

(a) From the definition (4.162), we readily find

$$E[Y_r] = r/\lambda,$$

and

$$\mathrm{Var}[Y_r] = r/\lambda^2.$$

(b) By differentiating $F_{Y_r}(y)$ of (4.163),

$$\begin{aligned} f_{Y_r}(y) = \frac{dF_{Y_r}(y)}{dy} &= \lambda e^{-\lambda y} \sum_{j=0}^{r-1} \frac{(\lambda y)^j}{j!} - e^{-\lambda y} \sum_{j=1}^{r-1} \frac{j\lambda(\lambda y)^{j-1}}{j-1!} \\ &= \lambda e^{-\lambda y} \sum_{j=0}^{r-1} \frac{(\lambda y)^j}{j!} - \lambda e^{-\lambda y} \sum_{i=0}^{r-2} \frac{(\lambda y)^i}{i!} \\ &= \frac{\lambda(\lambda y)^{r-1}}{(r-1)!} e^{-\lambda y}, \;\; y \geq 0. \end{aligned} \tag{1}$$

(c) Since $F_{S_r}(t) = P[S_r \leq t] = P[Y_r \leq rt] = F_{Y_r}(rt)$, we have

$$F_{S_r}(t) = 1 - e^{-\lambda rt} \sum_{j=0}^{r-1} \frac{(\lambda rt)^j}{j!} \tag{2}$$

$$= 1 - e^{-r\lambda t} \sum_{j=0}^{r-1} \frac{(r\lambda t)^j}{j!}. \tag{3}$$

By differentiating the above, we find

$$f_{S_r}(t) = \frac{r\lambda(r\lambda t)^{r-1}}{(r-1)!} e^{-r\lambda t}, \;\; t \geq 0, \tag{4}$$

(d) In Figure 4.1 we show the family of the $r$-stage Erlangian PDFs $S_{\lambda,r}$'s for $r = 1, 2, 4, 8, 16$. The case with $r = 1$ is an exponential distribution, and as $r$ increases, the distribution becomes more like a Gaussian distribution, which will be discussed in Section 4.2.4. In the limit $r \to \infty$, the PDF of $E_r$ approaches a unit impulse function at the point $t = 1/\lambda$.

**4.12 Hyper-exponential (or mixed exponential) distribution.**

(a)

$$f_S(t) = \sum_{i=1}^{k} \pi_i \mu_i e^{-\mu_i t}. \tag{5}$$

and

$$E[S] = \sum_{i=1}^{k} \frac{\pi_i}{\mu_i} = \frac{1}{\mu} \tag{6}$$

(b) In Figure 4.2 we plot the curves. The mean of the distribution is $1/\mu = \pi_1/\mu_1 + \pi_2/\mu_2 \approx 1.0$.

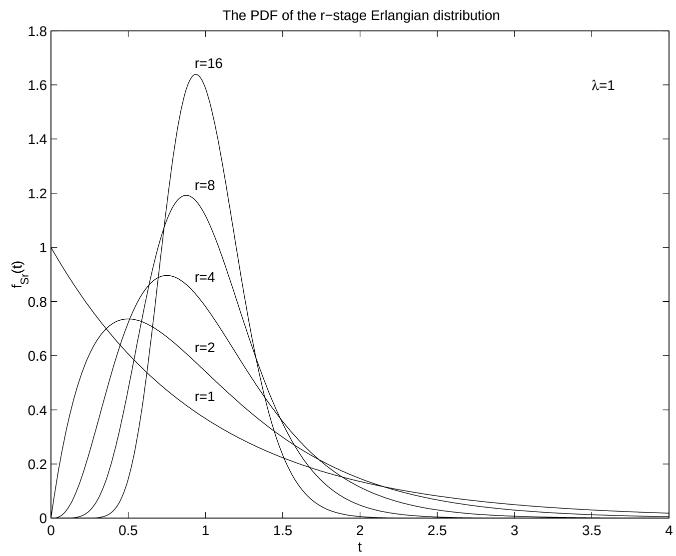



**Figure 4.1** The PDFs of the $r$-stage Erlangian distribution, $E_r$, for $r = 1, 2, 4, 8, 16$.

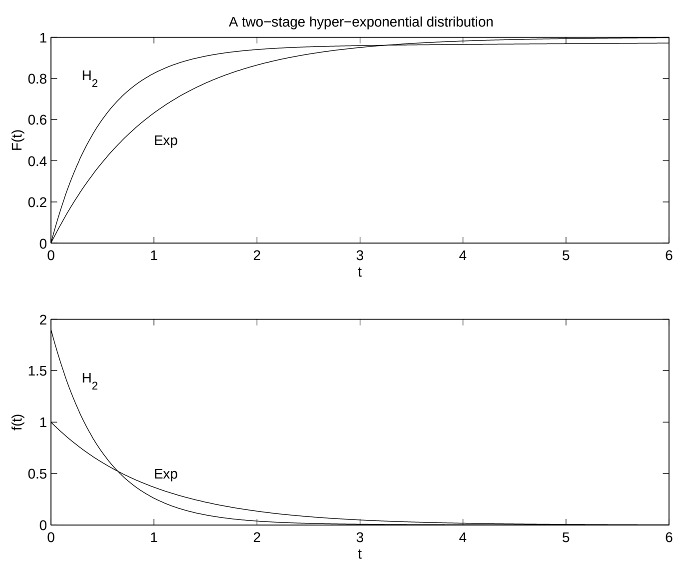



**Figure 4.2** An example of two-stage hyper-exponential distribution, $H_2$.

In the same figure we also plot the exponential distribution with the same mean, i.e., $1/\mu = 1$.

**4.13* Recursive formula for the gamma function.** The gamma function is defined by

$$\Gamma(\beta) = \int_0^\infty x^{\beta-1} e^{-x} dx.$$

Using integration by parts, we get

$$\Gamma(\beta) = x^{\beta-1}(e^{-x})\mid_0^\infty + \int_0^\infty (\beta-1)x^{\beta-2}e^{-x}dx = (\beta-1)\Gamma(\beta-1).$$

**4.14 Poisson distribution and the gamma distribution.**
From (4.30), the gamma distribution is given by

$$f_{\lambda,\beta}(y) = \frac{\lambda(\lambda y)^{\beta-1}}{\Gamma(\beta)}e^{-\lambda y}, \ \ y \geq 0.$$

Substituting $\lambda = 1$, $\beta = k+1$, and $y = \lambda$, we have

$$f_{Y_{1,k+1}}(\lambda) = \frac{\lambda^k}{\Gamma(k+1)}e^{-\lambda} = \frac{\lambda^k}{k!}e^{-\lambda} = p(k;\lambda).$$

**4.15* Mean and variance of the normal distribution.**

$$\begin{aligned}
E[X] &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} xe^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx + \mu\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx \\
&\overset{(a)}{=} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ye^{-\frac{y^2}{2}}\,dy + \mu\int_{-\infty}^{\infty}\phi(y)\,dy \\
&= \int_{-\infty}^{\infty} y\phi(y)dy + \mu = 0 + \mu = \mu, \\
\mathrm{Var}[X] = E[(X-\mu)^2] &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}(x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\
&\overset{(b)}{=} \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}}dy \\
&= \sigma^2\int_{-\infty}^{\infty} y^2\phi(y)dy = \sigma^2,
\end{aligned}$$

where in (a) and (b), the change of variables $y = \frac{x-\mu}{\sigma}$ is made.

**4.16*** $\mathbf{\Gamma}(1/2)$**.** Since $\phi(u)$ is a PDF, we have

$$\int_{-\infty}^{\infty}\phi(u)du = 1.$$

We have

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du \overset{(i)}{=} 2\int_0^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}du \\
&\overset{(ii)}{=} \frac{1}{\sqrt{\pi}}\int_0^\infty x^{-\frac{1}{2}}e^{-x}dx \overset{(iii)}{=} \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{1}{2}\right),
\end{aligned}$$

where $(i)$ follows because $e^{-z^2}$ is an even function of $z$, $(ii)$ follows from the change of variables $x = \frac{u^2}{2}$, and $(iii)$ follows from the definition of $\Gamma(\beta)$ in (3.164). Hence,

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

**4.17 The mean, second moment and variance of the Weibull distribution.** Show that the mean, second moment and variance of the standard Weibull distribution are given by (4.81) through (4.83).

The survivor function of the standard Weibull distribution is

$$F_X^c(x) = e^{-x^\alpha}, \ \ x \geq 0.$$

Using the formula for nonnegative RVs, we obtain

$$E[X] = \int_0^\infty e^{-x^\alpha}\, dx$$

By setting $x^\alpha = y$, or $x = y^{1/\alpha}$, we rewrite the above integration as

$$\begin{aligned} E[X] &= \int_0^\infty e^{-y}\frac{1}{\alpha}y^{\frac{1}{\alpha}-1}\, dy \\ &= \frac{1}{\alpha}\Gamma\left(\frac{1}{\alpha}\right) = \Gamma\left(\frac{1}{\alpha}+1\right), \end{aligned} \tag{7}$$

where we used the formulas for gamma function:

$$\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}\, dx \ \text{ and } \ \Gamma(t+1) = t\Gamma(t).$$

Similarly, the second moment is calculated as

$$\begin{aligned} E[X^2] &= 2\int_0^\infty xF_X^c(x)\, dx = 2\int_0^\infty xe^{-x^\alpha}\, dx \\ &= 2\int_0^\infty \frac{x^2}{\alpha y}e^{-y}\, dy = \frac{2}{\alpha}\int_0^\infty y^{\frac{2}{\alpha}-1}e^{-y}\, dy \\ &= \frac{2}{\alpha}\Gamma\left(\frac{2}{\alpha}\right) = \Gamma\left(\frac{2}{\alpha}+1\right). \end{aligned} \tag{8}$$

Therefore, the variance is given by

$$\text{Var}[X] = \sigma_X^2 = \left[\Gamma\left(\frac{2}{\alpha}+1\right) - \Gamma^2\left(\frac{1}{\alpha}+1\right)\right].$$

**4.18 The median and mode of the Weibull distribution.**

(a) Let $X$ be a Weibull RV, with complementary distribution function

$$F_X^c(x) = e^{-(x/\beta)^\alpha}, \quad x \geq 0. \tag{9}$$

The median is defined as a value $x$, which satisfies

$$P[X \leq x] = P[X \geq x].$$

For a continuous RV $X$, this condition is equivalent to:

$$F_X(x) = 1 - F_X(x) \Rightarrow F_X(x) = \frac{1}{2},$$

or

$$F_X^c(x) = \frac{1}{2}. \tag{10}$$

Substituting (9) into (10) and solving for $x$ yields the median as

$$x = \beta(\ln 2)^{1/\alpha}.$$

(b) The pdf of $X$ can be found to be

$$f_X(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-(x/\beta)^\alpha}. \tag{11}$$

The mode of the Weibull distribution occurs at the root of the equation

$$f'_X(x) = 0. \tag{12}$$

Solving (12) yields the mode as

$$x = \beta(1 - 1/\alpha)^{1/\alpha}.$$

**4.19 The expectation and variance of two-parameter and three-parameter Weibull distributions**

(a) The expectation and variance of the two-parameter Weibull are obtained by scaling those of the standard Weibull by $\beta$ and $\beta^2$, respectively.

(b) The expectation is shifted by $\gamma$, but the variance remains the same as that of the two-parameter distribution, because the shape of the distribution does not change by shifting.

**4.20 Residual lifetime of Weibull distribution.**

The quantity we are after is the probability that the **residual lifetime** (see p. 414 for a discussion on residual lifetime) $R = X - t$ falls between 0 and $dt$, given the current age is $t$. Since $0 \geq R < dt$ if and only if $t < X \leq t$, this probability can be written by the conditional probability.

$$P\left[t < X \leq t + \, dt | X > t\right] = \frac{f_X(t)\,dt}{P\left[X > t\right]} = \frac{f_X(t)\,dt}{1 - F_X(t)} \triangleq h_X(t)\,dt,$$

where $h_X(t)$ is called the **hazard function** of the distribution $F_X(t)$ (see pp. 148-149, Section 6.3.3). For the standard Weibull distribution with parameter $\alpha$, we find from (4.78) and (4.79)

$$h_X(t) = \alpha x^{\alpha-1}.$$

Thus, the probability that we are asked to obtain is given by

$$h_X(t)\,dt = \alpha x^{\alpha-1}\,dt.$$

## 4.3 Joint and Conditional Probability Density Functions

**4.21* Joint bivariate normal distribution and ellipses.** The level curves are determined by the locus of points $(u_1, u_2)$ satisfying

$$\phi_0(u_1, u_2) = K,$$

where $K$ is a constant. Substituting for $\phi_0(u_1, u_2)$, we have

$$\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(u_1^2 - 2\rho u_1 u_2 + u_2^2)\right\} = K.$$

Taking logarithms on both sides and re-arranging terms, we obtain

$$u_1^2 - 2\rho u_1 u_2^2 = K_1, \tag{13}$$

where

$$K_1 = -2(1-\rho^2)\log[2\pi\sqrt{1-\rho^2}].$$

Re-arranging the left-hand side of (13), we obtain

$$(u_1 - \rho u_2)^2 - \rho^2 u_2^2 + u_2^2 = (u_1 - \rho u_2)^2 + u_2^2(1-\rho^2).$$

Using the transformation $x = u_1 - \rho u_2$, $y = u_2$ in (13), we have

$$x^2 + y^2(1-\rho^2) = K_1,$$

which is equivalent to

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

where $a = \sqrt{K_1}$ and $b = \sqrt{\frac{K_1}{1-\rho^2}}$.

**4.22* Conditional multivariate normal distribution.**

From the definition of the conditional PDF we have

$$f_{\boldsymbol{X}_b|\boldsymbol{X}_a}(\boldsymbol{x}_b|\boldsymbol{x}_a) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{f_{\boldsymbol{X}_a}(\boldsymbol{x}_a)} = \frac{(2\pi)^{m/2}|\det\boldsymbol{\Sigma}_{aa}|^{1/2}}{(2\pi)^{n/2}|\det\boldsymbol{\Sigma}|^{1/2}} \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}}{\exp\left\{-\frac{1}{2}(\boldsymbol{x}_a-\boldsymbol{\mu}_a)^\top\boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{x}_a-\boldsymbol{\mu}_a)\right\}} \tag{14}$$

where

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}. \tag{15}$$

Since $\boldsymbol{x}_a$ is fixed, we need to analyse only the exponent of the numerator in the RHS of equation (14):

It is easy to verify that

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{aa}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1} & \boldsymbol{I} \end{bmatrix} \tag{16}$$

where $\boldsymbol{S} = [\boldsymbol{\Sigma}_{bb} - \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}]$ is the Schur complement of $\boldsymbol{\Sigma}_{aa}$. Using this expression we find

$$\begin{aligned} &(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}_a-\boldsymbol{\mu}_a)^\top\boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{x}_a-\boldsymbol{\mu}_a) \\ &+[-(\boldsymbol{x}_a-\boldsymbol{\mu}_a)^\top\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab} + (\boldsymbol{x}_b-\boldsymbol{\mu}_b)^\top]\boldsymbol{S}^{-1}[-\boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{x}_a-\boldsymbol{\mu}_a) + (\boldsymbol{x}_b-\boldsymbol{\mu}_b)]. \end{aligned} \tag{17}$$

Since $\boldsymbol{x}_a$ is fixed, we are interested only the in terms that depend on $\boldsymbol{x}_b$. The previous equation can be written as

$$(\boldsymbol{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}_b-\boldsymbol{\mu}_b)^\top\boldsymbol{S}^{-1}(\boldsymbol{x}_b-\boldsymbol{\mu}_b) - 2(\boldsymbol{x}_b-\boldsymbol{\mu}_b)^\top\boldsymbol{b} + const. \tag{18}$$

where

$$\boldsymbol{b} = \boldsymbol{S}^{-1}\boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{x}_a-\boldsymbol{\mu}_a).$$

It is not difficult to verify by direct multiplication that

$$(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}_b-\boldsymbol{\mu}_b-\boldsymbol{S}\boldsymbol{b})^\top \boldsymbol{S}^{-1}(\boldsymbol{x}_b-\boldsymbol{\mu}_b-\boldsymbol{S}\boldsymbol{b}) + const \tag{19}$$

where

$$\boldsymbol{S}\boldsymbol{b} = \boldsymbol{S}\boldsymbol{S}^{-1}\boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{x}_a-\boldsymbol{\mu}_a) = \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{x}_a-\boldsymbol{\mu}_a).$$

Thus,

$$f_{\boldsymbol{X}_b|\boldsymbol{X}_a}(\boldsymbol{x}_b|\boldsymbol{x}_a) \sim \exp\left\{-\frac{1}{2}(\boldsymbol{x}_b-\boldsymbol{\mu}_{b|a})^\top \boldsymbol{\Sigma}_{b|a}^{-1}(\boldsymbol{x}_b-\boldsymbol{\mu}_{b|a})\right\} \tag{20}$$

where

$$\begin{array}{l} \boldsymbol{\mu}_{b|a} = \boldsymbol{\mu}_b + \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{x}_a-\boldsymbol{\mu}_a) \\ \boldsymbol{\Sigma}_{b|a} = \boldsymbol{S} = \boldsymbol{\Sigma}_{bb} - \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab} \end{array}. \tag{21}$$

**4.23 Circularly symmetric and independent RVs [262]**

Differentiating both sides with respect (w.r.t.) to $x$, we have

$$\frac{\partial g(r)}{\partial x} = \frac{dg(r)}{dr}\frac{\partial r}{\partial x}, \ \text{ where } \frac{\partial r}{\partial x} = \frac{x}{r}.$$

From the independence assumption, we have

$$g(r) = f_X(x) f_Y(y).$$

By differentiating the above w.r.t. $x$, we find

$$\frac{x}{r}g'(r) = f'_X(x) f_Y(y),$$

which leads to

$$\frac{1}{r}\frac{g'(r)}{g(r)} = \frac{1}{x}\frac{f'_X(x)}{f_X(x)}.$$

The right side is a function of $x$ only and is independent of $y$, and the left side is a function of $r = \sqrt{x^2+y^2}$. Then the above must be a constant, i.e.,

$$\frac{1}{r}\frac{g'(r)}{g(r)} = \frac{1}{r}\frac{d\ln g(r)}{dr} = \alpha \text{ (constant)},$$

which leads to

$$\ln g(r) = \frac{\alpha r^2}{2} + c,$$

where $c$ is some constant. Thus

$$g(r) = Ae^{\frac{\alpha r^2}{2}} = Ae^{\frac{\alpha(x^2+y^2)}{2}}.$$

Thus, $X$ and $Y$ are normally distributed with zero mean and variance $-\alpha^{-1}$.

**4.24 Buffon's Needle Problem**

It is clear that the needle does not cross the upper line if

$$Y' = Y + \sin\Theta < 1.$$

Since the needle is dropped on the table in an random fashion, we can assume that the variables $Y$ and $\Theta$ that describe the position of the needle are independent RVs, i.e.,

$$f_{Y\Theta}(y,\theta) = f_Y(y) f_\Theta(\theta).$$

Furthermore, $Y$ is uniformly distributed in $(0,1]$ and $\Theta$ is uniform over $(0,\pi]$, that is

$$f_Y(y) = 1,\ \ 0 < y \le 1, \quad \text{and} \quad f_\Theta(\theta) = \frac{1}{\pi},\ \ 0 < \theta \le \pi.$$

Because of the statistical independence between $Y$ and $\Theta$, the conditional PDF of $Y$ given $\Theta$ is just the PDF of $Y$:

$$f_{Y|\Theta}(y|\theta) = f_Y(y),\ \ 0 < y \le 1,\ 0 < \theta \le \pi.$$

Thus, for given $\Theta$, the conditional probability that the needle does not cross a line is

$$P[Y < 1 - \sin\Theta] = \int_0^{1-\sin\Theta} f_Y(y)\,dy = 1 - \sin\Theta.$$

Taking the complement, the probability that the line will be hit, given the angle $\Theta$ is

$$h(\Theta) = P[Y > 1 - \sin\Theta] = \sin\Theta.$$

Then taking the expectation with respect to the angle variable $\Theta$, we obtain the probability that the needle hits a line as

$$\begin{aligned} E[h(\Theta)] &= \int_0^\pi \sin\theta\, f_\Theta(\theta)\,d\theta = \frac{1}{\pi}\int_0^\pi \sin\theta\,d\theta \\ &= \frac{1}{\pi}[-\cos\theta]_0^\pi = \frac{2}{\pi} \approx .6366197. \end{aligned} \tag{22}$$

Thus, from this result the value of $\pi$ can be estimated by conducting an experiment of dropping a needle many times (or many needles). Let $N$ be the total drops, and $N_h$ be the number of hits. Then

$$2N/N_h \approx \pi.$$

**4.25 Modifications of Buffon's needle experiment.**

(a) We proceed in the same way as in the original problem. The only modification is that now we have the condition of not hitting a line as

$$X < a - \ell\sin\Theta.$$

Hence, the conditional probability $h(\Theta)$ that the needle touches the line $Y = a$ is found by

$$1 - h(\Theta) = \int_0^{a-\ell\sin\Theta} f_Y(y)\,dy.$$

Now the RV $Y$ has the PDF $f_Y(y) = \frac{1}{a},\ \ 0 < y \le a$, we find

$$1 - h(\Theta) = 1 - \frac{\ell}{a}\sin\Theta,$$

from which we have

$$h = E[h(\Theta)] = \frac{2\ell}{a\pi}.$$

**Note:** If the needle length is greater than the line spacing, the problem becomes complicated.

(b) Now we assume $X < X'$ and $Y < Y'$, hence the angle variable $\Theta$ is uniformly distributed over $(0, \frac{\pi}{2}]$. This particular case and the other three cases, i.e., $\Theta \in (\frac{\pi}{2}, \pi], (\pi, \frac{3}{2}\pi]$ and $(\frac{3}{2}\pi, 2\pi]$ occur with equal probability (i.e., 1/4), so we can focus on this case.
The condition that the needle does not cross over any line is given by the following joint event

$$Y < a - \ell \sin\Theta \ \text{ and } \ X < b - \ell\cos\Theta.$$

The RVs $X$, $Y$, and $\Theta$ are statistically independent and are uniformly distributed over $(0, b], (0, a]$ and $(0, \frac{\pi}{2}]$, respectively. For given $\Theta$, let the conditional probability that the needle hits any line be denoted $h(\Theta)$. Then

$$1 - h(\Theta) = \left(1 - \frac{\ell}{a}\sin\Theta\right)\left(1 - \frac{\ell}{b}\cos\Theta\right).$$

Now that the PDF of $f_\Theta(\theta) = \frac{2}{\pi}, \ \ 0 < \theta \leq \frac{\pi}{2}$, we find the probability that the needle hits any line is given by

$$\begin{aligned} h = E[h(\Theta)] &= \frac{2}{\pi}\frac{\ell}{ab}\int_0^{\frac{\pi}{2}}\left(b\sin\theta + a\cos\theta - \frac{\ell}{2}\sin 2\theta\right)d\theta \\ &= \frac{\ell(2a + 2b - \ell)}{ab\pi}. \end{aligned}$$

If in particular $a = b = \ell$, then $h = \frac{3}{\pi}$.

## 4.4 Exponential Family of Distributions

**4.26* Exponential families of distributions**

(a) exponential distributions with PDF given by (4.25), parameterized by $\lambda$.
(b) gamma distributions with PDF given by (4.30), parameterized by $(\lambda, \beta)$.
(c) binomial distributions given by (3.62), parameterized by $(n, p)$.
(d) negative binomial (Pascal) distributions given by (3.98), parameterized by $(r, p)$.

**4.27 Uniform distribution.** TBD
**4.28 Weibull distribution.** TBD
**4.29 Pareto distribution.** TBD

## 4.5 Bayesian Inference and Conjugate Priors

**4.30* Posterior hyperparameters of the beta distribution associated with the Bernoulli distribution in Example 4.4.**

(a)

$$\begin{aligned} E[\Theta|\boldsymbol{x}] &= \frac{\alpha_1}{\alpha_1+\beta_1} = \frac{\alpha+\sum_{i=1}^n x_i}{\alpha+\beta+n} \\ &= \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\frac{\alpha}{\alpha+\beta} + \left(\frac{n}{\alpha+\beta+n}\right)\overline{x}_n. \end{aligned}$$

(b)

$$\begin{aligned} \mathrm{Var}[\Theta|\boldsymbol{x}] &= \frac{\alpha_1\beta_1}{(\alpha_1+\beta_1)^2(\alpha_1+\beta_1+1)} \\ &= \frac{(\alpha+\sum_{i=1}^n x_i)(\beta+n-\sum_{i=1}^n x_i)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)}. \end{aligned}$$

**4.31 Conjugate prior for a geometric distribution.**
We set the prior distribution to the beta distribution $\mathrm{Beta}(p;\alpha,\beta)$ defined in (4.140):

$$\pi(p) = \mathrm{Beta}(p;\alpha,\beta) \propto p^{\alpha-1}(1-p)^{\beta-1}, \;\; 0 \le p \le 1, \tag{23}$$

with prior hyperparameters $\alpha > 0$ and $\beta > 0$. Applying (4.137), we readily obtain the posterior distribution

$$\pi(p|x) \propto p^{\alpha+\sum_{i=1}^n x_i - 1}(1-p)^{\beta+n-1} \tag{24}$$

Hence, the posterior hyperparameters are

$$\alpha_1 = \alpha + \sum_{i=1}^n x_i \;\text{ and }\; \beta_1 = \beta + n.$$

**4.32 Conjugate prior for a multinomial distribution.**
The multinomial distribution is given by

$$p(\boldsymbol{x}|\boldsymbol{p}) = \frac{n!}{x_1!x_2!\cdots x_m!}p_1^{x_1}p_2^{x_2}\cdots p_m^{x_m}, \tag{25}$$

where $\sum_{i=1}^m p_i = 1$ and $x_i$'s are nonnegative integers such that $\sum_{i=1}^m x_i = n$. The Dirichlet distribution of $\boldsymbol{p}$ given by (4.172) and (4.173) can be written as

$$\pi(\boldsymbol{p}) = \frac{\Gamma\left(\sum_{i=1}^m \alpha_i\right)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\cdots\Gamma(\alpha_m)}p_1^{\alpha_1-1}p_2^{\alpha_2-1}\cdots p_m^{\alpha_m-1}, \tag{26}$$

which is an $m$-dimensional extension of the beta distribution given by (4.140). Note also its similarity to the multinomial distribution (25). The integers $x_i$ are replaced by real numbers $\alpha_i - 1$, hence the factorials $n_i!$, by $\Gamma(\alpha_i)$. For mathematically rigorous derivations of (4.173) see Notes 1 and 2 given below.

The posterior distribution of $\boldsymbol{p}$ is

$$\pi(\boldsymbol{p}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{p})\pi(\boldsymbol{p})}{p(\boldsymbol{x})} \propto \prod_{i=1}^{m} p_i^{x_i+\alpha_i-1}, \tag{27}$$

and it is not difficult to find its normalization constant from that in (26), and we obtain

$$\pi(\boldsymbol{p}|\boldsymbol{x}) = \frac{\Gamma\left(\sum_{i=1}^{m}(\alpha_i + x_i)\right)}{\Gamma(\alpha_1 + x_1)\Gamma(\alpha_2 + x_2)\cdots\Gamma(\alpha_m + x_m)} p_1^{\alpha_1+x_1-1} p_2^{\alpha_2+x_2-1} \cdots p_m^{\alpha_m+x_m-1}. \tag{28}$$

The distribution (26) can be interpreted as the distribution that event $i$ has been observed $\alpha_i$ times. By having $x_i$ additional occurrences, the posterior distribution has hyperparameters $\alpha_i + x_i$'s, as is shown in (28).

**Note 1: Derivation of Eq. (4.173).**

From the definition of the gamma-function we have

$$\prod_{i=1}^{m}\Gamma(\alpha_i) = \prod_{i=1}^{m}\int_0^\infty \exp\{-u_i\}u_i^{\alpha_i-1}\,du_i = \int_{\boldsymbol{U}} \exp\{-\sum_{i=1}^{m}u_i\}\prod_{i=1}^{m}u_i^{\alpha_i-1}\,du_i \tag{29}$$

where

$$\boldsymbol{U} : \{0 \le u_i < \infty\}. \tag{30}$$

In this multidimensional integral, perform the following variable substitutions

$$z = \sum_{i=1}^{m} u_i,\ u_i = t_i\,z,\ i = 1, 2, \ldots, m-1. \tag{31}$$

The Jacobian matrix of this substitution is

$$\boldsymbol{J} = \begin{bmatrix} t_1 & z & 0 & \cdots & 0 \\ t_2 & 0 & z & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{m-1} & 0 & 0 & \cdots & z \\ (1 - t_1 - \cdots - t_{m-1}) & -z & -z & \cdots & -z \end{bmatrix} \tag{32}$$

Decomposing the determinant of this matrix by the elements of the last row, we obtain

$$\det \boldsymbol{J} = (-1)^{m-1}[(1 - t_1 - \cdots - t_{m-1})z^{m-1} + t_1 z^{m-1} + \cdots + t_{m-1}z^{m-1}] = (-1)^{m-1}z^{m-1}. \tag{33}$$

Thus, the previous integral can be written as

$$\int_{z=0}^{\infty} e^{-z} z^{\sum_{i=1}^{m}\alpha_i - 1}\,dz \int_{\boldsymbol{D}} \prod_{i=1}^{m-1} t_i^{\alpha_i-1}(1 - \sum_{i=1}^{m-1} t_i)^{\alpha_m-1}\,dt_i. \tag{34}$$

where

$$\boldsymbol{D} : \{0 \le t_i,\ \sum_{i=1}^{m-1} t_i \le 1\}. \tag{35}$$

As we can see, this integral is $\Gamma(\sum_{i=1}^m \alpha_i)B(\boldsymbol{\alpha})$, i.e.,

$$\prod_{i=1}^m \Gamma(\alpha_i) = \Gamma(\sum_{i=1}^m \alpha_i)B(\boldsymbol{\alpha}),$$

from which (4.173) follows.

**Note 2: A Simple Derivation of Eq. (4.173) using the Laplace Transform.**
If the reader is already familiar with the Laplace transform discussed in Section 9.2 (pp. 226-230), here is a simpler and quicker derivation of the above result. From p. 228, Table 9.5 No. 8, the Laplace transform of a power function is given by

$$\int_0^\infty e^{-sx} x^{\alpha-1}\, dx = \frac{\Gamma(\alpha)}{s^\alpha} \tag{36}$$

The Laplace transform of the convolution of two functions is the product of their Laplace transforms (p. 230, Table 9.7, No. 2):

$$\mathcal{L}\left\{\int_0^x u^{\alpha_1-1}(x-u)^{\alpha_2-1}\, du\right\} = \frac{\Gamma(\alpha_1)}{s^{\alpha_1}}\frac{\Gamma(\alpha_2)}{s^{\alpha_2}} = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{s^{\alpha_1+\alpha_2}} \tag{37}$$

Thus, taking the inverse Laplace transform of the above and using the formula (36), we find

$$\begin{aligned} \int_0^x u^{\alpha_1-1}(x-u)^{\alpha_2-1}\, du &= \mathcal{L}^{-1}\left\{\frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{s^{\alpha_1+\alpha_2}}\right\} \\ &= \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}\, x^{\alpha_1+\alpha_2-1} \end{aligned} \tag{38}$$

Setting $x = 1$ gives (4.141) (which has a typo: the integration is from 0 to 1, not to $\infty$)

$$\int_0^1 u^{\alpha_1-1}(1-u)^{\alpha_2-1}\, du = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} \tag{39}$$

Equation (4.173) can be derived by applying the above convolution formula multiple times.

# 5 Solutions for Chapter 5: Functions of Random Variables and Their Distributions

## 5.1 A Function of one random variable

**5.1* Half-wave rectifier.**

$$F_Y(y) = \begin{cases} 0, & y < 0 \\ F_X(0) & y = 0 \\ F_X(y) & y > 0 \end{cases}$$

Hence

$$f_Y(y) = \begin{cases} 0, & y < 0 \\ F_X(0)\delta(y) & y = 0 \\ f_X(y) & y > 0 \end{cases}$$

**5.2 Square law detector-continued.**

(a) We note that $Y = g(X) = X^2$, and as $X \in [-1, 1]$ we must have $Y \in [0, 1]$. The PDF of $X$ is:

$$f_X(x) = \begin{cases} 1/2 & \mid x \mid \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

To calculate the CDF of $Y$, we use the CDF of $X$:

$$\begin{aligned} F_Y(y) &= P[X^2 \leq y] \\ &= P[-\sqrt{y} \leq X \leq \sqrt{y}] \\ &= \begin{cases} 2\frac{\sqrt{y}}{2} & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Taking derivatives with respect to $y$, we have

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) \\ &= \begin{cases} \frac{1}{2\sqrt{y}} & 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

(b) We first find the PDF of $Y = X^2$, then the PDF after normalization, $Z = Y/\sigma^2$.

(i) Our approach here will be very similar to the approach in **(a)**.
Since the support of $X$ is now: $X \in (-\infty, \infty)$, we have the support of $Y \in [0, \infty)$.
Calculating the CDF of $Y$:

$$\begin{aligned} F_Y(y) &= P[X^2 \leq y] \\ &= P[-\sqrt{y} \leq X \leq \sqrt{y}] \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ &= \Phi(\sqrt{y}/\sigma) - \Phi(-\sqrt{y}/\sigma) && y \geq 0 \\ &= \Phi(\sqrt{y}/\sigma) - (1 - \Phi(\sqrt{y}/\sigma)) && y \geq 0 \end{aligned}$$

where we use the well known property of symmetry in the normal distribution: $\Phi(-z) = 1 - \Phi(z)$.
The PDF of $Y$ is:

$$\begin{aligned} f_Y(y) &= \frac{\partial}{\partial y} F_Y(y) \\ &= 2\left(\frac{1}{2\sqrt{y}\sigma}\right)\phi(\sqrt{y}/\sigma)) && y \geq 0 \\ &= \frac{1}{\sqrt{y}\sigma}\phi(\sqrt{y}/\sigma)u(y) \\ &= \frac{1}{\sqrt{2\pi y}\sigma}\exp\left(-\frac{y}{2\sigma^2}\right)u(y) \end{aligned}$$

(ii) We have $Z = g(Y) = Y/\sigma^2$. $g(y)$ is monotonic in $y$ and therefore one-to-one. We note that the preimage is $g^{-1}(z) \triangleq \sigma^2 z$. There is one solution for any given $z$, and therefore $\forall z, m(z) = 1$. We can calculate $g'(y) = 1/\sigma^2$.
We can now directly calculate the PDF of $Z$:

$$\begin{aligned} f_Z(z) &= \frac{f_Y(g_1^{-1}(z))}{\mid g'(g^{-1}(z)) \mid} \\ &= \sigma^2 \frac{1}{\sqrt{2\pi\sigma^2 z}\sigma}\exp\left(-\frac{\sigma^2 z}{2\sigma^2}\right)u(\sigma^2 z) \\ &= \frac{1}{\sqrt{2\pi z}}\exp\left(-\frac{z}{2}\right)u(z) \end{aligned}$$

It is important to note that $Z = (X/\sigma)^2$ has the $\mathcal{X}_1^2$ distribution, i.e., the Chi-square distribution with one degree of freedom (df), which may be defined as the distribution that we get by squaring the unit normal RV $N(0, 1)$.

**5.3 Exponential-law detector.**
The exponential function is monotone increasing, so $Y \leq y$ if and only if $X \leq \ln y$. Thus

$$F_Y(y) = \begin{cases} 0, & y \leq 0 \\ F_X(\ln y), & y > 0. \end{cases}$$

By differentiating

$$f_Y(y) = \begin{cases} 0, & y \le 0 \\ \frac{f_X(\ln y)}{y}, & y > 0. \end{cases}$$

For $F_X(x) = 1 - e^{-\lambda x}$, we have $f_X(x) = \lambda e^{-\lambda x}$. For $x \ge 0$, we find $y \ge 1$. Also $F_X(\ln y) = 1 - e^{\lambda \ln y} = 1 - y^{-\lambda}$. Thus,

$$F_Y(y) = \begin{cases} 0, & y < 1 \\ 1 - y^{-\lambda}, & y \ge 1. \end{cases}$$

and

$$f_Y(y) = \begin{cases} 0, & y < 1, \\ \lambda y^{-(\lambda+1)}, & y \ge 1. \end{cases}$$

**5.4 Cauchy distribution.**

$$y = \tan x = \frac{\sin x}{\cos x} \Longrightarrow \frac{dy}{dx} = \frac{\cos^2 x + \sin^2}{\cos^2 x} = 1 + \tan^2 x = 1 + y^2.$$

Hence, by equating $f_Y(y)\,dy = f_X(x)\,dx$, we readily have

$$f_Y(y) = f_X(x)/|\frac{dy}{dx}| = \frac{1/\pi}{1+y^2}.$$

**5.5 Inverse of a random variable and the Cauchy distribution.**

(a) $F_Y(y) = P[Y \le y] = P[1/X \le y]$. If $y > 0$, we have

$$P[Y \le y] = P[X \ge 1/y \text{ or } X < 0] = F_X^c(1/y) + F_X(0).$$

If $y < 0$, we have

$$P[Y \le y] = P[1 \ge Xy] = P[1/y \le X] = 1 - F_X(1/y).$$

Therefore,

$$F_Y(y) = \begin{cases} 1 - F_X(1/y) + F_X(0^-), & y > 0, \\ 1 - F_X(1/y), & y < 0, \end{cases}$$

Hence, we obtain the pdf

$$f_Y(y) = \frac{1}{y^2} f_X(1/y).$$

(b) Using (**??**),

$$f_Y(y) = \frac{1}{y^2} \frac{\alpha/\pi}{(1/y)^2 + \alpha^2} = \frac{1/(\alpha\pi)}{(1/\alpha^2) + y^2},$$

which is a Cauchy distribution with parameter $1/\alpha$.

## 5.2 A Function of Two Random Variables

**5.6* Leibniz's rule.**

(a) The LHS of (5.95) equals

$$\frac{d}{dz}[H(b(z)) - H(a(z))] = H'(b(z))b'(z) - H'(a(z))a'(z)$$
$$= h((bz))b'(z) - h(a(z))a'(z).$$

(b) The LHS of (5.94) equals

$$\begin{aligned}\frac{d}{dz}[H(z,b(z)) - H(z,a(z))] &= [g(z,b(z)) + h(z,b(z))b'(z)] - [g(z,a(z)) + h(z,a(z))a'(z)]\\ &= h(z,b(z))b'(z) - h(z,a(z))a'(z) + [g(z,b(z)) - g(z,a(z))]\\ &= h(z,b(z))b'(z) - h(z,a(z))a'(z) + \int_{a(z)}^{b(z)} \frac{\partial h(z,y)}{\partial z}\,dy,\end{aligned}$$

where we used the following relation in the last step:

$$\frac{\partial g(z,y)}{\partial y} = \frac{\partial}{\partial y}\frac{\partial H(z,y)}{\partial z} = \frac{\partial}{\partial z}\frac{\partial H(z,y)}{\partial y} = \frac{\partial h(z,y)}{\partial z}.$$

(c) Then

$$\frac{\partial G}{\partial a} = -h(z,a), \quad \frac{\partial G}{\partial b} = h(z,b), \text{ and } \frac{\partial G}{\partial z} = \int_a^b \frac{\partial}{\partial z} h(z,y)\,dy.$$

Substitution of these into (5.96) yields the Leibniz's rule (5.94).

**5.7 Sum of uniform variables.** Let $X$ and $Y$ be independent uniform random variables.

(a) Since $f_Y(y) = 1$ for $y \in (0,1)$ and $f_X(z-y) = 1$ for $y_1(z-1,z)$, we have

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)\,dy = \int_{\max\{z-1,0\}}^{\min\{z,1\}} dy.$$

Thus,

$$f_Z(z) = \begin{cases} 0, & z < 0,\\ \int_0^z dy = z, & 0 \le z \le 1,\\ \int_{z-1}^1 dy = 1 - (z-1) = 2 - z. & 1 \le z \le 2,\\ 0, & z > 2.\end{cases}$$

$$F_Z(z) = \begin{cases} 0, & z < 0,\\ \int_0^z z\,dy = \frac{z^2}{2}, & 0 \le z \le 1,\\ \frac{1}{2} + \int_1^z (2-z)\,dy = 2z - 1 - \frac{z^2}{2} & 1 \le z \le 2,\\ 1, & z > 2.\end{cases}$$

(b) Since $f_Y(y) = 1/b$ for $y \in (0, b)$ and $f_X(z-y) = 1/a$ for $y_1(z-a, z)$, we have

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy = \frac{1}{ab} \int_{\max\{z-a,0\}}^{\min\{z,b\}} dy.$$

Thus,

$$f_Z(z) = \begin{cases} 0, & z < 0, \\ \frac{1}{ab}\int_0^z dy = \frac{z}{ab}, & 0 \le z \le a, \\ \frac{1}{ab}\int_{z-a}^z dy = \frac{a}{ab} = \frac{1}{b}. & a \le z \le b, \\ \frac{1}{ab}\int_{z-a}^b dy = \frac{a+b-z}{ab}, & b \le z \le a+b, \\ 0, & z > a+b. \end{cases}$$

$$F_Z(z) = \begin{cases} 0, & z < 0, \\ \frac{z^2}{2ab}, & 0 \le z \le a, \\ \frac{a}{2b} + \int_a^z \frac{1}{b}\, dz = \frac{2z-a}{2b} & a \le z \le b, \\ 1 - \frac{a}{2b} + \int_b^z \frac{a+b-z}{ab}\, dz = \frac{(a+b)z}{ab} - \frac{a^2+b^2}{2ab} - \frac{z^2}{2ab}, & b \le z \le a+b, \\ 1, & z > a+b. \end{cases}$$

**5.8 Sum of exponential variables.**

(a) Since $f_X(z-y)$ is nonzero for $y < z$ and $f_Y(y)$ is nonzero for $y \ge 0$, we have

$$\begin{aligned} f_Z(z) &= \int_0^z f_X(z-y) f_Y(y)\, dy = \lambda\mu e^{-\lambda z} \int_0^z e^{-(\mu-\lambda)y}\, dy \\ &= \lambda\mu e^{-\lambda z} \frac{1 - e^{-(\mu-\lambda)z}}{(\mu-\lambda)z} = \frac{\lambda\mu}{\mu-\lambda}\left(e^{-\lambda z} - e^{-\mu z}\right), \quad \text{for } \mu \ne \lambda. \end{aligned}$$

(b) From the second line in the above equation, we have for $\mu - \lambda = 0$,

$$f_Z(z) = \lambda\mu e^{-\lambda z} \int_0^z dy = \lambda\mu z e^{-\lambda z}$$

**5.9 Difference of two random variables.**

Let $Z = X - Y$ and $W = Y$. Then the transformation from $(X, Y)$ to $(Z, W)$ has the Jacobian of 1. Thus,

$$f_{ZW}(z, w) = f_{XY}(z+w, w).$$

The marginal distribution of $Z$ is

$$f_Z(z) = \int f_{XY}(z+w, w)\, dw = \int f_{XY}(z+y, y)\, dy.$$

If $X$ ad $Y$ are independent

$$f_{XY}(x, y) = f_X(x) f_Y(y).$$

Hence

$$f_Z(z) = \int f_X(z+y) f_Y(y)\, dy.$$

This is similar to convolution. Indeed if we write $Z = X + (-Y)$. The PDF of $Y' = -Y$ is $f_Y(-y')$. So the PDF of $Z$ is the convolution of $f_X(x)$ and $f_Y(-y')$:

$$\int f_X(z-y')f'_Y(y')\,dy' = \int f_X(z-y')f_Y(-y')\,dy' = \int f_X(z+y)f_Y(y)\,dy.$$

**5.10 Ratio of two random variables.**

(a) We have:

$$\begin{aligned} F_Z(z) = P[X/Y \le z] &= \int\int I\{x/y \le z\} f_{XY}(x,y)dx\,dy \\ &= \int_0^\infty \int_{-\infty}^\infty I\{x \le yz\} f_{XY}(x,y)dx\,dy + \int_{-\infty}^0 \int_{-\infty}^\infty I\{x \ge yz\} f_{XY}(x,y)dx\,dy \\ &= \int_0^\infty \int_{-\infty}^{yz} f_{XY}(x,y)dx\,dy + \int_{-\infty}^0 \int_{yz}^{-\infty} f_{XY}(x,y)dx\,dy \end{aligned}$$

Therefore,

$$f_Z(z) = \int_0^\infty y f_{XY}(yz,y)dy - \int_{-\infty}^0 y f_{XY}(x,y)dz,y)dy = \int_{-\infty}^\infty |y| f_{XY}(yz,y)dy.$$

(b) If $X$ and $Y$ are nonnegative RVs, then:

$$f_{XY}(x,y) = 0 \quad \text{for } x < 0 \text{ or } y < 0\ ,$$

which implies that

$$f_{XY}(yz,y) = 0 \quad \text{for } y < 0 \text{ or } yz < 0.$$

Therefore,

$$f_Z(z) = \begin{cases} \int_0^\infty y f_{XY}(yz,y)dy, & z \ge 0 \\ 0, & z < 0 \end{cases} \tag{1}$$

**5.11 Product of two random variables.**

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^\infty \int_{-\infty}^\infty I\{xy \le z\} f_{XY}(x,y)\,dxdy \\ &= \int_0^\infty \int_{-\infty}^\infty I\{y \le z/x\} f_{XY}(x,y)\,dxdy + \int_{-\infty}^0 \int_{-\infty}^\infty I\{y \ge z/x\} f_{XY}(x,y)\,dxdy \\ &= \int_0^\infty \int_{-\infty}^{z/x} f_{XY}(x,y)\,dxdy + \int_{-\infty}^0 \int_{z/x}^\infty f_{XY}(x,y)\,dxdy. \end{aligned}$$

Hence

$$f_Z(z) = \int_0^\infty \frac{1}{x} f_{XY}(x,z/x)\,dx - \int_{-\infty}^0 \frac{1}{x} f_{XY}(x,z/x)\,dx = \int_{-\infty}^\infty \frac{1}{|x|} f_{XY}(x,z/x)\,dx.$$

**5.12 Bivariate normal distribution and Cauchy distribution.**

(a) **Note:** The question should be corrected to $Z = (X_1 - \mu_1)/(X_2 - \mu_2)$ instead of $Z = X_1/X_2$.

We start with the simplest case first, and then generalize the result. **Case 1.** $\mu_1 = \mu_2 = 0$, $\rho = 0$ and $\sigma_1 = \sigma_2 = \sigma$.
The problem reduces to the case of Exercise **??**. (It is also related to the Rayleigh distribution discussed in Section **??**.) The RV $Z$ does not depend on $\sigma$ since the ratio $Z = X_1/X_2$ removes the scaling factor $\sigma$.
Note that $W = Z^{-1} = X_2/X_1$ and $Z = X_1/X_2$ have the same Cauchy distribution. In other words the inverse of a Cauchy RV has the same Cauchy distribution.
**Case 2.** $\mu_1 = \mu_2 = 0$, and $\rho = 0$, but $\sigma_1 \neq \sigma_2$.
Then we normalize $U_1 = X_1/\sigma_1,\ U_2/X_2/\sigma_2$. Then

$$Z = X_/X_2 = (\sigma_1/\sigma_2) \cdot (U_1/U_2).$$

The RV $T = U_1/U_2$ is the special case of Case 1, i.e., $\sigma = 1$, hence $T$ is Cauchy distributed:

$$f_T(t) = \frac{1}{\pi(1+t^2)},\ -\infty < t < \infty.$$

Since $dz/dt = \sigma_1/\sigma_2 \stackrel{\text{def}}{=} \alpha$, we obtain

$$\begin{aligned} f_Z(z) = \frac{f_T(t)}{\frac{dz}{dt}} &= \frac{1}{\pi\alpha(1+t^2)} \\ &= \frac{1}{\pi\alpha\left(1+\frac{z^2}{\alpha^2}\right)} \end{aligned}$$

**Case 3.** $\mu_1 = \mu_2 = 0$, but $\sigma_1 \neq \sigma_2$ and $\rho \neq 0$.

Since $\rho$ is non-zero, if we rotate the $X_1$ and $X_2$ axes so that under the new coordinates the correlation coefficient becomes zero. Then the problem reduces to Case 2. Let us rotate the axes by $\phi$ [radian] counter clockwise, and the new coordinates be $Y_1$ and $Y_2$. We can show that when

$$\tan\phi = \rho\sigma_1/\sigma_2 \stackrel{\text{def}}{=} \mu,$$

the correlation coefficients under the new coordinates becomes zero. Then the distribution is shifted by $\mu$, in other words the maximum of $Z = X_1/X_2$ occurs at $Z = \mu = \tan\phi$. Hence the distribution is simply translation of the case 2 distribution by $\mu$:

$$f_Z(z) = \frac{1}{\pi\alpha\left(1+\frac{(z-\mu)^2}{\alpha^2}\right)}$$

**Case 4.** No restriction.
The case $\mu_1 \neq 0$ and $\mu_2 \neq 0$ is no different from Case 3, since $Z = X_1 - \mu_1/X_2 - \mu_2$ is essentially the same random variable as $Z = X_1/X_2$ in case 3. Hence, the distribution is the same.

**5.13 Independent normal distribution and exponential distribution.**

We start with (5.34) of Example 5.4. Let $X_1 = X$ and $X_2 = Y$. Then for the given normal distribution $X, Y \sim N(0, \sigma^2)$, we find

$$f_{XY}(\sqrt{z-y^2}, y) = f_{XY}(-\sqrt{z-y^2}, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{z}{2\sigma^2}}.$$

Thus,

$$f_Z(z) = \frac{1}{2\pi\sigma^2} e^{-\frac{z}{2\sigma^2}} \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{z-y^2}}.$$

By setting $y/\sqrt{z} = t$,

$$f_Z(z) = \frac{1}{2\pi\sigma^2} e^{-\frac{z}{2\sigma^2}} \int_{-1}^{1} \frac{dt}{\sqrt{1-t^2}}\, dt = \frac{1}{2\sigma^2} e^{-\frac{z}{2\sigma^2}}.$$

In order to obtain the last expression, we used the formula $I = \int_{-1}^{1} \frac{dt}{\sqrt{1-t^2}}\, dt = \pi$, which can be easily derived by setting $t = \sin\theta$, because $I = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\cos\theta}{\sqrt{1-\sin^2\theta}}\, d\theta = \pi$.
Alternatively, we transform from the $(X_1, X_2)$ to $(R, \Theta)$, where

$$X_1 = R\cos\Theta, \quad X_2 = R\sin\Theta.$$

To find $f_{R\Theta}(r, \theta)$ from $f_{X_1X_2}(x_1, x_2)$, we need to compute the Jacobian matrix, which is

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}.$$

Hence, $|\boldsymbol{J}| = r$.

$$f_{X_1X_2}(x_1, x_2) J = f_{R\Theta}(r, \theta).$$

Thus,

$$f_{R\Theta}(r, \theta) = \frac{2}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}.$$

The marginal PDF of $R$ is

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}.$$

Finally, by using $Z = R^2$, which gives $dz = 2r\, dr$. From

$$f_R(r)\, dr = f_Z(z)\, dz,$$

we find

$$f_Z(z) = \frac{1}{2r} f_R(r) = \frac{1}{2\sigma^2} e^{-\frac{z}{2\sigma^2}}.$$

**5.14 Maximum of two random variables.**

(a)

$$\mathcal{D}_z = \{(x, y) : \max(x, y) \leq z\}$$

(b)

$$F_Z(z) = P[\max(X,Y) \le z] = P[X \le z \text{ and } Y \le z] = F_{XY}(z,z).$$

(c) If $X$ and $Y$ are independent, then

$$F_Z(z) = F_{XY}(z,z) = F_X(z)F_Y(z).$$

Therefore,

$$f_Z(z) = f_X(z)F_Y(z) + F_X(z)f_Y(z).$$

**5.15 Minimum of two random variables.**

(a)

$$\mathcal{D}_z = \{(x,y) : \min(x,y) \le z\}$$

Draw a vertical line $X = z$ and a horizontal line $Y = z$. Then the left side of $X = z$, and the area below $Y = z$ should constitute the shaded area $\mathcal{D}_z$.

(b)

$$F_Z(z) = P[\min(X,Y) \le z] = P[X \le z \text{ or } Y \le z] = F_X(z) + F_Y(z) - F_{XY}(z,z).$$

(c) If $X$ and $Y$ are independent, then

$$F_{XY}(z,z) = F_X(z)F_Y(z).$$

Therefore,

$$f_Z(z) = f_X(z) + f_Y(z) - f_X(z)F_Y(z) - F_X(z)f_Y(z).$$

(d)

$$f_X(z) = \lambda e^{-\lambda z}, \;\; f_Y(z) = \mu e^{-\mu z}.$$

Then, we find

$$f_Z(z) = (\lambda + \mu)e^{-(\lambda+\mu)z}.$$

**5.16* Maximum and minimum of two random variables.**

(a)

$$\mathcal{D}_{uv} = \{(x,y) : \min(x,y) \le u, \max(x,y) \le v\}.$$

For $u \le v$, the regions $\{x \le u\} \cap \{y \le v\}$ and $\{x \le v\} \cap \{y \le u\}$ constitute $\mathcal{D}_{uv}$. The region $\{x \le u\} \cap \{y \le u\}$ is included in both.
For $v < u$, the region $\{x \le v\} \cap \{y \le v\}$ defines $\mathcal{D}_{uv}$.

(b)

$$F_{UV}(u,v) = \begin{cases} F_{XY}(u,v) + F_{XY}(v,u) - F_{XY}(u,u) & v \ge u \\ F_{XY}(v,v) & v < u. \end{cases}$$

(c) The marginal distribution of $U$ is obtained by setting $v = \infty$ in the above equation for $v \geq u$:

$$F_U(u) = F_{UV}(u, \infty) = F_X(u) + F_Y(u) - F_{XY}(u, u).$$

The marginal distribution of $V$ is obtained by setting $u = \infty$ in the above expression for $v < u$:

$$F_V(v) = F_{UV}(\infty, v) = F_{XY}(v, v).$$

(d) We assume $a \leq b$. (The case $a > b$ can be treated in the same manner: we just exchange $a$ and $b$ in the final result.) The PDF $f_{XY}(x, y) = \frac{1}{ab},\;\; x \in [0, a] \cap y \in [0, b]$.
For $u \leq v$;

$$F_{UV}(u, v) = \begin{cases} 0, & u < 0 \\ \frac{2uv - u^2}{ab}, & 0 \leq u \leq v \leq a \\ \frac{ua + uv - u^2}{ab}, & 0 \leq u \leq a \leq v \\ \frac{v}{b}, & a \leq u \leq v \leq b \\ 1, & v > b. \end{cases}$$

For $v \leq u$

$$F(u, v) = \begin{cases} 0, & v < 0, \\ \frac{v^2}{ab}, & 0 \leq v \leq a, \\ \frac{v}{b}, & a < v, \\ 1, & v > b. \end{cases}$$

Thus, by combining the above results, we have

$$F_{UV}(u, v) = \begin{cases} 0, & \min(u, v)u < 0 \\ \frac{v^2}{ab}, & \{v < u\} \cap \{0 < v < a\} \\ \frac{u \min(a, v) + uv - u^2}{ab}, & \{v > u\} \cap \{0 \leq u \leq a\}, \\ \frac{v}{b}, & \{u > a\} \cap \{a < v < b\} \\ 1, & \{u > a\} \cap \{v > b\}. \end{cases}$$

**5.17 Area of a triangle.**

Consider the case $P_1 = (x_1, y_1) = (0, 0)$. Assume also $x_3 \geq x_2 > 0$ and $y_3 \geq y_2 > 0$ and $P_2$ lies below the line that connects $P_1$ and $P_3$. This layout corresponds to the where $P_1 - > P_2 - > P_3 - > P_1$ surrounds the triangle in a counter-clock wise. Then the area $A$ of the triangle can be represented as a rectangle minus three triangles:

$$A = x_3 y_3 - \frac{x_3 y_2}{2} - \frac{(x_3 - x_2) y_3}{2} - \frac{x_3 y_3}{2} = \frac{x_2 y_3 - x_3 y_2}{2} = \frac{1}{2} \det \begin{bmatrix} x_2 & x_3 \\ y_2 & y_3 \end{bmatrix}.$$

If $P_2$ lies above the line $P_1 P_2$, the above expression changes the sign, i.e.,

$$A = \frac{x_3 y_2 - x_2 y_3}{2} = -\frac{x_2 y_3 - x_3 y_2}{2}.$$

We now extend the above result to the case $P_1 \neq O$. By translating the coordinates by $x_1$ and $X-2$, we find

$$A = \frac{1}{2}\det\begin{bmatrix} x_2 - x_1 & x_3 - x_1 \\ y_2 - y_1 & y_3 - y_1 \end{bmatrix}.$$

Its equivalence to

$$A' = \frac{1}{2}\det\begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix},$$

can be proven by showing

$$A = A' = \frac{1}{2}(x_2y_3 - x_3y_2 + x_1y_2 - x_2y_1 + x_3y_1 - x_1y_3).$$

**5.18 Inverse of Jacobian matrix.** We are given

$$x = p(u,v),\ \ y = q(u,v),\ \ u = g(x,y),\ \ v = h(x,y).$$

Then

$$\boldsymbol{J}\left(\frac{p,q}{u,v}\right) = \begin{bmatrix} \frac{\partial p}{\partial u} & \frac{\partial p}{\partial v} \\ \frac{\partial q}{\partial u} & \frac{\partial q}{\partial v} \end{bmatrix},\ \ \boldsymbol{J}\left(\frac{g,h}{x,y}\right) = \begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{bmatrix}$$

Then by multiplying the two Jacobian matrices, we find

$$\boldsymbol{J}\left(\frac{p,q}{u,v}\right)\boldsymbol{J}\left(\frac{g,h}{x,y}\right) = \begin{bmatrix} \frac{\partial p}{\partial u}\frac{\partial g}{\partial x} + \frac{\partial p}{\partial v}\frac{\partial h}{\partial x}, & \frac{\partial p}{\partial u}\frac{\partial g}{\partial y} + \frac{\partial p}{\partial v}\frac{\partial h}{\partial y} \\ \frac{\partial q}{\partial u}\frac{\partial g}{\partial x} + \frac{\partial q}{\partial v}\frac{\partial h}{\partial x}, & \frac{\partial q}{\partial u}\frac{\partial g}{\partial x} + \frac{\partial q}{\partial v}\frac{\partial h}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

In order to obtain the last result, we observe

$$\begin{aligned} \frac{\partial p}{\partial u}\frac{\partial g}{\partial x} + \frac{\partial p}{\partial v}\frac{\partial h}{\partial x} &= \frac{\partial p}{\partial x} = \frac{\partial x}{\partial x} = 1, \\ \frac{\partial p}{\partial u}\frac{\partial g}{\partial y} + \frac{\partial p}{\partial v}\frac{\partial h}{\partial y} &= \frac{\partial p}{\partial y} = \frac{\partial x}{\partial y} = 0, \\ \frac{\partial q}{\partial u}\frac{\partial g}{\partial x} + \frac{\partial q}{\partial v}\frac{\partial h}{\partial x} &= \frac{\partial q}{\partial x} = \frac{\partial y}{\partial x} = 0, \\ \frac{\partial q}{\partial u}\frac{\partial g}{\partial x} + \frac{\partial q}{\partial v}\frac{\partial h}{\partial y} &= \frac{\partial q}{\partial y} = \frac{\partial y}{\partial y} = 1. \end{aligned}$$

## 5.3 Generation of Random Variates for Monte Carlo Simulation

**5.19* Use of a rejection method.**
Set $a = 0, b = 1, M = 2$ and $f_X(x) = 2x$ in Algorithm 5.1 of page 127. Then we obtain Algorithm 5.1 given below.

**5.20* Erlang variates.** From (5.75) we see

$$x_i = -\frac{\ln u_i}{k\mu}$$

**Algorithm 5.1 RNG Algorithm for $f_X(x) = 2x$**

1: Generate a uniform variate $u_1 \in [0,1]$, and set $x = u_1$.
2: Generate another uniform variate $u_2 \in [0,1]$.
3: If

$$2u_2 \leq 2x, \text{ i.e., } u_2 \leq x, \tag{2}$$

accept $x$, and reject otherwise.
4: Stop when the number of accepted variates $x$'s has reached a prescribed number. Otherwise, return to Step 1.

will be an exponential variate with mean $1/k\mu$. Thus,

$$x = \sum_{i=1}^{k} x_i = -\sum_{i=1}^{k} \frac{\ln u_i}{k\mu} = -\frac{\ln(\prod_{i=1}^{k} u_i)}{k\mu},$$

which is (5.82).

The algorithm is simply

1. Generate $k$ uniform variates $u_1, u_2, \ldots, u_k$.
2. Compute $x = -\frac{\ln(\prod_{i=1}^{k} u_i)}{k\mu}$.
3. Repeat the above until the desired number of Erlang variates $x$'s are generated.

**5.21 Poisson variates.**

Consider a Poisson process of rate $\lambda$, and observe $T$ time units. Then the number of arrivals in this period is Poisson distributed with mean $\lambda T$ (see p. 401, Theorem 14.1). The interarrival times of a Poisson process with rate $\lambda$ are exponentially distributed with mean $1/\lambda$ (see p. 403, Theorem 14.3).

Let $\tau_i$ be the interarrival time between $(i-1)$st and $i$th arrivals. Then consider an integer $X$ such that

$$\tau_1 + \tau_2 + \cdots + \tau_X \leq T < \tau_1 + \tau_2 + \cdots + \tau_X + \tau_{X+1}.$$

Then $X$ represents the number of Poisson arrivals in the interval $T$. As stated above, $X$ is Poisson distributed with mean $\lambda T$. So by setting $T = 1$, we find

$$\tau_1 + \tau_2 + \cdots + \tau_X \leq 1 < \tau_1 + \tau_2 + \cdots + \tau_X + \tau_{X+1}.$$

From (5.75), we can write the exponential random variates in terms of the uniform variates as

$$\tau_i = -\frac{\ln U_i}{\lambda}.$$

Hence the above inequalities can be written as

$$-\frac{\sum_{i=1}^{X} \ln U_i}{\lambda} \leq 1 < -\frac{\sum_{i=1}^{X+1} \ln U_i}{\lambda},$$

from which we obtain

$$\prod_{i=1}^{X} U_i \geq e^{-\lambda} > \prod_{i=1}^{X+1} U_i.$$

**5.22* The polar method for generating the Gaussian variate.** Let

$$X_1 = R\cos\Theta, \;\; X_2 = R\sin\Theta.$$

The joint PDF of $X_1, X_2$ is given as

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi}\exp\left\{-\frac{x_1^2 + x_2^2}{2}\right\}.$$

The PDF of $R, \Theta$ can be found as

$$f_{R\Theta}(r, \theta) = |J| f_{X_1X_2}(x_1, x_2),$$

where

$$J = \frac{\partial(x_1, x_2)}{\partial(r, \theta)} = \det\begin{bmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{bmatrix} = r.$$

Thus,

$$f_{R\Theta}(r, \theta) = \frac{r}{2\pi}\exp\left\{-\frac{r^2}{2}\right\}.$$

Since this joint PDF does not depend on $\theta$, the RVs $R$ and $\Theta$ are not only independent but also $\Theta$ is uniform. Thus the joint PDF is can be written as $f_\Theta(\theta)f_R(r)$, where

$$f_\Theta(\theta) = \frac{1}{2\pi}, \;\; 0 \leq \theta \leq 2\pi,$$

and

$$f_R(r) = r\exp\left\{-\frac{r^2}{2}\right\},$$

(a) By integrating the PDF obtained above

$$F_R(r) = \int_0^r f_R(s)\,ds = 1 - \exp\left\{-\frac{r^2}{2}\right\}.$$

The RV $\Theta$ is uniform in $[0, 2\pi]$ as obtained above.

(b) $R^2 = X_1^2 + X_2^2$ is exponentially distributed with mean 2. From (5.75) it then follows that $Y_1$ is uniformly distributed in $(0, 1)$. It is clear that since $\Theta$ is uniform in $[0, 2\pi]$, $Y_2$ is uniformly distributed in $(0, 1)$.

# 6 Solutions for Chapter 6: Fundamentals of Statistical Analysis

## 6.1 Sample Mean and Sample Variance

**6.1* Derivation of (6.11).**
Let $\overline{Y}$ denote the average of $Y_1, \ldots, Y_n$:

$$\overline{Y} \triangleq \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Note that

$$X_i - \overline{X} = (X_i - \mu) - (\overline{x} - \mu) = Y_i - \overline{Y}.$$

Then, the sample variance variable can be expanded as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right]. \tag{1}$$

By writing $\overline{Y}^2$ as

$$\overline{Y}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} Y_i Y_j = \frac{1}{n^2}\left[\sum_{i=1}^{n} Y_i^2 + \sum_{i=1}^{n}\sum_{j=1(j\neq i)}^{n} Y_i Y_j\right], \tag{2}$$

Then we can obtain (6.11)

**6.2 Recursive formula for sample mean and variance.**

(a) From

$$\overline{x}_i = \frac{1}{i}\sum_{j=1}^{i} x_j,$$

we can write

$$i\overline{x}_i = x_i + \sum_{j=1}^{i} x_j = x_i + (i-1)\overline{x}_{i-1} = i\overline{x}_{i-1} + x_i - \overline{x}_{i-1},$$

from which we find the recursive formula. The initial condition is obvious.

(b) Similarly,

$$
\begin{aligned}
(i-1)s_i^2 &= \sum j = 1^i [x_j - \overline{x}_{i-1} - (\overline{x}_i - \overline{x}_{i-1})]^2 \\
&= \sum_{j=1}^{i-1} [x_j - \overline{x}_{i-1} - \frac{1}{i}(x_i - \overline{x}_{i-1})]^2 + \left(1 - \frac{1}{i}\right)^2 (x_i - \overline{x}_{i-1})^2 \\
&= (i-2)s_{i-1}^2 + \frac{i-1}{i^2}(x_i - \overline{x}_{i-1})^2 + \frac{(i-1)^2}{i^2}(x_i - \overline{x}_{i-1})^2 \\
&= (i-2)s_{i-1}^2 + \frac{i-1}{i}(x_i - \overline{x}_{i-1})^2,
\end{aligned}
$$

from which we obtain the recursive formula for $s_i^2$. The initial conditions are again obvious.

## 6.2 Relative Frequency and Histograms

**6.3 Expectation and variance of the histogram.**

(a) Let $p_j = F_X(c_j) - F_X(c_{j-1}) \approx \Delta_j f_X(c_j)$. Then each sample falls in the $j$th class interval with probability $p_j$. Then $n_j(\boldsymbol{X})$ is binomially distributed with mean $E[n_j(\boldsymbol{X})] = np_j$ and variance $\mathrm{Var}[n_j(\boldsymbol{X}))] = np_j(1-p_j)$. Then

$$
E[h_j(\boldsymbol{X})] = \frac{E[n_j(\boldsymbol{X})]}{n\Delta_j} \approx \frac{p_j}{n\Delta_j} \approx f_X(c_j).
$$

(b) Similarly,

$$
\mathrm{Var}[h_j(\boldsymbol{X})] = \frac{\mathrm{Var}[n_j(\boldsymbol{X})}{n^2\Delta_j^2} = \frac{p_j(1-p_j)}{n\Delta_j^2} \approx \frac{f_X(c_j)}{n\Delta_j}.
$$

**6.4 Expectation and variance of the cumulative histogram.** Let $N_j(\boldsymbol{x})$ be the number of times that a sample value equal to or less than $c_j$ is observed, i.e.,

$$
N_j(\boldsymbol{x}) = n_1 + n_2 + \ldots, n_j.
$$

Then the cumulative histogram can be written as

$$
H_j(\boldsymbol{x}) = \frac{N_j(\boldsymbol{x})}{n} \leq 1.
$$

Note that both $H_j(\boldsymbol{x})$ and $N_j(\boldsymbol{x})$ depend on the observations $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, which is an instance of the RV $\boldsymbol{X}$. Then,

$$
E[N_j(\boldsymbol{X})] = nF_X(c_j), \text{ and } \mathrm{Var}[N_j(\boldsymbol{X})] = nF_X(c_j)(1 - F_X(c_j)).
$$

Then, the expectation and variance of $H_j(\boldsymbol{X})$ becomes

$$
E[H_j(\boldsymbol{X})] = F_X(c_j), \text{ and } \mathrm{Var}[H_j(\boldsymbol{X})] = \frac{1}{n}F_X(c_j)(1 - F_X(c_j)),
$$

which shows that the variance does not explicitly depend on the value of the interval size $\Delta_j$.

## 6.3 Graphical Presentations

**6.5 Log-survivor function curve of Erlang distributions.**

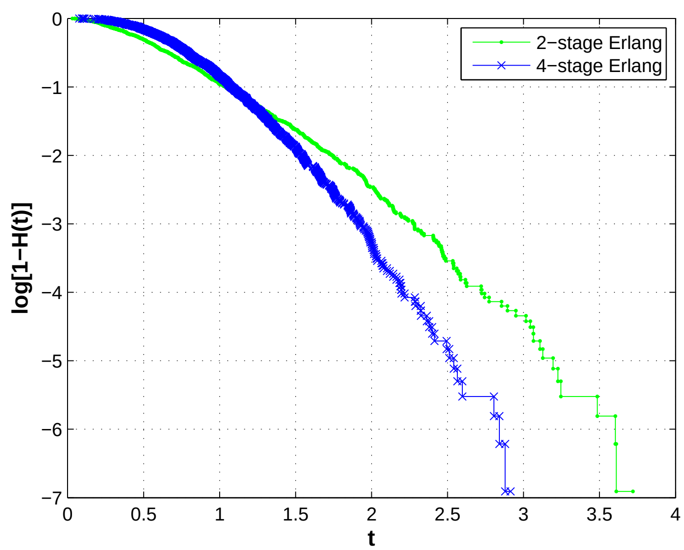



**Figure 6.1** Log-survivor function curve of 2-stage and 4-stage Erlang distributions (1000 samples).

In Figure 6.1, the log-survivor function curves for the 2-stage and 4-stage Erlang distributions with mean 1 are plotted by generating 1000 sample points for each curve. Each $k$-stage Erlang random variate $X$ is generated as a sum of $k$ random variates:

$$X = X_1 + X_2 + \cdots + X_k,$$

where $X_i$ $(i = 1, \ldots, k)$ is an exponentially distributed random variate with mean $1/k$.

**6.6* Log-survivor functions and hazard functions of a constant and uniform random variables.**

(a) For constant $X = a$, we have,

$$F_X(x) = u(x - a), \quad f_X(x) = \delta(x - a).$$

Hence

$$\log F_X^c(x) = \begin{cases} 0, & x < a \\ -\infty. & x \geq a. \end{cases}$$

and

$$h_X(x) == \begin{cases} 0, & x < a \\ \infty. & x \geq a. \end{cases}$$

(b) For the uniform distribution $X \in [a, b]$, we have

$$F_X(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \le x \le b, \\ 1, & x > b. \end{cases}$$

so,

$$f_X(x) = \begin{cases} 0, & x < a, \\ \frac{1}{b-a}, & a \le x \le b, \\ 0, & x > b. \end{cases}$$

Thus, the log-survivor function is

$$\log F_X^c(x) = \begin{cases} 0, & x < a, \\ \log \frac{b-x}{b-a}, & a \le x \le b, \\ -\infty, & x > b. \end{cases}$$

The hazard function is

$$h_X(x) = \frac{f_X(c)}{F_X^c(x)} = \begin{cases} 0, & x < a, \\ \log \frac{1}{b-x}, & a \le x \le b, \\ \infty, & x > b. \end{cases}$$

**6.7 Hazard function and distribution functions.** From (6.36),

$$h_X(t) = \frac{f_X(t)}{F_X^c(t)}.$$

Note that

$$\frac{d}{dt}[\ln F^c(t)] = \frac{F_X^{c\prime}(t)}{F^c(t)} = \frac{-f_X(t)}{F_X^c(t)} \overset{(a)}{=} -h_X(t).$$

Integrating both sides of equality $(a)$ from 0 to $x$, we obtain

$$\int_0^x \frac{d}{dt}[\ln F_X^c(t)]dt = -\int_0^x h_X(t)dt.$$

Hence,

$$\ln F_X^c(x) = -\int_0^x h_X(t)dt,$$

from which we obtain

$$F_X(x) = 1 - e^{-\int_0^x h_X(t)\, dt}.$$

By differentiating this, we obtain (6.55).

**6.8 Hazard function of a $k$-stage hyper-exponential distribution.**
We have

$$F_X(t) = 1 - \sum_{i=1}^{k} \pi_i e^{-\mu_i t}, \quad f_X(t) = \sum_{i=1}^{k} \pi_i \mu_i e^{-\mu_i t}.$$

Then

$$h_X(t) = \frac{\sum_{i=1}^k \pi_i \mu_i e^{-\mu_i t}}{\sum_{i=1}^k \pi_i e^{-\mu_i t}}.$$

By differentiating the above, we have

$$\frac{dh_X(t)}{dt} = \frac{-\left(\sum_i \pi_i \mu_i^2 e^{-\mu_i t}\right)\left(\sum_i e^{-\mu_i t}\right) + \left(\sum_i e^{-\mu_i t}\right)^2}{\left(\sum_i e^{-\mu_i t}\right)^2} \leq 0.$$

Thus, the function $h_X(x)$ is monotone decreasing.
The inequality in the above equation is based on the Cauchy-Schwartz inequality

$$\left(\sum_i x_i y_i\right)^2 \leq \left(\sum_i x_i^2\right)\left(\sum_i y_i\right)^2.$$

The equality holds if and only if

$$\sqrt{\pi_i}\mu_i e^{-\frac{\mu_i t}{2}} = \alpha\sqrt{\pi_i}e^{-\frac{\mu_i t}{2}}, \text{ for all } i \text{ and } t,$$

i.e., if and only if

$$\mu_i = \alpha, \text{ for all } i.$$

We readily find

$$h_X(0) = \sum_i \pi_i \mu_i, \text{ and } h_X(\infty) = \min\{\mu_i\}.$$

**6.9 Hazard function of the Pareto distribution.**
By substituting (4.84) and (4.85) into (6.36), we obtain

$$h_X(t) = \frac{f_X(t)}{F_X^c(t)} = \frac{\alpha}{t}, \ t \geq \beta.$$

**Note:** For the hazard function obtained above, we find from (6.38)

$$\begin{aligned} S_X(x) &= e^{-\alpha \int_\beta^x \frac{dt}{t}} = e^{-\alpha \ln x + \alpha \ln \beta} \\ &= \left(\frac{\beta}{x}\right)^\alpha, \ x > \beta. \end{aligned}$$

Then the mean residual life function of (6.41) is given as

$$R_X(t) = \frac{t}{\alpha - 1}, \ t > \beta.$$

So the solid curve given in Figure 6.5 corresponds to $\alpha = 2$ and $\beta = 1$. So $\alpha = 3$ in the figure caption is incorrect.

**6.10 Hazard function of the Weibull distribution.**

(a) From (6.37) it is apparent that for the standard Weibull distribution (see also Problem 4.20):

$$h_X(t) = \alpha t^{\alpha - 1}.$$

For $\alpha = 1$, $h_X(t) = 1$ and the Weibull distribution reduces to the exponential distribution with mean one.
For $\alpha = 2$, the hazard function becomes a linear function of $t$, i.e., $h_X(t) = 2t$.
**Remark:** For $\alpha = 2$, the (standard) Weibull distribution function and the PDF become

$$F_X(x) = 1 - e^{-x^2}, \text{ and } f_X(x) = 2xe^{-x^2}.$$

This distribution is equivalent to the Rayleigh distribution with $\sigma = \frac{1}{\sqrt{2}}$, which is discussed in Section 7.5.1

(b) Figure 4.5 (b) shows the hazard functions for $\alpha = 0.1, 0.5, 1, 2$ and 5 in the standard Weibull distribution.

**6.11* The mean residual life function and the hazard function**

The conditional survivor function can be written as

$$S_X(r|t) = \frac{S_X(r+t)}{S_X(t)} = \exp\left\{-\int_t^{t+r} h_X(u)\,du\right\}.$$

Then $S_X(r|t)$ is a monotone-increasing function of $t$ for all $r$, if and only if $h_X(t)$ is monotone non-decreasing; the inverse result holds if and only if $S_X(r|x)$ is non-decreasing. Since we can write

$$R_X(t) = E[R|X > t] = \int_0^\infty S_X(r|t)\,dr,$$

the stated property holds.

**6.12* Conditional survivor and mean residual life functions for standard Weibull distribution.**

(a)

$$\begin{aligned} S_X(r|t) = P[R > r|X > t] &= \frac{P[R > r, X > t]}{P[X > t]} = \frac{P[X > t + r]}{P[X > t]} \\ &= \frac{S_X(t+r)}{S_X(t)}, \end{aligned} \tag{3}$$

where $S_X(t) = e^{-t^\alpha}$ and $S_X(t+r) = e^{-(t+r)^\alpha}$ for the standard Weibull distribution. Thus,

$$S_X(r|t) = e^{-(t+r)^\alpha + t^\alpha}.$$

(b) Using the formula for the expectation of a nonnegative RV, we have

$$\begin{aligned} R_X(t) = E[R|X > t] &= \int_0^\infty P[R > r|X > t]\,dr = \int_0^\infty S_X(r|t)\,dr \\ &= \int_0^\infty \frac{S_X(t+r)}{S_X(t)}\,dr = \frac{\int_t^\infty S_X(u)\,du}{S_X(t)}, \end{aligned} \tag{4}$$

which is consistent with (6.41). Thus

$$R_X(t) = e^{t^\alpha} \int_t^\infty e^{-y^\alpha}\,dy.$$

Then by setting $y^\alpha = z$, or $y = z^{1/\alpha}$, we have

$$dy = \frac{1}{\alpha} z^{\frac{1}{\alpha} - 1}\,dz.$$

Thus,

$$R_X(t) = \frac{e^{t^\alpha}}{\alpha} \int_{t^\alpha}^{\infty} e^{-z} z^{\frac{1}{\alpha}-1}\, dz.$$

Then using the upper incomplete gamma function $\Gamma(\beta, x)$ defined by (4.34), we find

$$R_X(t) = \frac{e^{t^\alpha}}{\alpha} \Gamma\left(\frac{1}{\alpha}, t^\alpha\right).$$

For $t = 0$, we find

$$R_X(0) = \frac{1}{\alpha}\Gamma\left(\frac{1}{\alpha}, 0\right) = \frac{1}{\alpha}\Gamma\left(\frac{1}{\alpha}\right) = \Gamma\left(\frac{1}{\alpha} + 1\right),$$

where $\Gamma(x)$ is the gamma function defined by (4.31). This also agrees with (4.81) as expected, since $R_X(0) = E[X]$ as shown by (6.42).

**6.13 Mean residual life functions.**

(a)

$$F_X(x) = \pi_1 \left(1 - e^{-\alpha_1 x}\right) + \pi_2 \left(1 - e^{-\alpha_2 x}\right), \;\; \alpha_1 > \alpha_2.$$

Then

$$R_X(t) = \frac{\int_t^\infty \left(\pi_1 e^{-\alpha_1 u} + \pi_2 e^{-\alpha_2 u}\right)\, du}{\pi_1 e^{-\alpha_1 t} + \pi_2 e^{-\alpha_2 t}}.$$

For sufficiently large $t$, $\pi_1 e^{-\alpha_1 u} \ll \pi_2 e^{-\alpha_2 u}$ for $u > t$, since $\alpha_1 > \alpha_2$. Therefore,

$$\lim_{t\to\infty} R_X(t) = \lim_{t\to\infty} \frac{\int_t^\infty e^{-\alpha_2 u}\, du}{e^{-\alpha_2 t}}.$$

By applying l'Hôpital's rule, we have

$$\lim_{t\to\infty} R_X(t) = \lim_{t\to\infty} \frac{-e^{-\alpha_2 t}}{-\alpha_2 e^{-\alpha_2 t}} = \frac{1}{\alpha_2}.$$

(b) The CDF of the standard gamma distribution is

$$F_X(t;\beta) = 1 - \frac{\Gamma(\beta; t)}{\Gamma(\beta)},$$

where $\Gamma(t;\beta)$ is the upper incomplete gamma function defined by (7.126):

$$\Gamma(\beta, t) = \int_t^\infty u^{\beta-1} e^{-u}\, du.$$

Thus,

$$R_X(t) = \frac{\int_t^\infty \Gamma(\beta; u)\, du}{\Gamma(\beta; t)}. \tag{5}$$

By differentiating this we have

$$R'_X(t) = \frac{-\Gamma(\beta,t)\Gamma(\beta,t) - \int_t^\infty \Gamma(\beta,u)\,du\Gamma'(\beta,t)}{\Gamma^2(\beta,t)}$$
$$= -1 + \frac{t^{\beta-1}e^{-1}\int_t^\infty \Gamma(\beta,u)\,du}{\Gamma^2(\beta,t)}.$$

It does not seem so simple to determine when $R'_X(t)$ is positive or negative, so we will examine the hazard function instead. It is given by

$$h_X(t) = \frac{t^{\beta-1}e^{-t}}{\Gamma(\beta,t)}, \;\; t \geq 0, \;\; \beta > 0.$$

Thus,

- For $\beta > 1$, the hazard function $h_X(t)$ increases monotonically from $h_X(0) = 0$ to $h_X(\infty) = 1$ as $t$ goes from zero to $\infty$.
- For $\beta = 1$, $h_X(t) = 1$ for the standard gamma distribution (which becomes the exponential distribution with the parameter $\lambda = 1$).
- For $0 < \beta < 1$, the hazard function decreases monotonically from $h_X(0) = \infty$ to $h_X(\infty) = 1$. This property and the result of Problem 6.11 prove the property stated regarding $R_X(t)$. It is easy to calculate

  $$R_X(0) = E[X] = \beta, \;\text{ and }\; \lim_{t\to\infty} R_X(t) = 1,$$

  where the first result is from (6.42), whereas the second result is obtained by applying l'Hôpital's rule twice to (5).

**6.14 Mean residual life function – continued.**

(a) From (6.37), the hazard function for the Pareto distribution is

$$h_X(t) = \frac{\alpha}{t}, \;\; t \geq \beta.$$

Hence, the survivor function (6.38) becomes

$$S_X(x) = x^{-\alpha}\beta^\alpha, \;\; x > \beta.$$

and $S_X(x) = 1$ for $x < \beta$. Then by substituting the above into (6.41), we find:
For $t > \beta$

$$R_X(t) = \frac{\frac{\beta^\alpha}{\alpha-1}t^{-\alpha+1}}{\beta^\alpha t^{-\alpha}} = \frac{t}{\alpha-1}, \;\; t > \beta, \; \alpha > 1.$$

For $t < \beta$, we have, using $S_X(t) = 1$, that

$$R_X(t) = \int_t^\beta S_X(u)\,du + \int_\beta^\infty S_X(u)\,du = \beta - t + \frac{\beta}{\alpha-1} \tag{6}$$
$$= \frac{\alpha\beta}{\alpha-1} - t, \;\; t < \beta, \; \alpha > 1. \tag{7}$$

For $t = 0$, we have

$$R_X(0) = E[X] = \frac{\alpha\beta}{\alpha - 1}, \ \ \alpha > 1.$$

as expected from (4.87) and (6.42).

(b) For the Weibull distribution, we have from (6.37)

$$h_X(t) = \frac{\alpha}{\beta}\left(\frac{t}{\beta}\right)^{\alpha-1}, \ \ t \geq 0.$$

Then substituting this into (6.38), we find

$$S_X(x) = e^{-\frac{\alpha}{\beta}\int_0^x \left(\frac{t}{\beta}\right)^{\alpha-1} dt}.$$

Then it is not difficult to show

$$S_X(x) = e^{-\left(\frac{x}{\beta}\right)^{\alpha}}.$$

Thus, the mean residual life function of (6.41) takes the form

$$R_X(t) = \frac{\int_t^\infty e^{-\left(\frac{u}{\beta}\right)^{\alpha}} du}{e^{-\left(\frac{t}{\beta}\right)^{\alpha}}}.$$

The above can be represented in a close form if we use the upper incomplete gamma function defined by (4.34). This is left to the interested reader.

For $t = 0$, we have

$$R_X(0) = \int_0^\infty e^{-\left(\frac{u}{\beta}\right)^{\alpha}} du.$$

By setting

$$\left(\frac{u}{\beta}\right)^{\alpha} = y,$$

we find

$$R_X(0) = \frac{\beta}{\alpha}\int_0^\infty e^{-y} y^{\frac{1}{\alpha}-1} dy.$$

Recall the gamma function defined by (4.31):

$$\Gamma(p) = \int_0^\infty y^{p-1} e^{-y} dy.$$

Then

$$R_X(0) = \frac{\beta}{\alpha}\Gamma\left(\frac{1}{\alpha}\right) = \beta\Gamma\left(\frac{1}{\alpha} + 1\right),$$

where we used the identity (4.31), i.e.,

$$p\Gamma(p) = \Gamma(p+1).$$

The above result confirms that $R_X(0) = E[X]$. Recall that the mean of the two-parameter Weibull distribution is given by (4.168).

**6.15* Covariance between two random variables.** Since $X$ is uniformly distributed between $-\pi$ and $\pi$, $E[X] = 0$. Then

$$\text{Cov}[X, Y] = E\left[(X - E[X])(Y - E[Y])\right] = E[XY] - E[X]E[Y] = E[XY],$$

By substituting $Y = \cos X$, we have

$$E[XY] = \int_{-\pi}^{\pi} f_X(x) x \cos x \, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} x \cos x \, dx = 0.$$

Therefore, Cov $[X, Y] = 0$, hence $X$ and $Y$ are uncorrelated, but $X$ and $Y$ are not independent random variables.

**6.16 Correlation coefficient.**

(a) We write the normalized variables as $X^*$ and $Y^*$, respectively, and expand the right hand side (RHS), and take the expectation: Then

$$E[Z] = E\left[tX^* + Y^*\right]^2 = t^2 \text{Var}[X^*] + 2t\text{Cov}[X^*, Y^*] + \text{Var}[Y^*] = t^2 + 2\rho_{XY} t + 1.$$

(b) We write $E[Z] \triangleq f(t)$. $Z$ is nonnegative for any real number $t$, and so is the function $f(t) \geq 0$ for all real $t$. For any real-valued function $g(t) \triangleq at^2 + bt + c$, $g(t) \geq 0$ for all $t$, if and only if $a > 0$ and $D = b^2 - 4ac \leq 0$. For the problem at hand, this condition gives

$$D = 4\rho_{XY}^2 - 4 \leq 0.$$

Thus, $-1 \leq \rho_{XY} \leq 1$.

**6.17 Correlation coefficient-cont'd.** For $\rho_{XY} = 1$, then $f(t) = (t+1)^2$, thus $f(-1) = 0$. This means

$$E[-X^* + Y^*]^2 = \text{Var}[-X^* + Y^*] = 0.$$

Thus, with probability one $-X^* + Y^*$ assumes only one value, but this value must be zero, since $E[-X^* + Y^*] = E[Y^*] - E[X^*] = 0 - 0 = 0$. Thus,

$$P[-X^* + Y^* = 0] = 1.$$

Similarly for $\rho_{XY} = -1$, we have $f(t) = (t-1)^2$, which implies $E[X^* + Y^*]^2 = \text{Var}[X^* + Y^*] = 0$. Thus,

$$P[X^* + Y^* = 0] = 1.$$

By combining the above two, we have

$$P\left[\mp X^* + Y^* = 0\right] = 1.$$

**6.18* Sample covariance.**

$$\begin{aligned} s_{xy} &= \frac{1}{n-1}\sum_{i=1}^{n}\left[(x_i-\mu_X)-\frac{1}{n}\sum_{j=1}^{n}(x_j-\mu_X)\right]\left[(y_i-\mu_Y)-\frac{1}{n}\sum_{k=1}^{n}(y_k-\mu_X)\right] \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_X)(y_i-\mu_Y)-\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i-\mu_X)\sum_{j=1}^{n}(x_j-\mu_X) \\ &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_X)(y_i-\mu_Y)-\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}(x_i-\mu_X)(y_j-\mu_Y). \end{aligned}$$

Taking the expectations, we have

$$E[s_{xy}] = \sigma^2_{XY}.$$

**6.19 Recursive formula for sample covariance.** We denote by $s_{xy(i)}$ the sample covariance based on the first $i$ samples of $X$ and $Y$.

$$s_{xy}(i) = \frac{1}{i-1}\sum_{j=1}^{i}(x_j-\overline{x}_i)(y_j-\overline{y}_i),$$

where

$$\overline{x}_i = \frac{1}{i}\sum_{j=1}^{i}x_j = \overline{x}_{i-1}+\frac{1}{i}(x_i-\overline{x}_{i-1}).$$

By proceedings in the same manner as in Problem 6.2, we find

$$s_{xy(i)} = \frac{i-2}{i-1}s_{xy(i-1)}+\frac{1}{i}(x_i-\overline{x}_{i-1})(y_i-\overline{y}_{i-1}).$$

# 7 Solutions for Chapter 7: Fundamentals of Statistical Analysis

## 7.1 Chi-Squared Distribution

**7.1* Sample variance and chi-squared variable.**
We write

$$u_i = \frac{X_i - \overline{X}}{\sigma}, \quad i = 1, 2, \ldots, n.$$

Then

$$\chi^2 = \sum_{i=1}^{n} u_i^2, \quad \text{and} \quad \sum_{i=1}^{n} u_i = 0.$$

We use the last equation to eliminate $u_n$ from the expression for $\chi^2$:

$$\begin{aligned}
u_n &= -(u_1 + u_2 + \cdots + u_{n-1}),\\
u_n^2 &= u_1^2 + 2(u_1 u_2 + u_1 u_3 + \cdots + u_1 u_{n-1})\\
&\quad + u_2^2 + 2(u_2 u_3 + \cdots + u_2 u_{n-1})\\
&\qquad \vdots\\
&\qquad + u_{n-1}^2.
\end{aligned}$$

Then we can write $\chi^2$ as

$$\begin{aligned}
\frac{\chi^2}{2} &= u_1^2 + u_1u_2 + u_1u_3 + \cdots + u_1u_{n-1} \\
&\quad + u_2^2 + u_2u_3 + \cdots + u_2u_{n-1} \\
&\quad \vdots \\
&\quad + u_{n-1}^2 \\
&= \left[u_1 + \frac{1}{2}(u_2 + u_3 + \cdots + u_{n-1})\right]^2 \\
&\quad \frac{3}{4}u_2^2 + \frac{1}{2}(u_2u_3 + \cdots + u_2u_{n-1}) \\
&\quad \frac{3}{4}u_3^2 + \frac{1}{2}(u_3u_4 + \cdots + u_3u_{n-1}) \\
&\quad \vdots \\
&\quad + \frac{3}{4}u_{n-1}^2.
\end{aligned}$$

By writing

$$\begin{aligned}
\hat{u}_1 &= \sqrt{2}\left[u_1 + \frac{1}{2}(u_2 + u_3 + \cdots + u_{n-1})\right] \\
\hat{u}_2 &= \sqrt{\frac{3}{2}}\left[u_2 + \frac{1}{3}(u_3 + \cdots + u_{n-1})\right] \\
&\vdots \\
\hat{u}_i &= \sqrt{\frac{i+1}{i}}\left[u_i + \frac{1}{i+1}(u_{i+1} + \cdots + u_{n-1})\right] \\
&\vdots \\
\hat{u}_{n-1} &= \sqrt{\frac{n}{n-1}}u_{n-1},
\end{aligned}$$

We can write

$$\chi^2 = \sum_{i=1}^{n-1} \hat{u}_i^2.$$

Since $\hat{u}_i$s are linear functions of the normally distributed RVs, the distribution of $(\hat{u}_1, \ldots, \hat{u}_{n-1})$ is an $(n-1)$-dimensional normal distribution. In order to prove that $\hat{u}_i$s are statistically independent, we need only to prove that the covariances are zero.

We write

$$u_i = \frac{X_i - \overline{X}}{\sigma} = \frac{X_i - \mu}{\sigma} - \frac{\overline{X} - \mu}{\sigma} = U_i - \overline{U},$$

where $U_i = \frac{X_i - \mu}{\sigma}$ as in (7.20) and

$$\overline{U} = \frac{1}{n}\sum_{i=1}^{n} U_i.$$

Note $\hat{U}$ is different from $U$ of (7.30) by a factor $\frac{1}{\sqrt{n}}$. Then we have

$$\begin{aligned}
\hat{u}_i &= \frac{1}{\sqrt{i(i+1)}}\left[(i+1)u_i + u_{i+1} + \cdots + u_{n-1}\right] \\
&= \frac{1}{\sqrt{i(i+1)}}\left[(i+1)U_i + U_{i+1} + \cdots + U_{n-1} - n\overline{U}\right] \\
&= \frac{-1}{\sqrt{i(i+1)}}\left[U_1 + U_2 + \cdots + U_{i1} + U_n - iU_i\right].
\end{aligned}$$

Hence

$$E[\hat{u}_i] = 0,$$

and

$$\mathrm{Var}[\hat{u}_i] = \frac{1}{i+1}[1 + 1 + \cdots + 1 + 1 + i^2] = \frac{i + i^2}{i(i+1)} = 1.$$

The product

$$\hat{u}_i\hat{u}_j = \frac{1}{\sqrt{i(i+1)j(j+1)}}[U_1 + U_2 \cdots + U_{i1-} + U_n - iU_i][U_1 + U_2 + \cdots + U_{j-1} + U_n - jU_j]$$

shows that for $i < j$,

$$E[\hat{u}_i\hat{u}_j] = \frac{1}{\sqrt{i(i+1)j(j+1)}}E[U_1^2 + U_2^2 + \cdots + U + i - 1^2 + U_n^2 - iU_i^2],$$

since $E[U_rU_s] = 0$ for $r \neq s$. It is easy to see

$$\mathrm{Var}[\hat{u}_i\hat{u}_j] = E[\hat{u}_i\hat{u}_j] = \frac{1}{\sqrt{i(i+1)j(j+1)}}[1 + 1 + \cdots + 1 + 1 - i] = 0.$$

Thus, the $(n-1)$-variables $\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_{n-1}$ are statistically independent standard normal variables.

**7.2 Derivation of the $\chi_n^2$ distribution.**

(a) The joint density of $\boldsymbol{U} = (U_1, U_2, \ldots, U_n)$ of (7.20) has

$$\begin{aligned}
f_{\boldsymbol{U}}(\boldsymbol{u}) &= \frac{1}{(\sqrt{2\pi})^n}\exp\left[-\frac{1}{2}(U_1^2 + U_2^2 + \ldots + U_n^2)\right] \\
&= \frac{1}{(\sqrt{2\pi})^n}\exp\left(-\frac{\chi_n^2}{2}\right).
\end{aligned}$$

Hence,

$$f_{\chi_n}(\chi)d\chi = P[\chi < \chi_n < \chi + d\chi] = \frac{e^{-\frac{\chi^2}{2}}}{(\sqrt{2\pi})^n}d\chi.$$

(b)

$$
\begin{aligned}
f_{\chi_n^2}(\nu)d\nu &= P[\nu < \chi_n^2 < \nu + d\nu] = P[\sqrt{\nu} < \chi_n < \sqrt{\nu} + \frac{1}{s\sqrt{\nu}} d\nu] \\
&= \frac{A}{(2\pi)^{n/2}} \nu^{\frac{n-1}{2}} e^{-\frac{\nu}{2}} \frac{1}{2\sqrt{\nu}} d\nu \\
&= \frac{A}{2(2\pi)^{n/2}\sqrt{\nu}} \nu^{\frac{n-1}{2}} e^{-\frac{\nu}{2}} d\nu.
\end{aligned}
$$

(c)

$$
\begin{aligned}
1 = \int_0^\infty f_{\chi_n^2}(\nu)d\nu &= \frac{A}{2(2\pi)^{n/2}} \int_0^\infty \nu^{\frac{n}{2}-1} e^{-\frac{\nu}{2}} d\nu \\
&= \frac{A}{(2\pi)^{n/2}} \int_0^\infty (2t)^{\frac{n}{2}-1} e^{-t} dt \quad \text{(by setting } \frac{\nu}{2} = t) \\
&= \frac{2^{\frac{n}{2}-1} A}{(2\pi)^{n/2}} \int_0^\infty t^{\frac{n}{2}-1} e^{-t} dt = \frac{2^{\frac{n}{2}} A}{2(2\pi)^{n/2}} \Gamma\left(\frac{n}{2}\right),
\end{aligned}
$$

from which we can determine $A$.

**7.3* Moments of gamma and $\chi^2$-distributions.**

(a)

$$
E[X^m] = \int_0^\infty x^m f_X(x)dx = \frac{1}{\Gamma(\beta)} \int_0^\infty x^{m+\beta-1} e^{-x}\, dx = \frac{\Gamma(m+\beta)}{\Gamma(\beta)}.
$$

(b)

$$
f_{\chi_n^2}(\nu)d\nu = \frac{\nu^{\frac{n}{2}-1} e^{-\frac{\nu}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}, \;\; 0 \le \nu < \infty.
$$

$$
\begin{aligned}
E[(\chi_n^2)^m] &= \int_0^\infty \nu^m f_{\chi_n^2}(\nu)d\nu \\
&= \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty \nu^{m+\frac{n}{2}-1} e^{-\nu/2}\, d\nu \\
&= \frac{2^{m+\frac{n}{2}}}{2^{n/2}\Gamma(n/2)} \int_0^\infty t^{m+\frac{n}{2}-1} e^{-t}\, dt \\
&= \frac{2^m \Gamma\left(\frac{n}{2}+m\right)}{\Gamma(n/2)}.
\end{aligned}
$$

**7.4 $\chi^2$ distribution and exponential distribution.**

Transform the pair $(X_1, X_2)$ into the polar coordinates $(R, \Theta)$ according to

$$
X_1 = R\cos\Theta
$$

and

$$
X_2 = R\sin\Theta.
$$

Then from the result of Problem 5.22, the probability distribution function of $R$ is

$$F_R(r) = 1 - e^{-\frac{r^2}{2}}.$$

Hence the distribution of

$$Y = X_1^2 + X_2^2 = R^2$$

is given by

$$F_Y(y) = P[Y \leq y] = P[R \leq \sqrt{y}] = 1 - e^{-y/2}.$$

Hence,$Y$ is exponentially distributed with mean two.

**7.5 Non-central chi-squared distribution.**

(a) There are several equivalent formulas for the distribution function of $\chi_n^2(\mu^2)$: one uses a modified Bessel function of the first kind[1];
Another formula represents the PDF $f_{\chi_n^2(\mu^2)}(x)$ in terms of the following weighted sum of the PDFs of the regular chi-squared distributions $f_{\chi_{n+2j}^2}(x),\ j = 0, 1, 2, \ldots$, with the weights being equal to the Poisson distribution with parameter $\mu^2/2$:

$$f_{\chi_n^2(\mu^2)}(x) = e^{-\mu^2/2} \sum_{j=0}^{\infty} \frac{(\mu^2/2)^j}{j!} f_{\chi_{n+2j}^2}(x). \tag{1}$$

(b) For $n = 2$, the above reduces to

$$f_{\chi_2^2(\mu^2)}(x) = e^{-\mu^2/2} \sum_{j=0}^{\infty} \frac{(\mu^2/2)^j}{j!} f_{\chi_{2+2j}^2}(x). \tag{2}$$

(c) Recall that the distribution function of this chi-squared distribution with $2(j+1)$ degrees of freedom is found from the the $r$-stage Erlang distribution (4.165), by setting $r = j + 1$, and $\lambda = \frac{1}{2(j+1)}$, hence $\lambda r = \frac{1}{2}$. Thus

$$F_{\chi_{2+2j}^2}(y) = 1 - e^{-\frac{y}{2}} \sum_{i=0}^{j} \frac{\left(\frac{y}{2}\right)^i}{i!}, \ \ y \geq 0. \tag{3}$$

Thus the PDF is

$$f_{\chi_{2+2j}^2}(y) = \frac{1}{2} e^{-\frac{y}{2}} \frac{\left(\frac{y}{2}\right)^j}{j!}. \tag{4}$$

[1] See Section 7.5.2 on Rice distribution, which is a special case of the non-central chi-squared distribution.

Then

$$f_{\chi_2^2(\mu^2)}(y) = \frac{1}{2}e^{-\frac{\mu^2+y}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\mu^2 y}{4}\right)^j}{(j!)^2}$$
$$= \frac{1}{2}e^{-\frac{\mu^2+y}{2}} \sum_{j=0}^{\infty} \left[\frac{\left(\frac{\mu\sqrt{y}}{2}\right)^j}{j!}\right]^2$$
$$= \frac{1}{2}e^{-\frac{\mu^2+y}{2}} I_0(\mu\sqrt{y}),$$

which is the corrected (7.112), where $I_0(x)$ is the modified Bessel function of the first kind with zero degrees of freedom, and is defined by (7.83).

## 7.2 Student's $t$-Distribution

**7.6 Derivation of the $t$-distribution.**

(a) For an observation $X_i$, its deviation from the population mean $\mu$ is given as

$$X_i - \mu = (X_i - \overline{X}) + (\overline{X} - \mu).$$

Then $X_i - \overline{X}$ and $\overline{X} - \mu$ are orthogonal in the sense

$$E[(X_i - \overline{X})(\overline{X} - \mu)] = 0.$$

Hence, the normal variable $\overline{X} - \mu$ is orthogonal and hence independent of $X_i - \overline{X}$ for $i = 1, 2, \ldots, n$. Therefore, the variables

$$U = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$$

and

$$\chi^2_{n-1} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma}$$

are statistically independent:

$$f_{U,\chi^2_{n-1}}(u, \nu)dud\nu = f_U(u)du f_{\chi^2_{n-1}}(\nu)d\nu.$$

Since

$$f_U(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2}$$

and

$$f_{\chi^2_{n-1}}(\nu) = \frac{\nu^{\frac{n-1}{2}-1}e^{-\nu/2}}{2^{\frac{n-1}{n}}\Gamma\left(\frac{n-1}{2}\right)},$$

we obtain the desired result.

(b)

$$f_{t_{n-1}\chi^2_{n-1}}(t,\nu) = |J| f_{U\chi^2_{n-1}}(u,\nu),$$

where $J$ is Jacobian

$$J = \frac{\partial(U,\chi^2_{n-1})}{\partial(t_{n-1}\chi^2_{n-1})} = \sqrt{\frac{\chi^2_{n-1}}{n-1}}.$$

Hence,

$$f_{t_{n-1}\chi^2_{n-1}}(t,\nu) = \sqrt{\frac{\nu}{n-1}} f_{U\chi^2_{n-1}}\left(t\sqrt{\frac{\nu}{n-1}},\nu\right),$$

from which we obtain the desired result.

(c)

$$\begin{aligned}
f_{t_{n-1}}(t) &= \int_0^\infty f_{t_{n-1}\chi^2_{n-1}}(t,\nu)d\nu \\
&= \frac{1}{2\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)} \int_0^\infty \left(\frac{\nu}{2}\right)^{\frac{n-2}{2}} e^{-\frac{\nu}{2}\left(1+\frac{t^2}{n-1}\right)d\nu} \\
&= \frac{J}{2\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)}.
\end{aligned}$$

By transforming

$$z = \frac{\nu}{2}\left(1+\frac{t^2}{n-1}\right),$$

the integral becomes

$$\begin{aligned}
I &= \frac{2}{\left(1+\frac{t^2}{n-1}\right)^{n/2}} \int_0^\infty z^{\frac{n}{2}-1} e^{-z} dz \\
&= 2\left(1+\frac{t^2}{n-1}\right)^{-\frac{n}{2}} \Gamma\left(\frac{n}{2}\right).
\end{aligned}$$

Hence, we obtain $f_{t_{n-1}}(t)$ as given in (7.33).

## 7.3 Fisher's $F$-Distribution

**7.7* Moments of the $F$-distribution.**
From (7.39), the $r$th moment of $F$ is

$$E[F^r] = \left(\frac{n_2}{n_1}\right)^r E[V_1^r] E[V_2^{-r}].$$

From the result of Problem 7.3 (b)

$$E[V_1^r] = \frac{2^r \Gamma\left(\frac{n_1}{2}+r\right)}{\Gamma\left(\frac{n_1}{2}\right)}$$

and

$$E[V_2^{-r}] = \frac{2^{-r}\Gamma\left(\frac{n_2}{2} - r\right)}{\Gamma\left(\frac{n_2}{2}\right)}.$$

Hence,

$$E[F^r] = \left(\frac{n_2}{n_1}\right)^r \frac{\Gamma\left(\frac{n_1}{2} + r\right)\Gamma\left(\frac{n_2}{2} - r\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}.$$

From the conditions $\frac{n_1}{2} + r > 0$ and $\frac{n_2}{2} - r > 0$, we obtain $-n_1 < 2r < n_2$.

**7.8 The $F$-distribution and the $t$-distribution.** When $n_1 = 1, V_1 = U^2$, where $U$ is a standard normal variable. Then

$$F_{1,n_2} = \frac{U^2}{V_2/n_2} = \left(\frac{U}{\sqrt{\chi^2_{n_2}/n_2}}\right)^2 = t^2_{n_2}.$$

Therefore, the $F$ variable is simply the square of the $t$ variable of degree $n_2$. We can then derive the distribution of $F_{1,n_2}$ from that of $t_{n_2}$.

$$\begin{aligned} f_F(x)\,dx &= P[x < F < x + dx] \\ &= P\left[\sqrt{x} < t < \sqrt{x} + \frac{dx}{2\sqrt{x}}\right] + P\left[-\sqrt{x} > t > -\sqrt{x} - \frac{dx}{2\sqrt{x}}\right] \\ &= \frac{f_t(\sqrt{x}) + f_t(-\sqrt{x})}{2\sqrt{x}}\,dx \\ &= \frac{\Gamma\left(\frac{n_2+1}{2}\right)}{\Gamma\left(\frac{n_2}{2}\right)\sqrt{\pi n_2}}\left(1 + \frac{x}{n_2}\right)^{-\frac{n_2+1}{2}} \frac{dx}{\sqrt{x}}. \end{aligned}$$

It is not difficult to see that this is a special case of (7.40).

## 7.4 Lognormal Distribution

**7.9* Median and mode of the lognormal distribution.**

(a) The median of the distribution (7.46) is $y_{\mathrm{med}} = \mu_Y$. The corresponding $x_{\mathrm{med}}$ is

$$x_{\mathrm{med}} = e^{y_{\mathrm{med}}} = e^{\mu_Y}.$$

Using (7.52), we find

$$\begin{aligned} x_{\mathrm{med}} = e^{\mu_Y} &= e^{\ln\mu_X - \frac{1}{2}\ln\left(1 + \frac{\sigma_Y^2}{\mu_X^2}\right)} \\ &= \mu_X\left(1 + \frac{\sigma_Y^2}{\mu_X^2}\right)^{-\frac{1}{2}} = \frac{\mu_X}{\sqrt{\left(1 + \frac{\sigma_Y^2}{\mu_X^2}\right)}}. \end{aligned}$$

(b) Take the logarithm of (7.47):

$$\ln f_X(x) = -\frac{1}{2}\ln(2\pi) - \ln x - \frac{(\ln x - \mu_Y)^2}{2\sigma_Y^2}.$$

Differentiate the above with respect to $x$:

$$\frac{f'_X(x)}{f_X(x)} = -\frac{1}{x} - \frac{\ln x - \mu_X}{\sigma_Y^2 x}.$$

The mode $x_{\text{mode}}$ is such $x$ that maximizes $f_X(x)$, and thus $f'_X(x_{\text{mode}}) = 0$. Thus,

$$-1 - \frac{\ln x_{\text{mode}} - \mu_Y}{\sigma_Y^2} = 0,$$

from which we have

$$x_{\text{mode}} = e^{\mu_Y - \sigma_Y^2}.$$

By substituting the result of part (a) and (7.53), we obtain

$$x_{\text{mode}} = \frac{\mu_X}{\left(1 + \frac{\sigma_Y^2}{\mu_X^2}\right)^{\frac{3}{2}}}.$$

## 7.5 Rayleigh and Rice Distributions

**7.10* MGF of $R^2$ and $R$ variables in the Rayleigh distribution.**

(a)

$$M_Z(t) = E[e^{t(X^2+Y^2)}] = M_{X^2}(t)M_{Y^2}(t).$$

Let $X = \sigma U$ and $Y = \sigma V$, then $U$ and $V$ are independent unit normal variables. Then

$$\begin{aligned} M_{X^2}(t) = E[e^{t\sigma^2 U^2}] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t\sigma^2 u^2} e^{-\frac{u^2}{2}}\, du \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(u\sqrt{1-2\sigma^2 t})^2}{2}}\, du. \end{aligned}$$

By setting $u\sqrt{1-2\sigma^2 t} = w$, we have

$$M_{X^2}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{w^2}{2}} \frac{dw}{\sqrt{1-2\sigma^2 t}} = \frac{1}{\sqrt{1-2\sigma^2 t}}.$$

Since $M_{Y^2}(t)$ is the same as $M_{X^2}(t)$, we have

$$M_Z(t) = \frac{1}{1-2\sigma^2 t},$$

which leads to

$$m_Z(t) = -\ln(1-2\sigma^2 t).$$

Thus,

$$m'_Z(t) = \frac{2\sigma^2}{1-2\sigma^2 t}, \;\; m''_Z(t) = \frac{(2\sigma^2)^2}{(1-2\sigma^2 t)^2}.$$

Hence

$$E[Z] = m'_Z(0) = 2\sigma^2, \ \ \mathrm{Var}[Z] = 4\sigma^4.$$

(b)

$$M_R(t) = E\left[e^{t\sqrt{X^2+Y^2}}\right] = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{u^2+v^2}{2}} e^{t\sigma\sqrt{u^2+v^2}}\,du\,dv.$$

By writing

$$u = r\cos\theta, \ \ v = r\sin\theta, \ \text{ hence } du\,dv = rdr\,d\theta,$$

we have

$$M_R(t) = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{-\frac{r^2}{2}} e^{t\sigma r} r\,dr\,d\theta.$$

By setting $r - \sigma t = y$, i.e., $r = y + \sigma t$, we can write

$$\begin{aligned} M_R(t) &= \left[\int_{-\sigma t}^{\infty} (y+\sigma t) e^{-\frac{y^2}{2}}\,dy\right] e^{\frac{\sigma^2 t^2}{2}} \\ &= \left\{\left[-e^{-\frac{y^2}{2}}\right]_{-\sigma t}^{\infty} + \sqrt{2\pi}\sigma t\left[1-\Phi(-\sigma t)\right]\right\} e^{\frac{\sigma^2 t^2}{2}} \\ &= 1 + \sqrt{2\pi}\sigma t \Phi(\sigma t) e^{\frac{\sigma^2 t^2}{2}}, \end{aligned}$$

where $\Phi(x)$ is the distribution function of the unit normal variable, and we use the property $1 - \Phi(-x) = \Phi(x)$.

(c) To simplify the notation we define

$$\Phi \triangleq \Phi(\sigma t), \ \text{ and } \ F \triangleq \Phi e^{\frac{\sigma^2 t^2}{2}}.$$

So

$$\ln F = \ln \Phi + \frac{\sigma^2 t^2}{2}.$$

By differentiating the above with respect to $t$, we have

$$\frac{F'}{F} = \frac{\sigma\phi}{\Phi} + \sigma^2 t,$$

where $\phi \triangleq \phi(\sigma t)$, the PDF of the unit normal variable. By differentiating the above once again

$$\frac{F''F - (F')^2}{F^2} = \frac{\sigma^2(\phi'\Phi - \phi^2)}{\Phi^2} + \sigma^2,$$

where $\phi' \triangleq \phi'(\sigma t)$. By setting $t = 0$ in the functions,

$$F(0) = \Phi(0) = \frac{1}{2}, \ \text{ and } \frac{F'(0)}{F(0)} = \frac{\sigma\phi(0)}{\Phi(0)} = \sqrt{\frac{2}{\pi}}\sigma,$$

we find

$$M'_R(0) = \sqrt{2\pi}\sigma F(0) = \sqrt{\frac{\pi}{2}}\sigma, \ \text{ and } M''_R(0) = \sqrt{2\pi}\sigma(2F'(0)) = \sqrt{2\pi}\sigma\frac{2\sigma}{\sqrt{2\pi}} = 2\sigma^2.$$

Thus, we find the variance of $R$ as given above.

**7.11 Alternative derivation of the Rice distribution.** Set

$$X = \mu_X + \sigma U_1, \text{ and } Y = \mu_Y + \sigma U_2,$$

where $U_1, U_2$ are independent unit normal RVs. Then

$$R^2 = X^2 + Y^2 = \sigma^2 \left[ \left( U_1 + \frac{\mu_X}{\sigma} \right)^2 + \left( U_2 + \frac{\mu_Y}{\sigma} \right)^2 \right].$$

By setting $\mu_X^2 + \mu_Y^2 = \mu^2$, we have

$$R^2 = \sigma^2 \chi_2^2 \left( \frac{\mu^2}{\sigma^2} \right),$$

where $\chi_2^2\left(\frac{\mu^2}{\sigma^2}\right)$ is the noncentral chi-square variable with 2 degrees of freedom and noncentrality parameter $\mu^2/\sigma^2$. For notational convenience we denote this RV by $Y$. Then we know from (7.112)

$$f_Y(y) = \frac{1}{2} e^{-\frac{\left(\frac{\mu}{\sigma}\right)^2 + y}{2}} I_0 \left( \frac{\mu}{\sigma} \sqrt{y} \right).$$

Then we want to derive the PDF of the RV $R$, where

$$R^2 = \sigma^2 Y, \text{ that is, } R = \sigma \sqrt{Y}.$$

Setting

$$f_R(r) dr = f_Y(y) dy, \text{ and } r = \sigma \sqrt{y},$$

We find

$$\begin{aligned} f_R(r) &= f_Y(y) \frac{dy}{dr} = f_Y \left( \frac{r^2}{\sigma^2} \right) \frac{2r}{\sigma^2} \\ &= \frac{r}{\sigma^2} e^{-\frac{\mu^2 + r^2}{2\sigma^2}} I_0 \left( \frac{\mu r}{\sigma^2} \right). \end{aligned}$$

**7.12 Rotation of the Rice distribution.**
Consider the $X - Y$ coordinates and the point $(\mu_X, \mu_Y)$, and $\tan \psi \triangleq \frac{\mu_Y}{\mu_X}$. The result will be obvious once we recognize that the joint distribution is a circular bivariate normal distribution around the mean $(\mu_X, \mu_Y)$.
In order to prove algebraically, we note that the coordinates $(X', Y')$ and $(X, Y)$ are related by

$$\begin{aligned} X &= X' \cos \psi - Y' \sin \psi \\ Y &= X' \sin \psi + Y' \cos \psi. \end{aligned}$$

Then

$$
\begin{aligned}
&(X-\mu_X)^2+(Y-\mu_Y)^2\\
&=(X'\cos\psi-Y'\sin\psi-\mu_X)^2+(X'\sin\psi+Y'\cos\psi-\mu_Y)^2\\
&=X'^2\cos^2\psi+Y'^2\sin^2\psi+\mu_X^2-2X'Y'\sin\psi\cos\psi+2\mu_XY'\sin\psi-2\mu_XX'\cos\psi\\
&\quad+X'^2\sin^2\psi+Y'^2\cos^2\psi+\mu_Y^2+2X'Y'\sin\psi\cos\psi-2\mu_YX'\sin\psi-2\mu_YY'\cos\psi\\
&=X'^2+Y'^2+(\mu_X^2+\mu_Y^2)-2(\mu_X\cos\psi+\mu_Y\sin\psi)X'+2(\mu_XY'\sin\psi-\mu_YY'\cos\psi)\\
&=X'^2+Y'^2+\mu^2-2\mu X'+0=(X'-\mu)^2+Y'^2.
\end{aligned}
$$

Hence, by noting the Jacobian of the above linear transformation is one, we have

$$
\begin{aligned}
f'_XY'(x',y')=|J|f_XY(x,y)=|J|\frac{1}{2\pi\sigma^2}\exp\left[-\frac{(x-\mu_X)^2+(y-\mu_Y)^2}{2\sigma^2}\right]\\
=\frac{1}{2\pi\sigma^2}\exp\left[-\frac{(x'-\mu)^2+y'^2}{2\sigma^2}\right]
\end{aligned}
$$

**7.13* Nakagami distribution.**

(a)

$$E[Z]=2m\sigma^2\triangleq\Omega, \tag{5}$$

Then by writing $X_i=\sigma U_i$, we readily find

$$Z=\sigma^2\chi^2_{2m}, \tag{6}$$

where $\chi^2_{2m}$ is the chi-squared variable with $2m$ degrees of freedom[2]. Thus, we find the PDF of $Z$ as

$$f_Z(z)=\frac{1}{\sigma^2}f_{\chi^2_{2m}}\left(\frac{z}{\sigma^2}\right)=\frac{1}{\sigma^2}\frac{\left(\frac{z}{\sigma^2}\right)^{m-1}e^{-\frac{z}{2\sigma^2}}}{2^m\Gamma(m)},\ \ z\geq 0, \tag{7}$$

where $\Gamma(m)$ is the gamma function defined in (4.31), and $\Gamma(m)=(m-1)!$ when $m$ is an integer. By substituting (5), we have

$$f_Z(z)=\frac{m^m}{\Omega^m\Gamma(m)}z^{m-1}e^{-\frac{mz}{\Omega}},\ \ z\geq 0. \tag{8}$$

(b) define a random variable $R$, or the *envelope* of the The PDF of $R$ is obtained by setting $f_R(r)\,dr=f_Z(z)\,dz$. This leads to

$$f_R(r)=\frac{2m^m}{\Omega^m\Gamma(m)}r^{2m-1}e^{-\frac{mr^2}{\Omega}},\ \ r\geq 0. \tag{9}$$

[2] Recall that $Y_{2m}\triangleq\chi^2_{2m}/2$ is an $\mathrm{E}_m$ variable, i.e., is Erlangian distributed with mean $m$ (cf. (7.16)):

$$f_{Y_{2m}}(y)=\frac{y^{m-1}e^{-y}}{(m-1)!},$$

which can be seen as the gamma distribution with $\lambda=1$ and $\beta=m$ (cf. (4.30)).

An alternative expression is given in terms of $\sigma^2$ as

$$f_R(r) = \frac{2\left(\frac{m}{2\sigma^2}\right)^m}{\Gamma(m)} r^{2m-1} e^{-\frac{mr^2}{2\sigma^2}}, \quad r \geq 0, \tag{10}$$

(c)

$$E[R] = 2\left(\frac{m}{\Omega}\right)^m \frac{1}{\Gamma(m)} \int_0^\infty r r^(2m-1) e^{-\frac{mr^2}{\Omega}}\, dr$$

By setting $\frac{mr^2}{\Omega} = x$, we can write

$$\begin{aligned} E[R] &= 2\left(\frac{m}{\Omega}\right)^m \frac{1}{\Gamma(m)} \left(\frac{\Omega}{m}\right)^{\frac{2m-1}{2}} \frac{\Omega}{2m} \int_0^\infty x^{m+\frac{1}{2}-1} e^{-x}\, dx \\ &= \frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(m)} \left(\frac{\Omega}{m}\right)^{\frac{1}{2}}. \end{aligned}$$

The second moment is similarly obtained, but $E[R^2] = E[Z] = \Omega$ by definition. The variance is then readily found from $E[R^2] - E^2[R]$.

### 7.14 The CDF of the Nakagami-$m$ distribution.

(a) By using the relation between the Nakagami variable $R$ and the $\chi^2_{2m}$ variable given by (6), we find

$$F_R^c(r) = P[R > r] = P[Z > r^2] = P\left[\chi^2_{2m} > \frac{r^2}{\sigma^2}\right] \tag{11}$$

By noting the relation between the $\chi^2_{2m}$ distribution and the $m$-stage Erlang distribution, and setting $r^2/\sigma^2 \triangleq 2\lambda$, we find from (7.17)

$$F_R^c(r) = \int_\lambda^\infty P(m-1, y)\, dy = Q(m-1, \lambda). \tag{12}$$

By differentiating (7.121), verify that the Nakagami PDF obtained is equivalent to (9).

(b) When $m$ is a positive real number, not necessarily an integer, we cannot use the above nice relation between the chi-squared distribution and the Poisson distribution. By proceeding in a standard manner, we have

$$F_R(r) = \int_0^r f_R(x)\, dx = \frac{2m^m}{\Omega^m \Gamma(m)} \int_0^r x^{2m-1} e^{-\frac{mx^2}{\Omega}}\, dx. \tag{13}$$

By setting $mx^2/\Omega = y$, hence writing $dx = \frac{1}{2}\left(\frac{\Omega}{m}\right)^{\frac{1}{2}} y^{-\frac{1}{2}}\, dy$, we find

$$F_R(r) = \frac{1}{\Gamma(m)} \int_0^{\frac{mr^2}{\Omega}} y^{m-1} e^{-y}\, dy = \frac{\gamma\left(m, \frac{mr^2}{\Omega}\right)}{\Gamma(m)}, \tag{14}$$

### 7.15 Upper incomplete gamma functions.

(a) The two incomplete gamma functions are related to the ordinary gamma function $\Gamma(\beta)$ by

$$\Gamma(\beta) = \int_0^\infty y^{\beta-1} e^{-y}\, dy = \gamma(\beta, \lambda) + \Gamma(\beta, \lambda). \tag{15}$$

Thus we obtain (**??**).

(b) By comparing (**??**) and (7.121), we find **??**Q-and-Gamma) for an integer $m = k$.

## 7.6 Complex-valued normal variables

**7.16 Relations between $\mathbf{\Sigma}$ and $\boldsymbol{C}$.**

Property 1: It is straightforward to prove (7.101).

Property 2:

$$\mathbf{\Sigma}' = \begin{bmatrix} \boldsymbol{A}' & \boldsymbol{B}' \\ -\boldsymbol{B}' & \boldsymbol{A}' \end{bmatrix}$$

So it is clear $\mathbf{\Sigma}$ is symmetric if and only if $\boldsymbol{A}$ is symmetric (i.e., $\boldsymbol{A}' = \boldsymbol{A}$) and $\boldsymbol{B}$ is skew-symmetric (i.e., $\boldsymbol{B}' = -\boldsymbol{B}$). Similarly

$$\boldsymbol{C}^H = \boldsymbol{A}' - i\boldsymbol{B}'.$$

Thus $\boldsymbol{C}$ is self-adjoint or Hermitian if and only if $\boldsymbol{A}$ is symmetric and $\boldsymbol{B}$ is skew-symmetric.

Property 3:

$$\det \mathbf{\Sigma} = \det[\boldsymbol{A}^2 + \boldsymbol{B}^2],$$

and

$$\det \boldsymbol{C} = \det[\boldsymbol{A} + i\boldsymbol{B}]$$

$$|\det \boldsymbol{C}|^2 = \det[\boldsymbol{A} + i\boldsymbol{B}] \det[\boldsymbol{A} - i\boldsymbol{B}] = \det[(\boldsymbol{A} + i\boldsymbol{B})(\boldsymbol{A} - i\boldsymbol{B})] = \det[\boldsymbol{A}^2 + \boldsymbol{B}^2].$$

Property 4:

$$\boldsymbol{C}\boldsymbol{C}^{-1} = (\boldsymbol{A} + i\boldsymbol{B})(\boldsymbol{P} + i\boldsymbol{Q}) = \boldsymbol{AP} - \boldsymbol{BQ} + i(\boldsymbol{BP} + \boldsymbol{AQ}).$$

Hence

$$\boldsymbol{AP} - \boldsymbol{BQ} = \boldsymbol{I}, \text{ and } \boldsymbol{BP} + \boldsymbol{AQ} = \boldsymbol{0}.$$

Hence,

$$\begin{aligned} \mathbf{\Sigma}\mathbf{\Sigma}^{-1} &= \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{P} & -\boldsymbol{Q} \\ \boldsymbol{Q} & \boldsymbol{P} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{AP} - \boldsymbol{BQ} & -(\boldsymbol{AQ} + \boldsymbol{BP}) \\ \boldsymbol{AQ} + \boldsymbol{BP} & \boldsymbol{AP} - \boldsymbol{BQ} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \end{aligned}$$

Conversely, if $\mathbf{\Sigma}\mathbf{\Sigma}^{-1} = \boldsymbol{I}$, then $\boldsymbol{C}\boldsymbol{C}^{-1} = \boldsymbol{I}$ can be derived in a similar fashion.

Property 5: If $\boldsymbol{\Sigma}$ is symmetric, then $\boldsymbol{\Sigma}^\top = \boldsymbol{\Sigma}$. Hence

$$\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma}^{-1} = \boldsymbol{I}.$$

Taking the transpose of the above identity, we have

$$(\boldsymbol{\Sigma}^{-1})^\top \boldsymbol{\Sigma} = \boldsymbol{I},$$

hence, $\boldsymbol{\Sigma}^{-1}$ is also symmetric. Then

$$\boldsymbol{P}^\top = \boldsymbol{P}, \text{ and } \boldsymbol{Q}^\top = -\boldsymbol{Q}.$$

Then LHS of (7.105) is

$$\begin{aligned}
\text{LHS} &= [\boldsymbol{x}^\top \boldsymbol{y}^\top] \begin{bmatrix} \boldsymbol{P} & -\boldsymbol{Q} \\ \boldsymbol{Q} & \boldsymbol{P} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ by \end{bmatrix} \\
&= \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{y} + \boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{P}\boldsymbol{y} \\
&= \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{P}\boldsymbol{y} - 2x^\top \boldsymbol{Q}\boldsymbol{y}.
\end{aligned}$$

The RHS of (7.105) is

$$\begin{aligned}
\text{RHS} &= [\boldsymbol{x}^\top + i\boldsymbol{y}^\top](\boldsymbol{P} + i\boldsymbol{Q})[\boldsymbol{x} - i\boldsymbol{y}] \\
&= \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{p}\boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{y} \\
&\quad + i(\boldsymbol{y}^\top \boldsymbol{P}\boldsymbol{x} - \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{y}) + i(\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y}) \\
&= \boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{p}\boldsymbol{y} - 2\boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{x},
\end{aligned}$$

where we used $\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{y} = -\boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{x}$, $\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x} = 0$ and $\boldsymbol{y}^\top \boldsymbol{Q}\boldsymbol{y} = 0$, all due to the skew-symmetry of $\boldsymbol{Q}$, i.e., $\boldsymbol{Q}^\top = -\boldsymbol{Q}$. Hence (7.105) has been proved.

**7.17*** **Joint PDF of** $(\boldsymbol{Z}, \boldsymbol{Z}^*)$**.** $\boldsymbol{W}^\top = [\boldsymbol{X}^\top \boldsymbol{Y}^\top]$ is a $2M$-dimensional Gaussian variables with zero mean and the $2M \times 2M$ covariance matrix $\boldsymbol{\Sigma}$ of (7.98). Thus, the PDF of the vector variable $\boldsymbol{W}$ is given by (7.106). By writing $\boldsymbol{X}$ and $\boldsymbol{Y}$ explicitly, the joint PDF of $(\boldsymbol{X}, \boldsymbol{Y})$ is given by (7.107). Since $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$ and $\boldsymbol{z}^* = \boldsymbol{x} - i\boldsymbol{y}$, its Jacobian is given by the first expression of (7.108), i.e.,

$$\boldsymbol{J} = \frac{\partial(\boldsymbol{z}, \boldsymbol{z}^*)}{\partial(\boldsymbol{x}, \boldsymbol{y})} = \begin{bmatrix} \boldsymbol{I}_M & \boldsymbol{I}_M \\ i\boldsymbol{I}_M & -i\boldsymbol{I}_M \end{bmatrix},$$

where $\boldsymbol{I}_M$ is an $M \times M$ identity matrix. A nontrivial step is to show the second half of (7.108), i.e.,

$$\det J = (-2i)^M.$$

We want to show the following formula by mathematical induction:

$$\det \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{I}_k \\ \boldsymbol{I}_k & -\boldsymbol{I}_k \end{bmatrix} = (-2)^k. \tag{16}$$

For $k = 1$, we have

$$\det \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = -2.$$

Thus, the formula holds for $k = 1$. Suppose that it holds for $k = n - 1$, i.e.,

$$\det \begin{bmatrix} \boldsymbol{I}_{n-1} & \boldsymbol{I}_{n-1} \\ \boldsymbol{I}_{n-1} & -\boldsymbol{I}_{n-1} \end{bmatrix} = (-2)^{n-1}.$$

We can write the identity matrix $\boldsymbol{I}_n$

$$\boldsymbol{I}_n = \begin{bmatrix} 1 & \cdots \\ \vdots & \boldsymbol{I}_{n-1} \end{bmatrix}.$$

Then we can write

$$\begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{I}_n \\ \boldsymbol{I}_n & -\boldsymbol{I}_n \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 & \cdots \\ \vdots & \boldsymbol{I}_{n-1} & \vdots & \boldsymbol{I}_{n-1} \\ 1 & \cdots & -1 & \cdots \\ \vdots & \boldsymbol{I}_{n-1} & \vdots & -\boldsymbol{I}_{n-1} \end{bmatrix}.$$

Then the determinant is calculated as

$$\begin{aligned} \det \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{I}_n \\ \boldsymbol{I}_n & -\boldsymbol{I}_n \end{bmatrix} &= 1 \cdot \det \begin{bmatrix} \boldsymbol{I}_{n-1} & \vdots & \boldsymbol{I}_{n-1} \\ \cdots & -1 & \cdots \\ \boldsymbol{I}_{n-1} & \vdots & -\boldsymbol{I}_{n-1} \end{bmatrix} + (-1)^n \cdot \det \begin{bmatrix} \vdots & \boldsymbol{I}_{n-1} & \boldsymbol{I}_{n-1} \\ 1 & \cdots & \cdots \\ \vdots & \boldsymbol{I}_{n-1} & -\boldsymbol{I}_{n-1} \end{bmatrix} \\ &= 1 \cdot (-1) \cdot \det \begin{bmatrix} \boldsymbol{I}_{n-1} & \boldsymbol{I}_{n-1} \\ \boldsymbol{I}_{n-1} & -\boldsymbol{I}_{n-1} \end{bmatrix} + (-1)^n \cdot (-1)^{n-1} \cdot \det \begin{bmatrix} \boldsymbol{I}_{n-1} & \boldsymbol{I}_{n-1} \\ \boldsymbol{I}_{n-1} & -\boldsymbol{I}_{n-1} \end{bmatrix} \\ &= -2 \cdot \det \begin{bmatrix} \boldsymbol{I}_{n-1} & \boldsymbol{I}_{n-1} \\ \boldsymbol{I}_{n-1} & -\boldsymbol{I}_{n-1} \end{bmatrix} = (-2)(-2)^{n-1} = (-2)^n \end{aligned}$$

Thus, we have proved the formula (16) by mathematical induction. Then it is clear

$$\det \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{I}_k \\ i\boldsymbol{I}_k & -i\boldsymbol{I}_k \end{bmatrix} = (-2i)^k.$$

Thus (7.108) is proved, and (7.109) follows from $f_{\boldsymbol{XY}}(\boldsymbol{x}, \boldsymbol{y})$ of (7.107), the Jacobian $|\boldsymbol{J}| = 2^M$ of (7.108), and the quadratic form (7.105).

# 8 Solutions for Chapter 8: Moment Generating Function and Characteristic Function

## 8.1 Moment Generating Function (MGF)

**8.1* Properties of logarithmic MGF.** For simplify the notation we drop the subscript $X$ of $M_X(t)$.

$$M(t) = E\left[e^{tX}\right], \;\; M'(t) = E\left[Xe^{tX}\right], \;\; M''(t) = E\left[X^2 e^{tX}\right].$$

By setting $t = 0$, we have

$$M(0) = E[1] = 1, \;\; M'(0) = E[X], \;\; M''(0) = E[X^2].$$

Letting

$$m(t) = \ln M(t), \;\; m'(t) = \frac{M'(t)}{M(t)}, \;\; m''(t) = \frac{M''(t)M(t) - M'(t)^2}{M(t)^2}.$$

By setting $t = 0$, we have

$$m(0) = 0, \;\; m'(0) = M'(0) = E[X], \;\; m''(0) = M''(0) - M'(0)^2 = E[X^2] - E[X]^2 = \sigma^2.$$

**8.2 Uniform distributions.**

(a)

$$M(t) = E\left[e^{tX}\right] = \int_0^a \frac{e^{tx}}{a}\,dx = \left[\frac{e^{tx}}{at}\right]_0^a = \frac{e^{at}-1}{at}, \;\; \text{for } \; t \neq 0.$$

For $t = 0$, by definition $M(0) = E[e^0] = 1$.

(b)

$$M(t) = \frac{1}{2a}\int_{-a}^{a} e^{tx}\,dx = \frac{1}{2a}\left[\frac{e^{tx}}{t}\right]_{-a}^{a} = \frac{e^{at} - e^{-at}}{2at} = \frac{\sinh at}{at}, \;\; t \neq 0.$$

For $t = 0$, $M(0) = 1$ by definition.

**8.3* Exponential distribution.**

$$M(t) = \int_0^\infty e^{tx}\mu e^{-\mu x}\,dx = \mu\left[\frac{e^{(t-\mu)x}}{t-\mu}\right]_0^\infty = \begin{cases} \frac{\mu}{\mu - t} & t < \mu \\ \infty & t \geq \mu \end{cases}$$

**8.4 Bilateral exponential distribution.**

$$M(t) = \int_{-\infty}^{0} \frac{e^{tx}\mu e^{\mu x}}{2}\,dx + \int_{0}^{\infty} \frac{e^{tx}\mu e^{-\mu x}}{2}\,dx = \frac{\mu}{2}\left\{\left[\frac{e^{(\mu+t)x}}{\mu+t}\right]_{-\infty}^{0} + \left[\frac{e^{(t-\mu)x}}{t-\mu}\right]_{0}^{\infty}\right\}$$
$$= \begin{cases} \infty, & t \le -\mu \\ \frac{\mu^2}{\mu^2-t^2}, & |t| < \mu \\ \infty, & t > \mu. \end{cases}$$

**8.5 Triangular distribution.**

$$M(t) - \int_{-a}^{0} \frac{e^{tx}}{a}\left(1+\frac{x}{a}\right)\,dx + \int_{0}^{a} \frac{e^{tx}}{a}\left(1-\frac{x}{a}\right)\,dx$$
$$= \left[\frac{e^{tx}}{at}\right]_{-a}^{0} + \left[\frac{e^{tx}}{at}\right]_{0}^{a} + \frac{1}{a^2}\left(\int_{-a}^{0} xe^{tx}\,dx - \int_{0}^{a} xe^{tx}\,dx\right)$$

Now use the following integration by parts:

$$\int xe^{tx}\,dx = \int x\left(\frac{e^{tx}}{t}\right)'\,dx = \frac{xe^{tx}}{t} - \int \frac{e^{tx}}{t}\,dx = \frac{(tx-1)e^{tx}}{t^2}.$$

By substituting this into the above definite integrals, and after some manipulation we obtain

$$M(t) = \frac{e^{at}+e^{-at}-2}{a^2t^2} = \left(\frac{e^{\frac{at}{2}} - e^{-\frac{at}{2}}}{at}\right)^2 = \left(\frac{\sinh\frac{at}{2}}{\frac{at}{2}}\right)^2 \ge 0, \quad -\infty < t < \infty.$$

**8.6 Negative binomial distribution.**

$$M(t) = \sum_{i=r}^{\infty} e^{ti}\binom{i-1}{r-1}p^r(1-p)^{i-r}$$
$$= \left(e^t p\right)^r \sum_{i=r}^{\infty}\binom{i-1}{r-1}\left((1-p)e^t\right)^{i-r}.$$

Using the identity (3.96), we find

$$M(t) = \left(e^t p\right)^r\left(1-(1-p)e^t\right)^{-r},$$

where $0 < (1-p)e^t < 1$. This leads to (8.121).

**8.7* Multivariate normal distribution.** The derivation is essentially the same as that for the bivariate normal distribution. Let

$$Y \triangleq t_1X_1 + \ldots + t_mX_m = \langle \boldsymbol{t}, \boldsymbol{X}\rangle.$$

Then the joint MGF of the multivariate $\boldsymbol{X}$ is given by

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = M_Y(1),$$

where $M_Y(\xi)$ is the MGF of the scalar RV $Y$, i.e., $M_Y(\xi) = E\left[e^{\xi Y}\right]$. Since $Y$ is a linear sum of the normal variables, it is also a normal variable with mean and variance given by

$$\mu_Y = \langle \boldsymbol{t}, \boldsymbol{\mu}\rangle, \text{ and } \sigma_Y^2 = \boldsymbol{t}^\top \boldsymbol{C}\boldsymbol{t}.$$

Thus, its MGF is

$$M_Y(\xi) = e^{\xi\mu_Y + \frac{\xi\sigma_Y^2}{2}}.$$

Hence, the joint MGF of $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)$ is

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = M_Y(1) = e^{\mu_Y + \frac{\sigma_Y^2}{2}},$$

which leads to (8.54).

**8.8 Multinomial distributions**

$$\begin{aligned}
M_{\boldsymbol{k}}(\boldsymbol{t}) &= \sum_{k_1+k_2+\cdots+k_m=n} e^{t_1k_1+t_2k_2+\cdots+t_mk_m} \frac{n!}{k_1!k_2!\cdots k_m!} p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m} \\
&= \sum_{k_1+k_2+\cdots+k_m=n} \frac{n!}{k_1!k_2!\cdots k_m!} \left(p_1e^{t_1}\right)^{k_1} \left(p_2e^{t_2}\right)^{k_2} \cdots \left(p_me^{t_m}\right)^{k_m} \\
&= \left(p_1e^{t_1} + p_2e^{t_2} + \cdots + p_me^{t_m}\right)^n.
\end{aligned}$$

**8.9* Erlang distribution.** By definition the MGF is

$$M_{S_r}(t) = \int_0^\infty e^{xt} \frac{r\lambda(r\lambda x)^{r-1}}{(r-1)!} e^{-r\lambda x}\,dx = \frac{(r\lambda)^r}{(r-1)!} \int_0^\infty x^{r-1} e^{-(r\lambda-t)x}\,dx$$

We use the following integration by parts:

$$\begin{aligned}
I(r) &\triangleq \int_0^\infty x^{r-1} e^{-(r\lambda-t)x}\,dx = \int_0^\infty x^{r-1} \left(-\frac{e^{-(r\lambda-t)x}}{r\lambda-t}\right)' dx \\
&= -x^{r-1} \frac{e^{-(r\lambda-t)x}}{r\lambda-t}\bigg|_0^\infty + \frac{r-1}{r\lambda-t} \int_0^\infty x^{r-2} e^{-(r\lambda-t)x}\,dx,
\end{aligned}$$

where the first term is zero if $t < r\lambda$, and is infinity, otherwise. Hence

$$I(r) = \begin{cases} \frac{r-1}{r\lambda-t} I(r-1) & t < r\lambda, \\ \infty & t \geq r\lambda. \end{cases}$$

By solving this recursively we find for $t < r\lambda$,

$$I(r) = \frac{(r-1)(r-2)}{(r\lambda-t)^2} I(r-2) = \cdots = \frac{(r-1)!}{(r\lambda-t)^{r-1}} I(1),$$

where $I(1) = \int_0^\infty e^{-(r\lambda-t)x}\,dx = (r\lambda-t)^{-1}$. Hence

$$M_{S_r}(t) = \begin{cases} \left(\frac{r\lambda}{r\lambda-t}\right)^r, & t < r\lambda \\ \infty & t \geq r\lambda. \end{cases}$$

Alternatively, $S_r$ can be expressed as the sum of $r$ i.i.d. exponential RVs with mean $(r\lambda)^{-1}$, whose MGF is given by $\frac{r\lambda}{r\lambda-t}$, for $t < r\lambda$, as seen from the solution of Problem 8.3. Then from the product formula (8.41) for the sum of independent RVs, we readily obtain the above result.

**8.10 Gamma distribution.**

$$M_{Y_{\lambda,\beta}}(t) = \int_0^\infty \frac{\lambda^\beta}{\Gamma(\beta)} y^{\beta-1} e^{-\lambda y} e^{ty}\, dy = \frac{\lambda^\beta}{\Gamma(\beta)} I(\beta),$$

where

$$\begin{aligned}
I(\beta) &\triangleq \int_0^\infty y^{\beta-1} e^{(t-\lambda)y}\, dy = \int_0^\infty y^{\beta-1} \left(\frac{e^{(t-\lambda)y}}{t-\lambda}\right)' dy \\
&= \left.\frac{y^{\beta-1} e^{(t-\lambda)y}}{t-\lambda}\right|_0^\infty + \frac{\beta-1}{\lambda-t} I(\beta-1) \\
&= \begin{cases} \infty & t > \lambda, \\ \frac{\Gamma(\beta)}{(\lambda-t)^\beta} I(1), & t < \lambda. \end{cases}
\end{aligned}$$

Thus, we obtain (8.125).

## 8.2 Characteristic Function (CF)

**8.11* CF of the binomial distribution.** By definition

$$\begin{aligned}
\phi(u) &= \sum_{k=0}^n B(k; n, p) e^{iuk} = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{iuk} \\
&= \sum_{k=0}^n \binom{n}{k} (pe^{iu})^k (1-p)^{n-k} = (pe^{iu} + 1 - p)^n.
\end{aligned}$$

**8.12 CF of the Poisson distribution.**

$$\phi(u) = \sum_{k=0}^\infty \frac{\lambda^k}{k!} e^{-\lambda} e^{iuk} = e^{-\lambda} \sum_{k=0}^\infty \frac{(\lambda e^{iu})^k}{k!} = e^{-\lambda} e^{\lambda e^{iu}} = e^{\lambda(e^{iu}-1)}.$$

**8.13 Alternative derivation of (8.67).**

(a)

$$\begin{aligned}
\frac{d}{du}\phi_Y(u) &= \int_{-\infty}^\infty (iy) e^{iuy} \left(\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\right) dy = -i \int_{-\infty}^\infty e^{iuy} \frac{d}{dy} \left(\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\right) dy \\
&= \left[-i \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} e^{iuy}\right]_{-\infty}^\infty + \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \frac{d}{dy}\left(ie^{iuy}\right) dy \\
&= -u\phi_Y(u).
\end{aligned}$$

(b) From the above differential equation, we obtain

$$\frac{d}{du}\left(\ln \phi_Y(u)\right) = -u,$$

Hence,

$$\ln \phi_Y(u) = -\frac{u^2}{2} + c_1,$$

or

$$\phi_Y(u) = c_2 e^{-\frac{u^2}{2}}.$$

By using the property $\phi_Y(0) = 1$, we determine $c_2 = 1$.

**8.14 Contour integration.**

Since the function $e^{-s^2/2}$ is analytic in the entire $s$-plane, the integral around the contour of Figure 8.1(b) is also zero due to the Cauchy-Goursat theorem. We make the change of the variable $s = y - iu;\ \ u < 0$.

$$\int_{\alpha-iu}^{-\alpha-iu} e^{-\frac{s^2}{2}}\,ds + \int_{-\alpha-iu}^{-\alpha} e^{-\frac{s^2}{2}}\,ds + \int_{-\alpha}^{\alpha} e^{-\frac{x^2}{2}}\,dx + \int_{\alpha}^{\alpha-iu} e^{-\frac{s^2}{2}}\,ds = 0, \ \text{ for }\ u < 0.$$

The second and fourth terms approach zero as $\alpha \to \infty$, because the integral in both cases contains a factor $e^{-\frac{\alpha^2}{2}}$ (See the foot note given in the discussion for the contour integration for the case $u > 0$). Thus, by letting $\alpha \to \infty$,

$$\int_{\infty-iu}^{-\infty-iu} e^{-\frac{s^2}{2}}\,ds + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = 0.$$

By reversing the direction of integration in the first term, we find

$$\int_{-\infty-iu}^{\infty-iu} e^{-\frac{s^2}{2}}\,ds = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = \sqrt{2\pi}, \ \ u < 0.$$

(8.67) holds for $u < 0$ as well,

**8.15* CF of the exponential distribution** The CF is by definition

$$\phi_X(u) = \int_0^\infty \frac{1}{a} e^{-x/a}\, e^{iux}\,dx = \frac{1}{a}\lim_{R\to\infty}\int_0^R e^{-\frac{(1-iau)x}{a}}\,dx. \tag{1}$$

Defining the complex variable $z$, we write the above as an integral in the complex plane:

$$\phi_X(u) = \frac{1}{a(1-iau)}\lim_{R\to\infty}\int_0^{R-iauR} e^{-\frac{z}{a}}\,dz. \tag{2}$$

Then for $u > 0$, we consider the contour shown in Figure 8.1 (a). Since the function $e^{-\frac{z}{a}}$ is analytic in the entire plane, its integration along the contour is obviously zero:

$$0 = \oint_C e^{-\frac{z}{a}}\,dz = \int_0^{R-iauR} e^{-\frac{z}{a}}\,dz + \int_{-auR}^{0} e^{-\frac{R+iy}{a}}\,i\,dy + \int_R^0 e^{-\frac{x}{a}}\,dx. \tag{3}$$

The second integral on the right hand side can be bounded as follows:

$$\left|\int_{-auR}^{0} e^{-\frac{R+iy}{a}}\,i\,dy\right| \le e^{-\frac{R}{a}}\int_{-auR}^{0} dy = auRe^{-\frac{R}{a}} \ \stackrel{R\to\infty}{\longrightarrow}\ 0. \tag{4}$$

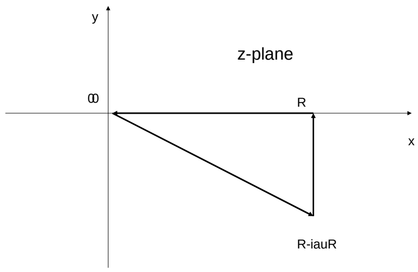
Therefore, we have

$$\lim_{R\to\infty}\int_0^{R-iauR} e^{-\frac{z}{a}}\,dz = \lim_{R\to\infty}\int_0^R e^{-\frac{x}{a}}\,dx = a. \tag{5}$$

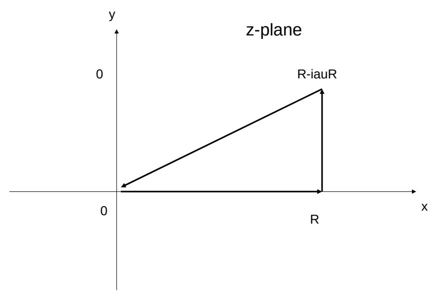
Thus, we find

$$\boxed{\phi_X(u) = \frac{1}{1 - iau}, \quad -\infty < u < \infty} \tag{6}$$

for $u > 0$.




(a) for $u > 0$




(b) for $u < 0$

**Figure 8.1** Contours for complex integral to obtain the CF of the exponential distribution.

For $u < 0$, we consider the contour shown in Figure 8.1 (b), and repeat similar steps as in the case for $u > 0$. In doing so, we can show that (6) holds for $u < 0$, as well. For $u = 0$, the integration in (1) reduces to an integration on the real parameter $x$, and is readily obtained as $\phi_X(0) = 1$, which satisfies (6).

As is the case with the normal distribution, the result (6) could have been obtained by substitution of $t = iu$ in the MGF $M_X(t) = \frac{1}{1-at}$ of the exponential distribution. The reader is cautioned again that such derivation is mathematically incorrect.

The *cumulant generating function*, given by

$$\psi_X(u) \triangleq \ln \phi_X(u) = -\ln(1 - iau).$$

By differentiating the above expression, we obtain

$$\mu_X = (-i)\psi'_X(0) = a, \text{ and } \sigma_X^2 = (-i)^2\psi''_X(0) = a^2.$$

**8.16 CF of the bilateral exponential distribution.** Write

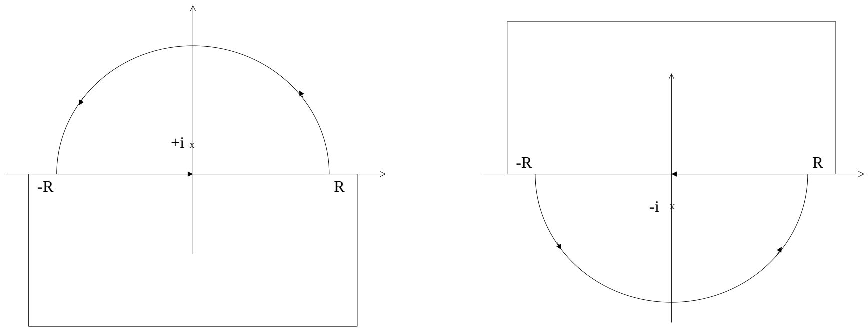
$$f(x) = \begin{cases} \frac{e^{-x}}{2} & x \geq 0 \\ \frac{e^{x}}{2} & x < 0 \end{cases}$$

Then

$$\phi(u) = \frac{1}{2}\left(\int_{-\infty}^{0} e^{x+iux}\, dx + \int_{0}^{\infty} e^{-x+iux}\, dx\right) \triangleq \frac{\phi_-(u) + \phi_+(u)}{2},$$

where $\phi_+(u)$ is the CF of the exponential distribution with mean $a = 2$ (see Example **??**), hence $\phi_+(u) = \frac{1}{1-iu}$. It is not difficult to see that $\phi_-(u) = \phi_+^*(u)$ for $-\infty < u < \infty$. Hence

$$\phi(u) = \frac{\frac{1}{1-iu} + \frac{1}{1+iu}}{2} = \frac{1}{1+u^2}.$$




(a) for $u > 0$ (b) for $u < 0$

**Figure 8.2** Contours for complex integral to obtain the CF of the Cauchy distribution.

**8.17 CF of the Cauchy distribution**

By integrating the PDF, we find

$$F_X(x) = \int_{-\infty}^{x} \frac{du}{\pi(1+u^2)} = \frac{1}{\pi} \tan^{-1} x + \text{const.} \tag{7}$$

From the condition $F(\infty) = 1$, we can determine the integration constant, obtaining

$$F_X(x) = \frac{\tan^{-1} x}{\pi} + \frac{1}{2}.$$

(b) Thus, $F_X(0) = \frac{1}{2}$. which confirms that the PDF $f_X(x)$ is symmetric and centered around $x = 0$. Thus, we may be tempted to hastily conclude that $\mu_X = E[X] = 0$. However, the mean $\mu_X$ does not exist, because the integral

$$\begin{aligned} M_1 &= \int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}\, dx = \lim_{a\to\infty} \int_{-a}^{0} \frac{x}{\pi(1+x^2)}\, dx + \lim_{b\to\infty} \int_{0}^{b} \frac{x}{\pi(1+x^2)}\, dx \\ &= \frac{1}{2\pi}\left[ -\lim_{a\to\infty} \ln(1+a^2) + \lim_{b\to\infty} \ln(1+b^2) \right]. \end{aligned} \tag{8}$$

The integral does not exist because both the first integral and the second integral are infinite. If we let $b = a$ and then take the limit $a \to \infty$, then the above $M_1$ becomes zero. and this is called the **Cauchy's principal value**.

The second moment is

$$M_2 = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x^2}{1+x^2}\, dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \left(1 - \frac{1}{1+x^2}\right) dx = \frac{1}{\pi} \int_{-\infty}^{\infty} dx = \infty - 1.$$

Hence $M_2$ does not exist. Similarly the Cauchy's principal value for the third moment $M_3$ is zero, but otherwise ill-defined. Thus, we can conclude that the $n$th moment of the Cauchy distribution is infinite for even values of $n$ and its Cauchy's principal value is zero for odd $n$. Whether or not we assign the principal value to the odd moments, the MGF does not exist for the Cauchy distribution, because not all moments exist.

(c) By defining a complex variable $z = x + iy$, consider a complex function

$$g(z) \triangleq \frac{e^{iuz}}{\pi(1+z^2)},$$

which is not an analytic function; it has **poles** at $z = i$ and $z = -i$. For $u > 0$, the numerator approaches zero as $y \to +\infty$, since $|e^{iuz}| = |e^{iux-uy}| = e^{-uy}$. For $u < 0$, the same term tends to zero as $y \to -\infty$. Therefore, we take different contours depending on $u \gtrless 0$, as shown in Figure 8.2.

For $u > 0$, we integrate the function $g(z)$ along the contour shown in Figure 8.2 (a) and the line integration along the upper semi-circle approaches zero as $R \to \infty$, and the contour includes a singular point at $z = i$. Thus, from the **Cauchy's residue theorem** we have

$$\oint g(z)\,dz = \lim_{R\to\infty}\int_{-R}^{R}\frac{e^{iux}}{\pi(1+x^2)}\,dx = 2\pi i(z-i)\frac{e^{iuz}}{\pi(z^2+1)}\bigg|_{z=i} = e^{-u}, \quad u > 0. \tag{9}$$

Similarly for $u < 0$, we integrate $g(z)$ along the contour shown in Figure 8.2(b). Then the pole $z = -i$ is surrounded by this contour. Hence, we find

$$\oint g(z)\,dz = -\lim_{R\to\infty}\int_{-R}^{R}\frac{e^{iux}}{\pi(1+x^2)}\,dx = 2\pi i(z+i)\frac{e^{iuz}}{\pi(z^2+1)}\bigg|_{z=-i} = -e^{u}, \quad u < 0. \tag{10}$$

Combining the above results we have

$$\phi_X(u) = e^{-|u|}, \quad -\infty < u < \infty.$$

An alternative approach is to use the Fourier inverse formula of the characteristic function (Problem 8.18).

**8.18 Alternative derivation of the CF of the Cauchy distribution.**

From the result of the bilateral exponential density function, we have

$$\phi(u) \triangleq \mathcal{F}\left\{\frac{e^{-|x|}}{2}\right\} = \frac{1}{1+u^2}.$$

Thus, the corresponding Fourier inversion is

$$\frac{e^{-|x|}}{2} = \mathcal{F}^{-1}\left\{\frac{1}{1+u^2}\right\} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-iux}}{1+u^2}\,du.$$

By interchanging the symbols $x$ and $u$, and taking the complex conjugate of both sides, we have

$$\frac{e^{-|u|}}{2} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{e^{ixu}}{\pi(1+x^2)}\,dx.$$

which implies

$$\mathcal{F}\left\{\frac{1}{\pi(1+x^2)}\right\} = e^{-|u|}.$$

**8.19 CF of the gamma and $\chi^2$ distributions.**

(a)

$$\phi_X(u) = \int_0^{\infty}\frac{x^{\beta-1}e^{-(1-iu)x}}{\Gamma(\beta)}\,dx.$$

Similar to the example on the exponential distribution, set $z = (1 - iu)x$, or $x = \frac{z}{1-iu}$, we have

$$\phi_X(u) = \int_0^\infty \frac{z^{\beta-1} e^{-z}\, dz}{\Gamma(\beta)(1-iu)^\beta} = \frac{1}{(1-iu)^\beta} \int_0^\infty g(z)\, dz.$$

where $g(z) \triangleq \frac{z^{\beta-1}e^{-z}}{\Gamma(\beta)}$. Then for $u > 0$, we take the same sector counter as done in the example on the exponential distribution, and following the same step, we have

$$\int_0^\infty g(z)\, dz = \int_0^\infty g(x)\, dx = \int_0^\infty \frac{x^{\beta-1} e^{-x}\, dx}{\Gamma(\beta)} = 1,$$

where we used the fact that the integrand of the above is the gamma density function. As in the exponential case, the same holds for $u < 0$. Hence, we have shown

$$\phi_X(u) = \frac{1}{(1-iu)^\beta}.$$

(b)

$$\phi_{\chi_n^2}(u) = \int_0^\infty \frac{x^{n/2-1} e^{-x/2} e^{iux}\, dx}{2^{n/2}\Gamma(n/2)}.$$

If we set $\chi_n^2 = 2Y_n$, then

$$f_{Y_n}(y) = \frac{y^{n/2-1} e^{-y}}{\Gamma(n/2)},$$

which is the gamma distribution with $\beta = n/2$ and $\lambda = 1$.

$$\begin{aligned}
\phi_{\chi_n^2}(u) &= \int_0^\infty \frac{2^{n/2-1} y^{n/2-1} e^{-y} e^{2iuy} 2\, dy}{2^{n/2}\Gamma(n/2)} \\
&= \int_0^\infty \frac{y^{n/2-1} e^{-(1-2iu)y}\, dy}{\Gamma(n/2)} = \frac{1}{(1-2iu)^{n/2}},
\end{aligned}$$

which is obtained by setting $\beta = n/2$ and replacing $u$ by $2u$ in the CF of the gamma distribution in part (a).

**8.20 CF of the non-central $\chi^2$ distribution.**

$$M_{\chi_n^2(\mu^2)}(u) = E\left[e^{iu\chi_n^2(\mu^2)}\right] = \prod_{k=1}^n E\left[e^{iu(X_k+\mu_k)^2}\right],$$

where $X_k \sim N(0,1)$ for all $k$.

$$\begin{aligned}
E\left[e^{iu(X_k+\mu_k)^2}\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{iu(x+\mu_k)^2} e^{-\frac{x^2}{2}}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left\{ -\frac{(1-2iu)}{2} \left( x - \frac{2-u\mu_k}{1-2iu} \right)^2 + \frac{iu\mu_k^2}{1-2iu} \right\} \\
&= \frac{\exp\left\{\frac{iu\mu_k^2}{1-2iu}\right\}}{\sqrt{1-2iu}}.
\end{aligned}$$

Thus,

$$M_{\chi_n^2(\mu^2)}(u) = \frac{1}{(1-2iu)^{\frac{n}{2}}} \exp\left\{\frac{iu\mu^2}{1-2iu}\right\}.$$

**8.21 Independent random variables.**

(a) Since the joint PDF is the product of the marginal PDFs,

$$f_{XY}(x,y) = f_X(x) f_Y(y),$$

so we have

$$\begin{aligned} E[g(X)h(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)\,dx\,dy \\ &= \int_{-\infty}^{\infty} g(x)f_X(x)\,dx \int_{-\infty}^{\infty} h(y)f_Y(y)\,dy = E[g(X)]E[h(Y)]. \end{aligned}$$

(b) Set $g(X) = e^{iuX}, h(Y) = e^{iuT}$. Then

$$E\left[e^{iuZ}\right] = E\left[e^{iuX}e^{iuY}\right] = E\left[e^{iuX}\right]E\left[e^{iuY}\right],$$

because $X$ and $Y$ are independent.

**8.22 CF of a symmetric distribution.**

$$\phi_X(u) = \int_{-\infty}^{\infty} \cos ux f_X(x)\,dx + i\int_{-\infty}^{\infty} \sin ux f_X(x)\,dx.$$

The second term is zero, because $\sin ux$ is an odd (i.e., skew-symmetric) function of $x$.

**8.23 Sum of independent unit normal variables.** Let

$$Y = \frac{U_1 + U_2 + \cdots + U_n}{\sqrt{n}}.$$

The CF of $Y$ is

$$\begin{aligned} \phi_Y(u) = E\left[e^{iuY}\right] &= \prod_{k=1}^{n} E\left[e^{\frac{iuX_k}{\sqrt{n}}}\right] = \prod_{k=1}^{n}\phi_X\left(\frac{u}{\sqrt{n}}\right) \\ &= \left(e^{-\frac{u^2}{2n}}\right)^n = e^{-\frac{u^2}{2}}. \end{aligned}$$

Thus, the PDF of $Y$ is also the unit normal distribution.

**8.24 Poisson distribution.**

(a)

$$\phi_X(u) = e^{\lambda(e^{iu}-1)}.$$

(b)

$$\psi(u) = \ln\phi(u) = \lambda\left(e^{iu}-1\right).$$

(c)

$$\phi_Y(u) = \prod_{k=1}^{n} e^{\lambda_k\left(e^{iu}-1\right)},$$

where $\lambda = \sum_{k=1}^{n} \lambda_k$. Hence, $Y$ has the Poisson distribution with parameter $\lambda = \sum_{k=1}^{n} \lambda_k$.

**8.25 Bernoulli trials.**

(a) The first $(k-1)$ trials are failure, and then the success:

$$p_k = q^k p, \quad k = 0, 1, 2, \ldots.$$

(b)

$$\phi_X(u) = \sum_{k=0}^{\infty} pq^k e^{iuk} = \frac{p}{1 - qe^{iu}}.$$

$$\psi_X(u) = \ln p - \ln(1 - qe^{iu}).$$

$$\psi'_X(0) = \frac{iq}{1-q}, \quad \psi''_X(0) = -\frac{q}{(1-q)^2} = -\frac{1-p}{p^2}.$$

Hence

$$\mu_X = \frac{q}{p}, \quad \text{and} \quad \sigma_X^2 = \frac{q}{p^2}.$$

**8.26 The joint CF of a multinomial distribution.**

(a) Use the method of mathematical induction. The case $m = 2$ corresponds to the binomial distribution.

(b)

$$\begin{aligned}
\phi(u) &= e\left[e^{i(u_1k_1 + u_2k_2 + \cdots + u_mk_m)}\right] \\
&= \sum_{k_1 + k_2 + \cdots + k_m = n} \prod_{j=1}^{m} \left(p_j e^{iu_j}\right)^{k_j} \\
&= \left(p_1 e^{iu_1} + p_2 e^{iu_2} + \cdots + p_m e^{iu_m}\right)^n.
\end{aligned}$$

(c)

$$\phi(u_1, 0, \ldots, 0) = \left(p_1 e^{iu_1} + p_2 + \cdots * p_m\right)^n = \left[p_1 e^{iu_1} + 1 - p\right]^n.$$

Hence,

$$E[k_1] = (-i)\left[\frac{\partial \phi(u_1, 0, \ldots, 0)}{\partial u_1}\right]_{u_1=0} = np_1.$$

Similarly,

$$\begin{aligned}
E\left[k_1^2\right] &= np_1(1-p_1)\phi(u_1, u_2, 0, \ldots, 0) = \left(p_1 e^{iu_1} + p_2 e^{iu_2} + p_3 + \cdots + p_m\right)^n \\
&= \left[\left(e^{iu_1} - 1\right)p_1 + \left(e^{iu_2} - 1\right)p_2 + 1\right]^n,
\end{aligned}$$

and

$$E[k_1k_2] = (-i)^2\left[\frac{\partial^2 \phi(u_1, u_2, 0, \ldots, 0)}{\partial u_1 \partial u_2}\right]_{u_1=u_2=0} = n(n-1)p_1p_2.$$

Hence

$$\text{Cov}[k_1, k_2] = n(n-1)p_1p_2 - np_1np_2 = -np_1p_2.$$

**8.27 Sample mean of the Cauchy variables.** Let $X_k$'s be i.i.d. RVs with the Cauchy distribution. Then from Problem 8.17, we know their CFs, i.e.,

$$\phi_{X_k}(u) = e^{-|u|}, \quad -\infty < u < \infty, \text{ for all } k = 1, 2, \ldots, n.$$

Let

$$S_n = X_1 + X_2 + \ldots + X_n,$$

Then, its CF is

$$\phi_{S_n}(u) = \prod_{k=1}^{n} \phi_{X_k}(u) = e^{-n|u|}, \quad -\infty < u < \infty.$$

The sample mean of the $n$ RVs $X_k$'s is

$$\overline{X}_n = \frac{S_n}{n},$$

so its CF is

$$\phi_{\overline{X}_n}(u) = E\left[e^{\frac{uS_n}{n}}\right] = \phi_{S_n}\left(\frac{u}{n}\right) = e^{-|u|}.$$

Thus $\overline{X}_n$ has the same Cauchy distributions as the individual $X_k$'s. Hence the CLT does not apply to the independent RVs drawn from the Cauchy distribution.

**8.28 Moments of complex-valued multivariate normal variables.**
We have

$$E\left[e^{\frac{i}{2}(\boldsymbol{\gamma}^\top \boldsymbol{Z}^* + \boldsymbol{Z}^\top \boldsymbol{\gamma})}\right] = e^{Q(\boldsymbol{\gamma})},$$

where

$$Q(\boldsymbol{\gamma}) \triangleq -\frac{1}{2}\boldsymbol{\gamma}^\top \boldsymbol{C}\boldsymbol{\gamma}^*.$$

(a) By differentiating both side of the first equation with respect to $\gamma_m$ and setting $\boldsymbol{\gamma} = \mathbf{0}$,

$$\frac{i}{2}E[Z_m^*] = \left.\frac{\partial Q(\boldsymbol{\gamma})}{\partial \gamma_m} e^{Q(\boldsymbol{\gamma})}\right|_{\boldsymbol{\gamma}=\mathbf{0}} = -\frac{1}{2}\left.\sum_n C_{mn}\gamma_n^* e^{Q(\boldsymbol{\gamma})}\right|_{\gamma_n=0} = 0.$$

If we differentiate with respect to $\gamma^*$, we have $E[Z_m] = 0$.

(b) Similarly, by taking partial derivative with respect to $\gamma_m$ and $\gamma_n^*$ in the first equation and setting $\boldsymbol{\gamma} = \mathbf{0}$, we have

$$-\frac{1}{4}E[Z_m^* Z_n] = \left[\frac{\partial^2 Q(\boldsymbol{\gamma})}{\partial \gamma_m \gamma_n^*} e^{Q(\boldsymbol{\gamma})} + \frac{\partial Q(\boldsymbol{\gamma})}{\partial \gamma_m}\frac{\partial Q(\boldsymbol{\gamma})}{\partial \gamma_n^*} e^{Q(\boldsymbol{\gamma})}\right]_{\boldsymbol{\gamma}=\mathbf{0}} = -\frac{1}{2}C_{mn} + 0.$$

Hence $E[Z_m^* Z_n] = 2C_{mn}$.

(c) If we differentiate with respect to $\gamma_p$ in the above procedure before we set $\boldsymbol{\gamma} = \mathbf{0}$, we find

$$-\frac{i}{8}E[Z_m^* Z_n Z_p^*] = \left[-\frac{1}{2}C_{mn}\left(-\frac{1}{2}\sum_q C_{pq}\gamma_q^*\right)e^{Q(\boldsymbol{\gamma})} + \frac{\partial^2 Q(\boldsymbol{\gamma})}{\partial\gamma_m\partial\gamma_p}\frac{\partial Q(\boldsymbol{\gamma})}{\partial\gamma_n^*}e^{Q(\boldsymbol{\gamma})}\right.$$
$$\left. + \frac{\partial Q(\boldsymbol{\gamma})}{\partial\gamma_n^*\partial\gamma_p}e^{Q(\boldsymbol{\gamma})} + \frac{\partial Q(\boldsymbol{\gamma})}{\partial\gamma_m}\frac{\partial Q(\boldsymbol{\gamma})}{\partial\gamma_n^*}\frac{\partial Q(\boldsymbol{\gamma})}{\partial\gamma_p}e^{Q(\boldsymbol{\gamma})}\right]_{\boldsymbol{\gamma}=\mathbf{0}} = 0+0+0+0 = 0$$

(d) Similarly, by differentiating with respect to $\gamma_n^*$, and setting $\boldsymbol{\gamma} = \mathbf{0}$,

$$\frac{1}{16}E[Z_m^* Z_n Z_p^* Z_q] = \left[\left(-\frac{1}{2}C_{mn}\right)\left(-\frac{1}{2}C_{pq}\right)e^{Q(\boldsymbol{\gamma})}0 + \frac{\partial^2 Q(\boldsymbol{\gamma})}{\partial\gamma_m\gamma_q^*}\frac{\partial^2 Q(\boldsymbol{\gamma})}{\partial\gamma_n^*\partial\gamma_p}e^{Q(\boldsymbol{\gamma})} + 0 + 0\right]_{\boldsymbol{\gamma}=\mathbf{0}}$$
$$= \frac{1}{4}[C_{mn}C_{pq} + C_{mq}C_{pn}]$$

# 9 Solutions for Chapter 9: Generating Functions and Laplace Transform

## 9.1 Generating Functions

**9.1* Region of convergence for PGF, generating function and $Z$-transform.**

(a) If $|f_k| \le M$ for all $k$ and some constant $M$, then

$$|F(z)| \le \sum_{k=0}^{\infty} |f_k z^k| \le M \sum |z^k| = M\frac{1}{1-|z|}, \text{ for } |z| < 1.$$

(b) Similarly

$$|\tilde{F}(z)| \le \sum_{k=0}^{\infty} |f_k z^{-k}| \le M \sum |z^{-k}| = M\frac{1}{1-|z|^{-1}}, \text{ for } |z^{-1}| < 1, \text{ or } |z| > 1.$$

(c) Since $\sum_k p_k = 1$ by definition

$$|P(z)| \le P(1) = \sum_k p_k = 1, \text{ for } |z| \le 1.$$

**9.2* Derivation of PGFs in Table 9.1.**

(a) The PGF is given by

$$P(z) = \sum_{k=0}^{n} \binom{n}{k} (pz)^k q^{n-k} = (pz+q)^n, \ |z| < \infty.$$

(b)

$$P(z) = \sum_{k=1}^{\infty} z(zq)^{k-1} p = pz \sum_{j=0}^{\infty} (zq)^j = C, \ |z| < q^{-1}.$$

(c) We can write

$$Z_r = X_1 + X_2 + \ldots + X_r.$$

where $X_i$ is the number of failures until the $i$ success is attained after the $(i-1)$st success. Then $X_i$ has the shifted geometric distribution with its PGF $\frac{pz}{1-qz}$, as obtained in Example 9.1. Since $X_i$'s are i.i.d., we have the PGF of $Z_r$ given by $\left(\frac{pz}{1-qz}\right)^r$.

**9.3 Derivation of (9.10).**

The coefficient of $z^k$ in $(1-z)Q(z)$ is equal to $q_k - q_{k-1} = -p_k$ when $k \geq 1$, and is equal to $q_0 = p_1 + p_2 + \ldots = 1 - p_0$ when $k = 0$. Therefore

$$(1-z)Q(z) = 1 - P(z).$$

**9.4 Moments of binomial distribution.** The PGF of the binomial distribution is $P(z) = (pz+q)^n$. Thus, $Q(z) = \frac{1-(pz+q)^n}{1-z}$. Then

$$E[X] = P'(1) = Q(1) = np, \text{ and } E[X(X-1)] = E[X^2] - E[X] = P''(1) = 2Q'(1) = n(n-1)p^2.$$

Thus

$$\sigma_X^2 = (n-1)p^2 + np - (np)^2 = np(1-p) = npq.$$

**9.5 Examples of a generating function.**

(a) $F(z) = \frac{1}{1-z}, \;\; |z| < 1.$
(b) $F(z) = \frac{z^n}{1-z}, \;\; |z| < 1.$
(c) $F(z) = e^z.$
(d) $F(z) = (1+z)^n.$

**9.6 Some properties of PGF.**

(a) Since $P(z) = \sum_{k=0}^{\infty} p_k z^k$, it is apparent that $P(1) = \sum_k p_k = 1$ from the definition of the probability distribution.
(b) It is also apparent that $P(0) = p_0 z^0 = p_0$.

**9.7 Generating function of a sequence.**

(a)

$$F(z) = \sum_{k=0}^{\infty} (\alpha z)^k,$$

Hence $f_k = \alpha^k, \;\; |\alpha| < 1.$
(b) Let $G(z) = \frac{1}{1-\alpha z}$ Then from the result of part (a), $G(z) = \sum_{k=0}^{\infty} \alpha^k z^k$. Since $G'(z) = \frac{\alpha}{(1-\alpha z)^2} = \sum_{k=1}^{\infty} k\alpha^k z^{k-1}$. Hence,

$$F(z) = \frac{1}{(1-\alpha z)^2} = \frac{G'(z)}{\alpha} = \sum_{k=1}^{\infty} \alpha^{k-1} k z^{k-1} = \sum_{j=0}^{\infty} \alpha^j (j+1) z^j.$$

Hence, $f_k = (k+1)\alpha^k, \;\; k = 0, 1, 2, \ldots.$
(c) With $G(z)$ as defined in part (b), we have

$$F(z) = \frac{\alpha z}{(1-\alpha z)^2} = zG'(z) = \sum_{k=1}^{\infty} k\alpha^k z^k.$$

Hence

$$f_k = k\alpha^k, \;\; k = 0, 1, 2, \ldots.$$

**9.8** 1. **PGF of a sum of random variables.**
For $W$ defined by (9.25), we have

$$P_W(z) = E\left[z^W\right] = E\left[\prod_{i=1}^n z^{X_i}\right].$$

Since $X_i$'s are independent, we can take the expectation of the individual factors $z^{X_i}$ separately:

$$P_W(z) + \prod_{i=1}^n E\left[z^{X_i}\right] = \prod_{i=1}^n P_i(z).$$

**9.9 Convolution of Poisson Distributions** Consider $Y = X_1 + X_2$, where $X_1$ and $X_2$ are independent Poisson variables with means $\lambda_1$ and $\lambda_2$. Then from (9.8) of Example 9.2, we have

$$P_Y(z) = P_{X_1}(z)P_{X_1}(z) = e^{(\lambda_1+\lambda_2)(z-1)}.$$

Thus, the RV $Y$ is also a Poisson variable, but with mean $\lambda = \lambda_1 + \lambda_2$.
This result implies that the convolution of two Poisson distributions is a Poisson distribution, which can be written in terms of the notation $P(k;\lambda)$ defined by (3.77):

$$\{P(k;\lambda)\} = \{P(k;\lambda_1)\} \circledast \{P(k;\lambda_2)\}.$$

Thus, the Poisson distribution possesses the **reproductive property** similar to the normal distribution discussed in Example 8.6.

**9.10* Shifted negative binomial distributions**

(a) The distribution of $X$ is the shifted geometric distribution discussed in Example 9.1. Using the result (9.6), we find

$$E[X] = \frac{q}{p} \text{ and } \mathrm{Var}[X] = \frac{q}{p^2}.$$

(b) We can write $Z_r$ as

$$Z_r = X_1 + X_2 + \ldots + X_r,$$

where $X_k$ denotes the number of failures after the $(k-1)$st success and prior to the $k$th success. Since $X_k$'s are independent, the PGF of the RV $Z_r$ is

$$P_{Z_r}(z) = \left(\frac{p}{1-qz}\right)^r = p^r(1-qz)^{-r}. \tag{1}$$

The mean and variance are readily obtained as

$$E[Z_r] = \frac{rq}{p} \text{ and } \mathrm{Var}[Z_r] = \frac{rq}{p^2}.$$

(c) From the identity of the hint

$$(1-qz)^{-r} = \sum_{j=0}^{\infty}\binom{-r}{j}(-qz)^j, \;\; |z| < q^{-1}.$$

Thus, from (1),

$$P_{Z_r}(z) = p^r \sum_{j=0}^{\infty} \binom{-r}{j} (-qz)^j.$$

(d) The coefficient of $z^k$ is given by (9.117), where the second expression was obtained by using the identity (3.97):

$$\binom{-n}{i} = \frac{(-n)!}{i!(-n-i)!} = (-1)^i \frac{n(n+1)\cdots(n+i-1)}{i!} = (-1)^i \binom{n+i-1}{i}, \quad (2)$$

(e) The expression (1) suggests that the (shifted) negative binomial distribution $f(j;r,p)$ is $r$-fold convolutions of the (shifted) geometric distribution:

$$\{f(k;r,p)\} = \{q^k p\}^{r\circledast},$$

which implies the reproductive property.
From the definition of $Z_r$, it is apparent that

$$Z_{r_1} + Z_{r_2} = Z_{r_1+r_2},$$

where $Z_{r_1}$ and $Z_{r_2}$ are independent in the Bernoulli trials.
Recall that the negative binomial distribution can be extended to the case for a positive real $r$ (but still $0 < p < 1$) as defined in (3.109). The generating function remains the same as (1).

**9.11 Formula (9.63).**

$$\frac{1}{1-\alpha z} = \sum_{i=0}^{\infty} \alpha^i z^i$$

$$\frac{d}{dz}\frac{1}{1-\alpha z} = \frac{\alpha}{(1-\alpha z)^2} = \sum_{i=1}^{\infty} i\alpha^i z^{i-1}$$

$$\frac{d^j}{dz^j}\frac{1}{1-\alpha z} = \frac{(j-1)!\alpha^{j-1}}{(1-\alpha z)^j} = \sum_{i=j-2} i(i-1)\cdots(i-j+2)\alpha^i z^{i-j+1}$$

Thus,

$$\frac{1}{(1-\alpha z)^j} = \sum_{i=j-2}^{\infty} \binom{i}{j-1} (\alpha z)^{i-j+1}.$$

Setting $i - j + 1 = k$, we find

$$\frac{1}{(1-\alpha z)^j} = \sum_{k=0}^{\infty} \binom{k+j-1}{j-1} (\alpha z)^k = \sum_{k=0}^{\infty} \binom{k+j-1}{k} (\alpha z)^k.$$

**9.12 Final value theorem.**

As (9.64) for $p_k$ indicates, the partial fraction of $F(z)$ consists of a series of terms, each of which has a component time sequence of the form $k^j \alpha_i^k$ for $k \geq 0$. If $f_k$ is to approach a finite limit as $k \to \infty$, it is necessary that

$$|\alpha_i| < 1, \quad \text{for all } i.$$

The limit of all components of $f_k$ for which $j > 0$ will be zero. The only term that has a finite a non-zero limit is a term of the form

$$\frac{c}{1-z}.$$

Then

$$\lim_{k\to\infty} f_k = c,$$

and it is clear that

$$\lim_{z\to 1}(z-1)F(z) = z.$$

**9.13 Joint PGF.**

(a) Differentiate the given definition of $P_{\boldsymbol{X}}(\boldsymbol{z})$ with respect $z_1, z_2, \ldots, z_m$, $k_1$ times, $k_2$ times, $\ldots$, $k_m$ times, respectively.
(b) $P_{\boldsymbol{X}}(\boldsymbol{z}) = (p_1 z_1 + p_2 z_2 + \cdots + p_m z)m)^n$.

**9.14 Negative binomial (or Pascal) distribution.**

(a) The equivalence is obvious, but we could show the equivalence in terms of the PGF.

$$\begin{aligned} P[Y^{(i)} = k] &= pq^{k-1}, \ \ k \geq 1. \\ P_{Y^{(i)}}(z) &= \sum_{k=1}^{\infty} pq^{k-1} z^k = p\left(z^{-1} - q\right)^{-1}. \\ P_{S_r}(z) &= p^r \left(z^{-1} - q\right)^{-r}. \end{aligned}$$

From (9.120)

$$\begin{aligned} \sum_{k=r}^{\infty} p_k z^k &= \sum_{k=r}^{\infty} \binom{k-1}{r-1} (pz)^r (qz)^{k-r} \\ &= (pz)^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} (qz)^{k-r} \\ &= (pz)^r (1-qz)^{-r} = p^r (z^{-1} - q)^{-r}, \end{aligned}$$

where we used the Newton's generalized binomial formula (9.116) to obtain the second line of the above.
(b) It will be more convenient to use the CF than PGF in this case. By setting $z = e^{iu}$, we find

$$\phi_{S_r}(u) = p^r \left(e^{-iu} - q\right)^{-r}.$$

Taking the logarithm

$$\psi_{S_r}(u) = r \ln p - r \ln\left(e^{-iu} - q\right).$$

Then

$$\psi'_{S_r}(u) = \frac{rie^{-iu}}{e^{-iu} - q}, \ \text{ and } \ \psi''_{S_r}(u) = \frac{-rie^{-iu}}{\left(e^{-iu} - q\right)^2}.$$

Thus,

$$E[S_r] = (-i)\psi'_{S_r}(0) = \frac{n}{1-q} = \frac{n}{q}, \text{ and } \mathrm{Var}[S_r] = -\psi''_{S_r}(0) = \frac{nq}{(1-q)^2} = \frac{nq}{p^2}.$$

**9.15* Derivation of the binomial distribution via a two-dimensional generating function $\boldsymbol{C(z,w)}$.**

(a)

$$\begin{aligned} b(k;n,p) &= b(k-1;n-1,p)p + b(k;n-1,p)q \\ b(0;n,p) &= b(0;n-1,p)q, \quad n \geq k \geq 1. \end{aligned}$$

(b)

$$\begin{aligned} B(z;n,p) &= pz\sum_{k=1}^{\infty} b(k-1;n-1,p)z^{k-1} + q\sum_{k=0}^{\infty} b(k;n-1,p)z^k \\ &= pzB(z;n-1,p) + qB(z;n-1,p) \\ &= (pz+q)B(z;n-1,p), \quad n \geq 1. \\ B(z;0,p) &= 1. \end{aligned}$$

(c) From the result in (b) we find immediately

$$C(z,w;p) = (pz+q)wC(z,w;p) + 1,$$

from which we have

$$C(z,w;p) = \frac{1}{1-w(pz+q)} = \sum_{n=0}^{\infty}(pz+q)^n w^n.$$

Therefore, we find

$$B(z;n,p) = (pz+q)^n,$$

and

$$b(k;n,p) = \binom{n}{k} p^k q^{n-k}.$$

**9.16 Example of the recursion method.**

$$\sum_{k=0}^{\infty} p_k z^k = \frac{66 - 69z + 3z^2 + 16z63 - 4z^4}{2(18 - 33z + 20z^2 - 4z^3)} = \frac{\sum_{i=0}^{n} a_i z^i}{\sum_{j=0}^{d} b_j z^j}.$$

Thus,

$$\begin{aligned} p_0 &= \frac{a_0}{b_0} = \frac{66}{12 \times 18} = \frac{11}{36} \\ p_1 &= \frac{a_1 - b_1 p_0}{b_0} = \frac{-69 + 12 \times 33\frac{11}{36}}{12 \times 18} = \frac{13}{54} \\ p_2 &= \frac{a - 2 - b_1 p_1 - b_2 p0}{b_0} = \frac{25}{216}, \text{etc..} \end{aligned}$$

## 9.2 Laplace Transform Method

**9.17 Derivation of the Erlang distribution.**

(a)

$$\Phi_Y(s) = \frac{n\lambda}{s+n\lambda}, \text{ and } \Phi_X(s) = (\Phi_y(s))^n = \left(\frac{n\lambda}{s+n\lambda}\right)^n.$$

(b) Using (9.17), we find

$$f_X(x) = \frac{(n\lambda)^n x^{n-1}}{(n-1)!} e^{-n\lambda x}, \;\; x \geq 0.$$

(c)

$$\ln \Phi_X(s) = n \ln n\lambda - n \ln(s+n\lambda).$$

Thus,

$$\begin{aligned} \frac{\Phi_X'(s)}{\Phi_X(s)} &= -\frac{n}{s+n\lambda} \\ \frac{\Phi_X''(s) - (\Phi_x'(s))^2}{\Phi_X(s)^2} &= \frac{n}{(s+n\lambda)^2}. \end{aligned}$$

Hence,

$$E[X] = \frac{1}{\lambda}, \text{ and } \mathrm{Var}[X] = \Phi''(0) - \Phi'(0)^2 = \frac{1}{n\lambda^2}.$$

**9.18* Convolution and the Laplace transform.**

$$\Phi_g(s) = \int_0^\infty g(x) e^{-sx}\, dx = \int_0^\infty \left(\int_0^x f_1(x-y) f_2(y)\, dy\right) e^{-sx}\, dx.$$

Define a new variable $z = x - y$, then $0 \leq z < \infty$, because $0 \leq y \leq x$. Then,

$$\Phi_g(s) = \int_0^\infty f_1(z) e^{-sz}\, dz \int_0^\infty f_2(y) e^{-sy}\, dy = \Phi_{f_1}(s)\Phi_{f_2}(s).$$

**9.19 Laplace transforms of the distribution function and survivor function.**

(a)

$$\Phi(s) = \int_0^\infty e^{-sx} f(x)\, dx = \int_0^\infty e^{-sx} F'(x) dx = \left[e^{-sx}F(x)\right]_0^\infty + \int_0^\infty s e^{-sx} F(x)\, dx.$$

The first term of the last equation is zero, so we have

$$\int_0^\infty e^{-sx} F(x)\, dx = \frac{\Phi(s)}{s}.$$

(b)

$$\begin{aligned}\int_0^\infty e^{-sx}[1-F(x)]\,dx &= \int_0^\infty \frac{partial}{\partial x}\left(-\frac{e^{-sx}}{s}\right)[1-F(x)]\,dx \\ &= -\left[\frac{e^{-sx}}{s}[1-F(x)]\right]_0^\infty - \frac{1}{s}\int_0^\infty e^{-sx}f(x)\,dx \\ &= \frac{1}{s} - \frac{\Phi(s)}{s}.\end{aligned}$$

**9.20 The $n$-fold convolutions of the uniform distribution**

$$f_X(x) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{elsewhere} \end{cases}$$

$$\Phi_X(s) = \int_0^1 e^{-sx}\,dx = \frac{1-e^{-s}}{s}$$

$$\Phi_Y(s) = \left(\frac{1-e^{-s}}{s}\right)^n = \sum_{k=0}^n (-1)^k \binom{n}{k} s^{-n} e^{-ks}$$

Using the hint, we find

$$f_Y(y) = \sum_{k=0}^n (-1)^k \binom{n}{k} \frac{[y-k]_+^{n-1}}{(n-1)!}.$$

**9.21* Discontinuities in a distribution function.** If $F_X(x)$ has a discontinuity only at $x = 0$, the corresponding $\Phi_X(s)$ is a rational function of the form (9.97). Then it is clear from (9.102) and (9.106) that

$$\lim_{s\to\infty} \Phi_X(s) = \lim_{x\to 0^+} F_X(x) = \frac{a_d}{b_d},$$

which is the magnitude of a jump in $F_X(x)$ at the origin.
If $F_X(x)$ contains discontinuities of $p_k$'s at $x = x_k$'s, we can write

$$F_X(x) = \sum_k p_k u(x - x_k) + G(x),$$

where $u(x)$ is the unit step function, and $G(x)$ is a continuous and piecewise differentiable function with $g(x) = G'(x)$. Then the corresponding density function is

$$f_X(x) = \sum_k p_k \delta(x - x_k) + g(x),$$

where $\delta(x)$ is Dirac's delta function or the impulse function. The Laplace transform is then

$$\Phi_X(s) = \sum_k p_k e^{-sx_k} + \int g(x) e^{-sx}\,dx.$$

**9.22 Derivation of the cosine transform of (9.110).** By setting $= c + i\omega$, the inverse transform (9.95) become

$$\begin{aligned} f_X(x) &= \frac{e^{cx}}{2\pi}\int_{-\infty}^{\infty} \Phi(c+i\omega)e^{i\omega x}\,d\omega \\ &= \frac{e^{cx}}{2\pi}\int_{-\infty}^{\infty} [\Re\{\Phi(c+i\omega)\}\cos\omega x - \Im\{\Phi(c+i\omega)\}\sin\omega x]\,d\omega \\ &+ i\frac{e^{cx}}{2\pi}\int_{-\infty}^{\infty} [\Re\{\Phi(c+i\omega)\}\sin\omega x + \Im\{\Phi(c+i\omega)\}\cos\omega x]\,d\omega, \quad -\infty < x < \infty. \end{aligned}$$

Since $f_X(x)$ is a real-valued function, $\Phi(c+i\omega)$ must satisfy

$$\int_{-\infty}^{\infty} [\Re\{\Phi(c+i\omega)\}\sin\omega x + \Im\{\Phi(c+i\omega)\}\cos\omega x]\,d\omega = 0, \quad -\infty < x < \infty.$$

Furthermore, for a nonnegative RV $X$, the PDF $f_X(x) = 0$ for $x < 0$. By setting $x = -x'$ $(x' > 0)$, we have

$$\int_{-\infty}^{\infty} [\Re\{\Phi(c+i\omega)\}\cos\omega x' + \Im\{\Phi(c+i\omega)\}\sin\omega x']\,d\omega = 0, \;\; x' > 0.$$

Then, for $x > 0$, we have

$$\begin{aligned} f_X(x) &= \frac{e^{ct}}{\pi}\int_{-\infty}^{\infty} \Re\{\Phi(c+i\omega)\}\cos\omega x\,d\omega \\ &= \frac{e^{ct}}{\pi}\left[\int_0^{\infty} \Re\{\Phi(c+i\omega)\}\cos\omega x + \int_{-\infty}^{0} \Re\{\Phi(c+i\omega)\}\cos\omega x\right]d\omega \\ &= \frac{2e^{ct}}{\pi}\int_0^{\infty} \Re\{\Phi(c+i\omega)\}\cos\omega x\,d\omega, \;\; x \geq 0, \end{aligned}$$

which is (9.110) that we are after.

**9.23 Trapezoidal approximation of an integration.** We write (9.110)

$$f_X(x) = \frac{2e^{cx}}{\pi}\int_0^{\infty} \cos(\omega x)\Re\{\Phi(c+i\omega)\}\,d\omega, \;\; x \geq 0.$$

We want to evaluate the above over a finite range $x \in [0, T)$ at regularly spaced points $x = 0, \frac{T}{N}, \frac{2T}{N}, \ldots, T$. First we approximate the integration over $\omega$ by the trapezoid method. By choosing the strip width $h = \frac{\pi}{2T}$, we obtain the approximation given by (9.111).
Evaluation at $x = j\delta = j\frac{T}{N}$ is given by

$$f_X(j\delta) \approx \frac{e^{jc\delta}}{T}\left[\frac{1}{2}\Re\{\Phi(c)\} + \sum_{n=1}^{\infty}\Re\left\{\Phi\left(c + i\frac{\pi n}{2T}\right)\right\}\cos\left(\frac{\pi jn}{2N}\right)\right].$$

The term $\cos\left(\frac{\pi jn}{2N}\right)$ is periodic in $n$ with period $4N$.

1. $\cos\left(\frac{\pi jn}{2N}\right)$ takes on the value $\cos 0 = 1$ at $n = 0, 4N, 8N, \ldots, 4mN, \ldots$ $(m \geq 1)$. Collecting these terms in the square bracket, we denote it as $\frac{1}{2}g_0$,

$$g_0 = \Re\{\Phi(c)\} + 2\sum_{m=1}^{\infty}\Re\left\{\Phi_X\left(c + i\frac{2\pi m}{\delta}\right)\right\}.$$

2. $\cos\left(\frac{\pi jn}{2N}\right)$ takes on the value $\cos j\pi = (-1)^j$ at $n = 2N, 6N, \ldots, 2N = 4mN, \ldots (m \geq 0)$. Collecting these terms in the square bracket, we denote it as $(-1)^j g_{2N}$:

$$g_N = \sum_{m=0}^{\infty} \Re\left\{\Phi_X\left[c + i\frac{2\pi}{\delta}\left(\frac{1}{2} + m\right)\right]\right\}.$$

3. Similarly, $\cos\left(\frac{\pi jn}{2N}\right)$ takes on the value $\cos\frac{\pi jk}{2N}$ at $n = k, 4N \pm k, \ldots, 4mN \pm k, \ldots (m \geq 1)$. Collecting these terms in the square bracket, we denote it as $g_k \cos\left(\frac{\pi jk}{2N}\right)$:

$$\begin{aligned} g_k = &\sum_{m=0}^{\infty} \Re\left\{\Phi_X\left[c + i\frac{2\pi}{\delta}\left(\frac{k}{4N} + m\right)\right]\right\} \\ &+ \sum_{m=0}^{\infty} \Re\left\{\Phi_X\left[c + i\frac{2\pi}{\delta}\left(1 - \frac{k}{4N} + m\right)\right]\right\}, k = 1, 2, \ldots, 2N-1. \end{aligned}$$

Treating Case 2 as a special case of Case 3, we can write somewhat concisely as in (9.114). Note: In (9.113), the summation should be from $m = 1$ (not $m = 0$) to $\infty$.

# 10 Solutions for Chapter 10: Inequalities, Bounds and Large Deviation Approximation

## 10.1 Inequalities frequently used in Probability Theory

**10.1 Lagrange identity.** We prove the identity formula by mathematical induction. It is straightforward to show the identity formula is valid for $n = 2$. Suppose that the formula is correct for $n = k - 1$, where $k \geq 3$. All we need to show that it is true for $n = k$, as well.
Define

$$\begin{aligned}
A(k) &\triangleq \sum_{i=1}^{k} |x_i|^2 = A(k-1) + |x_k|^2 \\
B(k) &\triangleq \sum_{i=1}^{k} |y_i|^2 = B(k-1) + |y_k|^2 \\
C(k) &\triangleq \sum_{i=1}^{k} x_i \overline{y}_i = C(k-1) + x_k \overline{y}_k \\
D(k) &\triangleq \sum_{i=1}^{k} \sum_{j=i+1}^{k} |x_i y_j - x_j y_i|^2 = D(k-1) + \sum_{i=1}^{k-1} |x_i y_k - x_k y_i|^2 \\
&= D(k-1) + A(k-1)|y_k|^2 + B(k-1)|x_k|^2 - x_n \overline{y}_k \overline{C}(k-1) - \overline{x}_k y_k C(k-1).
\end{aligned}$$

Then the LHS of (**??**) for $n = k$ is

$$\begin{aligned}
\text{LHS}(k) &= A(k)B(k) - C(k)\overline{C}(k) = [A(k-1) + |x_k|^2][B(k-1) + |y_k|^2] - x_k \overline{y}_k \overline{C}(k-1) - \overline{x}_k y_k C(k-1) \\
&= A(k-1)B(k-1) + A(k-1)|y_k|^2 + B(k-1)|x_k|^2 + |x_k|^2 |y_K|^2 - x_k \overline{y}_k \overline{C}(k-1) - \overline{x}_k y_k C(k-1),
\end{aligned}$$

and the RHS is

$$\text{RHS}(k) = D(k-1) + A(k-1)|y_k|^2 + B(k-1)|x_k|^2 - x_n \overline{y}_k \overline{C}(k-1) - \overline{x}_k y_k C(k-1).$$

By using LHS($k-1$)=RHS($k-1$), where

$$\begin{aligned}
\text{LHS}(k-1) &= A(k-1)B(k-1) - C(k-1)\overline{C}(k-1) \\
\text{RHS}(k-1) &= D(k-1),
\end{aligned}$$

we can show LHS($k$)=RHS($k$).

**10.2 Proof of Jensen's inequality**
Suppose that (10.15) is true for $n = k \geq 2$, then we need to show it for $n = k + 1$. We assume $p_1 < 1$, without loss of generality: otherwise $p_i = 0$ for all $i \geq 2$, and the problem becomes

meaningless. If we write $\frac{p_i}{1-p_1} = \beta_i, \quad i = 2, 3, \ldots, k$, then $\beta_i \geq 0$, and $\sum_{i=2}^{k+1} = 1$. We then write

$$g\left(\sum_i p_i x_{i=1}^{k+1}\right) = g\left(p_1 x_1 + (1-p_1)\sum_{i=2}^{k+1} \beta_i x_i\right) \leq p_1 g(x_1) + (1-p_1) g\left(\sum_{i=2}^{k+1} \beta_i x_i\right)$$
$$\leq p_1 g(x_1) + (1-p_1)\left(\sum_{i=2}^{k+1} \beta_i g(x_i)\right) = \sum_{i=1}^{k+1} p_i g(x_i), \quad (1)$$

where the first inequality is due to the definition of convex function (10.14) defined for two points $x_1$ and $x_2' = \sum_{i=2}^{k+1} \beta_i x_i$, and the second inequality is due to the assumption that (10.15) holds for $k$ points $x_2, \ldots, x_{k+1}$. Thus, we have proved that the inequality (10.15) holds for $n = k = 1$ as well.

**10.3 Inequality for covariance.** Let $E[X] = \mu_X, E[Y] = \mu_Y$.

$$|\mathrm{Cov}[X,Y]|^2 = |E[(X-\mu_X)(\overline{Y} - \overline{\mu}_Y)]|^2 = |\langle X - \mu_X, Y - \mu_Y \rangle|^2$$
$$\leq \|X - \mu_X\|^2 \|Y - \mu_Y\|^2 = E\left[|X - \mu_X|^2\right] E\left[|Y - \mu_Y|^2\right] = \mathrm{Var}[X]\mathrm{Var}[Y].$$

**10.4 Arithmetic mean and geometric mean.** If any one of $x_i$ is zero, the inequality obviously holds. So we assume that all $x_i$'s are positive below. Since $g(x) = \log x, \quad x > 0$ is a concave function, we have

$$g(E[X]) \geq E[g(X)],$$

or

$$\log(p_1 x_1 + p_2 x_2 + \ldots + p_n x_n) \geq \sum_i p_i \log x_i.$$

If we choose $p_i = 1/n$ for all $i$, then

$$\log\left(\frac{\sum_i x_i}{n}\right) \geq \frac{1}{n}\sum_i \log x_i = \frac{1}{n}\log\prod_i x_i = \log\left(\prod_i x_i\right)^{1/n}.$$

If we take the exponentiation of both sides, we arrive at (10.128).

**10.5 Convex function is continuous.**

(a)

$$p = \frac{x-2-x_1}{x-2-x_0}, \text{“}1 - p = \frac{x_1 - x_0}{x_2 - x_0}.$$

(b) Because $g(x)$ is convex, we have using the result of (a)

$$\frac{x_2 - x_1}{x_2 - x_0} g(x_0) + \frac{x_1 - x_0}{x_2 - x_0} g(x_2) \geq g(x_1),$$

where

$$x_1 = \frac{x_2 - x_1}{x_2 - x_0} x_0 + \frac{x_1 - x_0}{x_2 - x_0} x_2.$$

Then by rearranging the above we can show

$$\frac{g(x_1)-g(x_0)}{x_1-x_0} \leq \frac{g(x_2)-g(x_0)}{x_2-x_0}.$$

We can also obtain in a similar fashion

$$\frac{g(x_0)-g(x_{-1})}{x_0-x_{-1}} \leq \frac{g(x_1)-g(x_0)}{x_1-x_0}.$$

As $x \downarrow x_0$, i.e., approaches $x_0$ from the right, the slope $\frac{g(x)-g(x_0)}{x-x_0}$ monotonically decreases, but is bounded from below by $\frac{g(x_0)-g(x_{-1})}{x_0-x_{-1}}$. Thus the limit $f'_+(x_0)$ exists. Similarly, the LHS derivative $f'_-(x_0)$ exists, and obviously $g'_-(x_0) \leq g'_+(x_0)$.

(c) Since the derivatives exist at all points, $g(x)$ must be continuous at all point. If there were a discontinuity, then the derivative cannot be finite.

**10.6 A convex function is above its tangent.** By referring to the result of Problem 10.5, $a$ be such that

$$g'_-(x_0) \leq a \leq g'_+(x_0).$$

Then we can claim (10.129), i.e.,

$$g(x) \geq g(x_0) + a(x-x_0) \text{ for all } x,$$

for if $x > x_0$, we have from the result of Problem 10.5,

$$\frac{g(x)-g(x_0)}{x-x_0} \geq g'_+(x_0) \geq a.$$

and for $x < x_0$,

$$\frac{g(x)-g(x_0)}{x-x_0} \leq g'_-(x_0) \leq a,$$

which leads also to (10.129), by noting $x - x_0 < 0$.

**10.7 A twice-differentiable function, and a convex function** By expanding the observations we made in Problem 10.5, if the second derivatives are all positive, it follows that for the points $x_0 < x_1 < x_2$, we will have

$$\frac{g(x_1)-g(x_0)}{x_1-x_0} \leq \frac{g(x_2)-g(x_1)}{x_2-x_1},$$

which leads to

$$(x_2-x_0)g(x_1) \leq (x_2-x_1)g(x_0) + (x_1-x_0)g(x_2),$$

which implies that $g(x)$ is convex. The reverse argument holds if the second derivatives exist. Thus $g''(x) \geq 0$ is the necessary and sufficient condition for convexity of functions that are continuous and possess second derivatives.

**10.8 Another derivation of Jensen's inequality** Apply Taylor' series expansion to $g(x)$ at $x = \mu_X = E[X]$:

$$g(x) = g(\mu_X) + g'(\mu_X)(x-\mu_X) + \frac{g''(\xi)(x-\mu_X)^2}{2},$$

where $\xi$ is some value between $x$ and $\mu_X$. Since $g''(\xi) \geq 0$. we have

$$g(x) \geq g(\mu_X) + g'(\mu_X)(x - \mu_X).$$

Thus,

$$g(X) \geq g(\mu_X) + g'(\mu_X)(X - \mu_X).$$

Taking the expectation, we find

$$E[g(X)] \geq g(E[X]) + g'(\mu_X)E[X - \mu_X] = g(E[X]),$$

which is Jensen's inequality.

**10.9 Alternative derivation of Shannon's lemma.**

(a) Since $-\log x$ is convex, we have from Jensen's inequality

$$E[-\log X] \geq -\log E[X].$$

Assume the RV $X$ that takes on values $x_i = \frac{g_i}{f_i}$ with probability $f_i$. Then

$$-\sum_{i=1}^{n} f_i \log \frac{g_i}{f_i} \geq -\log \sum_{i=1}^{n} f_i \frac{g_i}{f_i} = -\log 1 = 0.$$

Thus,

$$\sum_{i=1}^{n} f_i \log \frac{f_i}{g_i} \geq 0.$$

(b) Let us assume 's are fixed and we find $\boldsymbol{g}$'s that maximizes $\sum_i f_i g_i$ under the constraint $\sum g_i = 1$, namely

$$J(\boldsymbol{g}) = \sum_i f_i \log g_i - \lambda(\sum_i g_i - 1).$$

Differentiate $J$ with respect to $g_i$ and set it to zero.

$$\frac{\partial J}{\partial g_i} = \frac{f_i}{g_i} - \lambda = 0, \quad \text{for all} \;\; i. \tag{2}$$

Then $f_i = \lambda g_i$. By summing this over $i$, we find $\lambda = 1$, and $\boldsymbol{g} =$ gives

$$J() = \sum_{\mathbf{i}} \mathbf{f_i} \log \mathbf{f_i}.$$

Since

$$H_{ij} \triangleq \frac{\partial^2 J}{\partial g_i \partial g_j} = -\frac{f_i}{g_i^2}\delta_{i,j},$$

the Hessian[1]matrix is negative definite, the stationary point determined by (2) is a maximal point.

**10.10 Inequalities in information theory [278]**

[1] named after Ludwig Otto Hesse (1811 1874), a German mathematician.

(a) It can be proved in the same way as Shannon's lemma was derived. Or set $a = b$ in the log sum inequality.

(b) Taylor's expansion of $\log x$ at $x = 1$ gives

$$\log x = (x-1) - \frac{(x-1)^2}{2\xi^2},$$

where $\xi$ is some value between 1 and $x$. Set $x = \frac{b_i}{a_i}$ in the above:

$$\log \frac{b_i}{a_i} = \left(\frac{b_i}{a_i} - 1\right) - \frac{\left(\frac{b_i}{a_i} - 1\right)^2}{2\xi_i^2}.$$

Multiply both sides by $a_i$ and sum them over $i$, obtaining

$$\sum_i a_i \log \frac{b_i}{a_i} = \sum_i b_i - \sum_i a_i - \sum_i a_i \frac{\left(\frac{b_i}{a_i} - 1\right)^2}{2\eta_i^2} \leq 0,$$

which also proves (10.130).

In the last expression $\eta_i$ is between $a_i$ and $b_i$. If $a_i, b_i \leq 1$, the maximum value of $\eta_i^2$ is not greater than one. Thus,

$$-\sum_i a_i \frac{\left(\frac{b_i}{a_i} - 1\right)^2}{2\eta_i^2} \leq -\frac{\sum_i a_i (b_i - a_i)^2}{2},$$

Combining the last two inequalities, we obtain (10.131).

**10.11 Coin tossing and Markov and Chebyshev inequalities**

(a) Consider Markov's inequality:

$$P[S_n \geq n\beta] \leq \frac{E[S_n]}{n\beta} = \frac{n/2}{n\beta} = \frac{1}{2\beta}. \tag{3}$$

This upper-bound is independent of $n$, hence remains a loose bound. For $\beta = 0.8$, the upper bound is $\frac{1}{1.6} = 0.625$ for all $n$.

(b) Writing $b = n\beta = \frac{n}{2} + \left(\beta - \frac{1}{2}\right) n = E[S_n] + \left(\beta - \frac{1}{2}\right) n$, we obtain

$$\begin{aligned} P[S_n \geq n\beta] &= P\left[S_n - E[S_n] \geq \left(\beta - \frac{1}{2}\right) n\right] \leq P\left[|S_n - E[S_n]| \geq \left(\beta - \frac{1}{2}\right) n\right] \\ &\leq \frac{\sigma_{S_n}^2}{\left[\left(\beta - \frac{1}{2}\right) n\right]^2} = \frac{n/4}{\left(\beta - \frac{1}{2}\right)^2 n^2} = \frac{1}{4n\left(\beta - \frac{1}{2}\right)^2}, \quad \frac{1}{2} < \beta < 1. \end{aligned} \tag{4}$$

Thus, the probability that $S_n$ exceeds $b = n\beta$ approaches zero as $n$ increases. This is a special case of (10.30), in which we set $p = \frac{1}{2}$ and $\epsilon = (\beta - \frac{1}{2})n$ For $\beta = 0.8$, the upper bound is $\frac{1}{0.36n}$, which is $2.78 \times 10^{-2}$ for $n = 100$, and $2.78 \times 10^{-3}$ for $n = 1000$. Compared with the Chernoff bound discussed in Example 10.4, these upper bound by Chebyshev's inequality are too loose.

**10.12 Bienaymé's inequality.**

(a) The proof is essentially the same as that for Chebyshev's inequality, which is a special case of (10.134) obtained with $r = 2$. Let $Y = |X|^r$ and $a = b^r$ in Markov' inequality (10.23). Then inequality (10.133) is readily obtained.

(b) Similarly, substitution of $Y = |X - E[X]|^r$ and $a = b^r$ into Markov's inequality results in (10.134). Some authors call this inequality Bienaymé-Chebyshev's theorem (see e.g. [318]).

**10.13 Markov-Chebyshev-Bienaymé's inequality.**
We set $Y = g(|X|)$ and $a = g(b)$ in Markov's inequality. Then using the relation

$$|X| \geq b \iff g(|X|) \geq g(b),$$

for any increasing non-negative function $g(x)$, the generalized inequality (10.135) is obtained. An alternative proof of the above inequality is, in referring to Figure 10.1, to use the following inequality:

$$\mathbf{1}_{|x|\geq a} \leq \frac{g(|x|)}{g(a)} \tag{5}$$

where

$$\mathbf{1}_{|x|\geq a} = u(-x-a) + u(x-a) = \begin{cases} 1, & x \geq a, \\ 0, & x \leq -a. \end{cases} \tag{6}$$

By multiplying both sides of (5) by $f_X(x)$ and integrating, we find

$$P[|x| \geq a] \leq \frac{1}{g(a)} \int_{-\infty}^{\infty} g(|x|) f_X(x)\, dx \tag{7}$$

yielding the inequality (10.135).

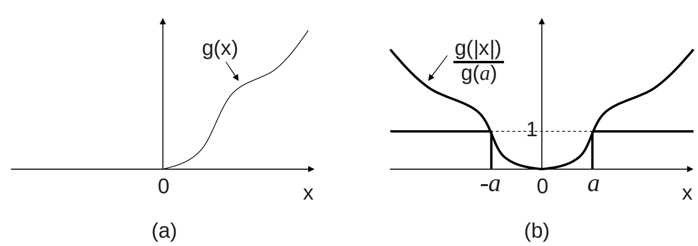



**Figure 10.1** (a) An increasing non-negative function $g(x)$; (b) $g(|x|)/g(a)$ and $1_{|x|\geq a}$.

**10.14 One-sided Chebyshev's inequality.**

(a) Let $b > 0$. Hence

$$\begin{aligned} P[X \geq a] &= P[X + b \geq a + b] = P[(X+b)^2 \geq (a+b)^2] \\ &\leq \frac{E[(X+b)^2]}{(a+b)^2} = \frac{\sigma^2 + b^2}{(a+b)^2}. \end{aligned}$$

The minimum of the right side is obtained when $b = \sigma^2/a$.

$$P[X \geq a] \leq \frac{\sigma^2}{\sigma^2 + a^2}.$$

(b) $X-\mu$ has mean zero and variance $\sigma^2$, we obtain the first inequality from part (a). Similarly $\mu - X$ has mean zero and variance $\sigma^2$ and we obtain the second inequality.

**10.15 Submartingale derived from a martingale**

(a) For any given non-negative convex function $g(x)$, we define $Y_k = g(S_k)$, where $\{S_k\}$ is a martingale. Then

$$\begin{aligned} E[Y_k|Y_1, Y_2, \dots, Y_{k-1}] &= E[g(S_k)|S_1, S_2, \dots, S_{k-1}] \\ &\geq g\left(E[S_k|S_1, S_2, \dots, S_{k-1}]\right) = g(S_{k-1}) = Y_{k-1}. \end{aligned}$$

which shows $Y_k$ is a submartingale. Since $g(x) = |x|$ is a convex function, $Y_k = |S_k|$ is a submartingale.

(b) We can generalize the proof of Theorem 10.7

Defining $\xi$ as in the proof of the text, we can write for any $p \geq 1$

$$E[Y_n^p] = \sum_{j=0}^{n} E[Y_n^p|\xi = j]P[\xi = j] \geq \sum_{j=1}^{n} E[Y_n^p|\xi = j]P[\xi = j]. \tag{8}$$

For $j \geq 1$ the event $\{\xi = j\}$ depends only on $Y_1, Y_2, \dots, Y_j$, and therefore

$$E[Y_n^p|\xi = j] = E[Y_n^p|Y_1, Y_2, \cdots, Y_j] \geq Y_j^p \geq a^p. \tag{9}$$

By substituting (9) into (8), it is evident that

$$\sum_{j=1}^{n} P[\xi = j] \leq \frac{E[Y_n^p]}{a^p}, \;\; p \geq 1. \tag{10}$$

The LHS can be written as

$$\begin{aligned} \sum_{j=1}^{n} P[\xi = j] &= P[Y_j \geq a \text{ for some } j = 1, 2, \dots, n.] \\ &= P[\max\{Y_1, Y_2, \cdots, Y_n\} \geq a]. \end{aligned}$$

Thus, we have proved the inequality (10.41).

**10.16* Bernstein's inequality [21, 131].**

(a)

$$
\begin{aligned}
P\left[\frac{S_n}{n} - p \geq \epsilon\right] &= \sum_{k \geq n(p+\epsilon)} P[S_n = k] = \sum_{k=m}^{n} \binom{n}{k} p^k q^{n-k} \\
&\leq \exp\{\lambda[k - n(p+\epsilon)]\} \binom{n}{k} p^k q^{n-k} \\
&= e^{-\lambda n \epsilon} \sum_{k=m}^{n} \binom{n}{k} \left(pe^{\lambda q}\right)^k \left(qe^{-\lambda p}\right)^{n-k} \\
&\leq e^{-\lambda n \epsilon} \sum_{k=0}^{n} \binom{n}{k} \left(pe^{\lambda q}\right)^k \left(qe^{-\lambda p}\right)^{n-k} \\
&= e^{-\lambda n \epsilon} \left(pe^{\lambda q} + qe^{-\lambda p}\right)^n
\end{aligned}
$$

(b) Using the inequality in the hint

$$
e^{\lambda q} \leq \lambda q + e^{\lambda^2 q^2}, \text{ and } e^{-\lambda p} \leq -\lambda q + e^{\lambda^2 p^2}.
$$

Thus,

$$
pe^{\lambda q} + qe^{-\lambda p} \leq e^{\lambda^2 q^2} + e^{\lambda^2 p^2}.
$$

Thus,

$$
\begin{aligned}
P\left[\frac{S_n}{n} - p \geq \epsilon\right] &\leq e^{-\lambda n \epsilon} \left(e^{\lambda^2 q^2} + e^{\lambda^2 p^2}\right)^n \\
&\leq e^{-\lambda n \epsilon} \left(pe^{\lambda^2} + qe^{\lambda^2}\right)^n = \exp(-n\lambda(\epsilon - \lambda)).
\end{aligned}
$$

(c) Since

$$
\lambda(\epsilon - \lambda) \leq \frac{\lambda^2}{4},
$$

We finally find

$$
P\left[\frac{S_n}{n} - p \geq \epsilon\right] \leq \exp\left(-\frac{n\epsilon^2}{3}\right), \ \ \epsilon > 0.
$$

Since the distribution of $\frac{S_n}{n}$ should be symmetric around $p$, we have

$$
P\left[\frac{S_n}{n} - p \leq -\epsilon\right] \leq \exp\left(-\frac{n\epsilon^2}{3}\right).
$$

Thus combining the above we obtain **Bernstein's inequality** (10.136).

**10.17* Hoeffding's inequality for a martingale [152, 288].**

(a) Suppose $\mu = 0$. Then for any $\lambda > 0$, we have from Markov inequality

$$
P[Y_n \geq t] = P\left[\exp(\lambda Y_n) \geq e^{\lambda t}\right] \leq e^{-\lambda t} E[\exp(\lambda Y_n)].
$$

Let $W_n = \exp(\lambda Y_n)$ Then $W_0 = e^{\mu} = 1$, and

$$
W_n = e^{\lambda Y_{n-1}} e^{\lambda(Y_n - Y_{n-1})}.
$$

Thus,

$$E[W_n|Y_{n-1}] = e^{\lambda Y_{n-1}} E\left[e^{\lambda(Y_n - Y_{n-1})}|Y_{n-1}\right]$$
$$\leq W_{n-1} \frac{b_n e^{-\lambda a_n} + a_n e^{\lambda b_n}}{a_n + b_n},$$

where we used the hint since $f(x) = e^{ix}$ is a convex function, and that $X = Y_n - Y_{n-1}$ satisfies $E[X] = 0$, because $E[Y_n - Y_{n-1}|Y_{n-1}] = E[Y_n|Y_{n-1}] - E[Y_{n-1}|Y_{n-1}] = Y_{n-1} - y_{n-1} = 0$.

(b) Taking the expectation of the above,

$$E[W_n] \leq E[W_{n-1}] \frac{b_n e^{-\lambda a_n} + a_n e^{\lambda b_n}}{a_n + b_n},$$

which lead to

$$E[W_n] \leq \prod_{i=1}^{n} \frac{b_i e^{-\lambda a_i} + a_i e^{\lambda b_i}}{a_i + b_i}.$$

Then from (10.139)

$$P[Y_n \geq t] \leq e^{-\lambda t} \prod_{i=1}^{n} \frac{b_i e^{-\lambda a_i} + a_i e^{\lambda b_i}}{a_i + b_i} \leq e^{-\lambda t} \prod_{i=1}^{n} \exp\left(\frac{\lambda^2 (a_i + b_i)^2}{8}\right),$$

where we set $\theta = \frac{a_i}{a_i + b_i}$ and $x = \lambda(a_i + b_i)$ in the hint. Hence,

$$P[Y_n \geq t] \leq \exp\left[\lambda\left(\lambda \frac{\sum_{i=1}^{n}(a_i + b_i)^2}{8} - t\right)\right],$$

(c) The expression in [ ] takes the minimum when $\lambda = \frac{4t}{\sum_{i=1}^{n}(a_i + b_i)^2}$, and we have

$$P[Y_n \geq t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(a_i + b_i)^2}\right).$$

Then the Azuma-Hoeffding inequalities (10.137) and (10.138) follow by applying the above to the zero-mean martingale $Y_i - \mu$ and to a zero-mean martingale $\mu - Y_i$, respectively.

**10.18* Upper bound on the waiting time in a G/G/1 queuing system [196].**

(a) By expanding (10.143) recursively, we have

$$W_n = \max\{0, X_{n-1}, X_{n-1} + X_{n-2}, \ldots, X_{n-1} + X_{n-2} + \ldots + X_0\}.$$

For $\theta > 0$, $e^{\theta W_n}$ is a monotone increasing function of $W_n$. Hence,

$$e^{\theta W_n} = \max\left\{1, e^{\theta X_{n-1}}, e^{\theta(X_{n-1}+X_{n-2})}, \ldots, e^{\theta(X_{n-1}+X_{n-2}+\ldots+X_0)}\right\}$$
$$\max\left\{Y_0, Y_1, \ldots, Y_n\right\}.$$

(b)

$$M_X(\theta) = E\left[e^{\theta X_n}\right].$$

The MGF is defined over an interval $I_\theta$, in which the MGF is bounded. This domain $I_\theta$ includes $\theta = 0$. The function $M_X(\theta)$ is a convex function. Furthermore, $M_X(0) = 1$ and $M'_X(0) = E[X] < 0$. Let $\theta > 0$ be any value in $I_\theta$ that satisfies

$$M_X(\theta) \geq 1. \tag{11}$$

Then

$$\begin{aligned} E[Y_n|Y_1, Y_2, \dots, Y_{n-1}] &= E\left[e^{\theta(X_{n-1}+X_{n-2}+\cdots+X_0)}\middle|e^{\theta(X_{n-1}+X_{n-2})}, \dots, e^{\theta(X_{n-1}+X_{n-2}+\cdots+X_1)}\right] \\ &= e\left[e^{\theta X_0}\right] e^{\theta(X_{n-1}+X_{n-2}+\cdots+X_1)} = M_X(\theta)Y_{n-1} \geq Y_{n-1}, \end{aligned}$$

which shows that $Y_n$ is a submartingale.

(c) By applying the Doob-Kolmorogov' inequlaity, we obtain

$$\begin{aligned} F^c_{W_n}(t) = P[W_n > t] &= P\left[e^{\theta W_n \geq e^{\theta t}}\right] \\ &= P\left[\max\{Y_0, Y_1, \dots, Y_n\} \geq e^{\theta t}\right] \\ &\leq \frac{P[Y_n]}{e^{\theta t}} = e^{-\theta t + n m_X(\theta)}, \end{aligned}$$

(c) The tightest bound is attained by finding

$$\min\left\{m_X(\theta) - \frac{t\theta}{n}\right\}$$

with the constraint (11), or equivalently

$$m_X(\theta) \geq 0. \tag{12}$$

## 10.2 Chernoff's bound

**10.19 Altenative derivation of (10.47).**

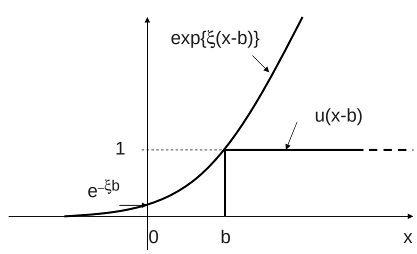



**Figure 10.2** An exponential function $e^{\xi(x-b)}$ vs. $1_{x\geq b} = u(x-b)$.

Consider the inequality between the exponential function $e^{\xi(x-b)}$ ($\xi \geq 0$) and the unit-step function $\mathbf{1}_{x\geq b} = u(x-b)$, as sketched in Figure 10.2:

$$\mathbf{1}_{x\geq b} \leq e^{\xi(x-b)} \text{ for } -\infty < x < \infty. \tag{13}$$

Multiplying both sides by $f_X(x)$ and integrating, the inequality (10.47) ensues. Note that the intercept of the exponential function at $x = 0$ in Figure 10.2 is $e^{-\xi b}$.

**10.20 Derivation of the Chernoff bound (10.50)**

Consider a function

$$U(\xi) = e^{-\xi b} M_X(\xi). \tag{14}$$

Taking its derivative, we have

$$U'(\xi) = e^{-\xi b}[M'_X(\xi) - bM_X(\xi)], \tag{15}$$

By setting (15) to be zero, we have

$$b = \frac{M'_X(\xi)}{M_X(\xi)} = \frac{d(\ln M_X(\xi)))}{d\xi}. \tag{16}$$

In order to show the solution $\xi^*$ of the above equation provides the **minimum**, it suffices to show that $U(\xi)$ is a convex $\cup$ function. By taking the second derivative, we obtain

$$U''(\xi) = e^{-\xi b}[b^2 M_X(\xi) - 2bM'_X(\xi) + M''_X(\xi)], \tag{17}$$

Then noting

$$M_X(\xi) = E[e^{\xi X}], \;\; M'_X(\xi) = E[Xe^{\xi X}], \;\; M''_X(\xi) = E[X^2 e^{\xi X}], \tag{18}$$

we can show that

$$U''(\xi) = e^{-\xi b} E[(b - X)^2 e^{\xi X}] \geq 0. \tag{19}$$

Thus, $U(\xi)$ is a convex function, and takes its the minimum value at $\xi = \xi^*$.

**10.21 Alternative proof of Chernoff's bound.**

(a) For a real valued function $g(x) = ax^2 + bx + c$, it is nonnegative for all $x$ if and only if $a \geq 0$ and $D = b^2 - 4ac \leq 0$. Using this result to $f(b, \xi)$, the necessary and sufficient condition is

$$M_X(\xi) \geq 0, \;\text{ and }\; D = 4(M'_X(\xi)^2 - M_X(\xi)M''_X(\xi)) \leq 0.$$

(b) From (18) in the solution of Problem **??**, the inequality $D/4 \leq 0$ can be written as

$$E\left[Xe^{\xi X}\right]^2 \leq E\left[e^{\xi X}\right] E\left[X^2 e^{\xi X}\right].$$

By defining

$$A = Xe^{\xi X/2}, \;\text{ and }\; B = e^{\xi X/2},$$

we have

$$E[AB]^2 \leq E[A^2]E[B^2],$$

which is the Cauchy-Schwarz inequality for RVs. The equality holds when $A = kB$ for some constant $k$. It then implies that $k = A/B = X$. Then, $M_X(\xi) = e^{k\xi}$ and $m_X(\xi) = k\xi$. Then (10.49) gives $k = b$. When $X = b$ constant, then the Chernoff bound will degenerate to $P[X \geq b] \leq e^{-\xi b} e^{\xi b} = 1$, and indeed the equality holds.

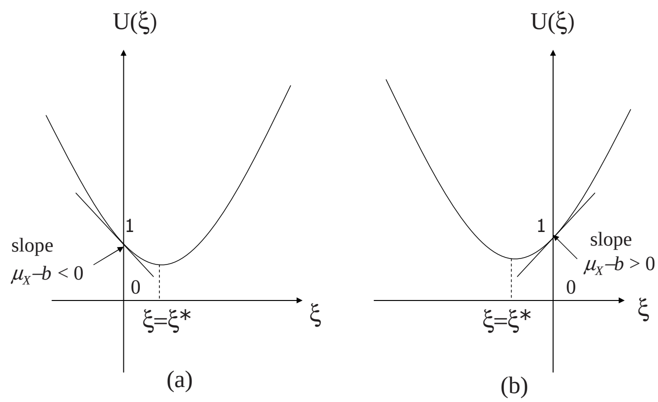



**Figure 10.3** The Chernoff's upper bound function $U(\xi) = e^{-\xi b} M_X(\xi)$ vs. $\xi$: (a) when $b > \mu_X$; (b) when $b < \mu_X$.

**10.22 When is the Chernoff bound meaningful?**
In Figure 10.3 we sketch the upper bound function $U(\xi) = e^{-\xi b} M_X(\xi)$ vs. $\xi$. The left curve (a) shows the case where $b > \mu_X = E[X]$. By setting $\xi = 0$ in (15), we find that $U'(0) = M'_X(0) - bM_X(0) = \mu_X - b < 0$, hence $U(\xi)$ intersects the vertical axis $\xi = 0$ with a negative slope and $U(0) = M_X(0) = 1$. The optimum value $\xi^*$ is positive, and $U(\xi^*)$ is indeed smaller than one. In the region $\xi < 0$, we have $U(\xi) > 1$, hence it is useless as an upper bound for the probability $P[X \geq b]$, which cannot be greater than one in any case.
If we choose the threshold parameter $b$ such that $b < \mu_X$, then the function $U(\xi)$ has a positive slope $\mu_X - b$ at $\xi = 0$, and the minimum of $U(\xi)$ occurs at $\xi = \xi^* < 0$, as is shown in the right curve (b). The bound (10.47) is useless, since it is greater than one in the region $\xi > 0$.

**10.23 Chernoff's bound for sum of normal RVs.**

$$M_U(\xi) = e^{\xi^2/2}, \;\; m_U(\xi) = \xi^2/2, \;\; m'_U(\xi) = \xi$$

So, $\xi^*$ that achieves a minimum is the root of

$$m'_U(\xi) = \xi = \frac{b}{n} = \beta, \;\; \text{i.e.,} \;\; \xi^* = \beta.$$

$$P[S \geq b] \leq e^{-n[\xi^* m'_U(\xi^*) - m_U(\xi^*)]} = e^{-n(b^2 - b^2/2)} = e^{-nb^2/2}, \;\; b \geq 0.$$

Similarly

$$P[S \leq b] \leq e^{-nb^2/2}, \;\; b < 0.$$

**10.24** 2. **Chernoff's bound for the sum of Poisson variables.**

$$M_X(\xi) = e^{\lambda(e^\xi - 1)}, \;\; m_X(\xi) = \lambda(e^\xi - 1), \;\; m'_X(\xi) = \lambda e^\xi.$$

So, $\xi^*$ is obtained from

$$\lambda e^{\xi^*} = \beta, \;\; \text{thatis} < \;\; \xi^* = \ln \beta - \ln \lambda.$$

So

$$P[S > b] \le e^{-n[(\ln\beta - \ln\lambda)\beta - (\beta-\lambda)]} = \left(\frac{\lambda}{\beta}\right)^{n\beta} e^{-n(\lambda-\beta)} = \frac{e^{-n\lambda}(\lambda e)^{n\beta}}{\beta^{n\beta}}.$$

Hence setting $n\beta = b$,

$$P[S > b] \le \frac{e^{-n\lambda}(\lambda e)^{b}}{\left(\frac{b}{n}\right)^{b}}.$$

**10.25 Numerical evaluation of the Chernoff bound.** For $\beta = 0.51$, we find $1 - \mathcal{H}(0.51) = 2.8856 \times 10^{-4}$, Thus, we find Chernoff's bound as follows for different numbers of tossing:

$$P[S_{100} \ge 51] \le 0.9802 \tag{20}$$

$$P[S_{1000} \ge 510] \le 0.8187 \tag{21}$$

$$P[S_{10^4} \ge 5,100] \le 2^{-2.8856} = 0.1353 \tag{22}$$

$$P[S_{10^5} \ge 51,000] \le 2^{-28.856} = 2.0584 \times 10^{-9} \tag{23}$$

$$P[S_{10^6} \ge 510,000] \le 2^{-288.56} = 1,3656 \times 10^{-87} \tag{24}$$

**10.26 Assessment of Chernoff's bound**

(a) The total number of distinct "head-tail" sequences is clearly $2^n$, and in fair coin tossing each sequence occurs with equal probability $2^{-n}$. Hence, by writing $b = \lceil n\beta \rceil$, we have

$$P[S_n \ge b] = 2^{-n} \sum_{k=b}^{n} \binom{n}{k} = 2^{-n} C_b, \tag{25}$$

where $C_b$ is the sum of the binomial coefficient from $k = b$ to $k = n$. Because the binomial coefficients for $k$ and $n - k$ are equivalent, we can write

$$C_b = \sum_{k=b}^{n} \binom{n}{k} = \sum_{k=0}^{n-b} \binom{n}{k} = 1 + n + \frac{n(n-1)}{2} + \ldots + \binom{n}{n-b}. \tag{26}$$

(b) For $n - b < \frac{n}{2}$, the last term $\binom{n}{n-b}$ is the largest term in the summation $C_b$. So if we discard all the other terms, we get a simple lower-bound of $C_b$. Using inequalities $\binom{n}{k} < \binom{n}{n-b}$ for all $k = 0, 1, 2, \ldots, n-b-1$, we can bound $C_b$ from the above by replacing each term in (**??**) by the largest, i.e., $\binom{n}{n-b} = \binom{n}{b}$. Thus we have (10.149). The lower-bound becomes a good and tight bound as $n$ increases, because among the $(n - b + 1)$ terms in the right side of (**??**), the last term $\binom{n}{b}$ becomes increasingly dominant, and the total sum of discarded terms becomes insignificant compared with this dominant term. The upper-bound of (10.149), on the other hand, is a loose bound, since we have replaced all the terms in (**??**) by $\binom{n}{b}$.

(c)

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

By substituting the Stirling formula for $n!$, $k!$ and $(n-k)!$, we readily find (10.151).

(d) Also straightforward.

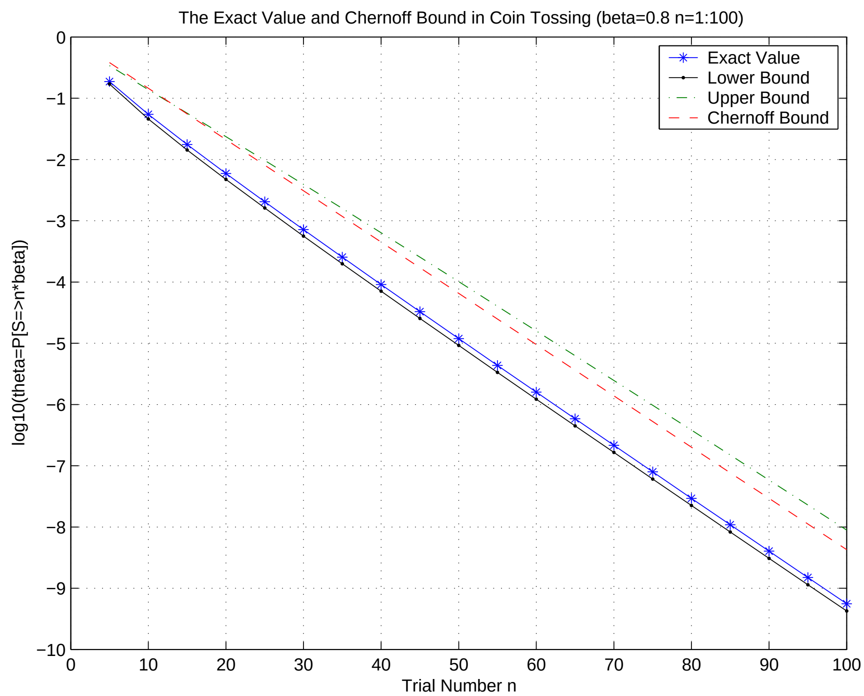



**Figure 10.4** The logarithm of $P[S_n \geq \lceil n\beta \rceil]$ in the coin tossing experiment ($\beta = 0.8$) as a function of $n$, and its bounds: the exact value, Chernoff's bound, and the upper and lower bounds based on Stirling's approximation.

(e) We plot the exact value, lower bound, upper bound and Chernoff bound in Figure **??**, for $\beta = 0.8$.

## 10.3 Large Deviation Theory

**10.27 Derivation of (10.82) and (10.83)**

Equation (10.82) can be derived as follows:

$$\begin{aligned} E[Y] &= \int_{-\infty}^{\infty} y \, dF_Y(y) = \int_{-\infty}^{\infty} \frac{y e^{\xi y} \, dF_X(y)}{M_X(\xi)} \\ &= \frac{1}{M_X(\xi)} \frac{d}{d\xi} \int_{-\infty}^{\infty} e^{\xi y} \, dF_X(y) = \frac{d}{d\xi} \ln M_X(\xi) \\ &= m'_X(\xi). \end{aligned} \tag{27}$$

A better approach is to compute the MGF of $Y$:

$$M_Y(t) = E[e^{tY}] = \int e^{ty} f_Y(y) \, dy = \int e^{ty} \frac{e^{\xi y} f_X(y)}{M_X(\xi)} \, dy = \frac{M_X(\xi + t)}{M_X(\xi)}.$$

Taking the logarithm

$$m_Y(t) = m_X(\xi + t) - m_X(\xi).$$

Then

$$E[Y] = m_Y'(0) = m_X'(\xi), \text{ and } \mathrm{Var}[Y] = m_Y''(0) = m_X''(\xi).$$

**10.28 Application of large deviation approximation to coin tossing.**

(a)

$$F_{B_i}(x) = \frac{1}{2}\left(u(x) + u(x-1)\right), \tag{28}$$

and

$$dF_{Y_i}(x) = \frac{e^{\xi x}}{M_{B_i}(\xi)} dF_{B_i}(x) = \frac{1}{2M_{B_i}(\xi)} e^{\xi x}\left(du(x) + du(x-1)\right), \tag{29}$$

where $u(x)$ is the unit-step function. Thus $du(x) = \delta(x)\,dx$, where $\delta(x)$ is Dirac's delta function. $M_{B_i}(\xi) = \frac{1}{2}\left(1 + e^{\xi}\right)$ of (10.66) into the above equation, we have

$$F_{Y_i}(x) = \frac{1}{1+e^{\xi}} u(x) + \frac{e^{\xi}}{1+e^{\xi}} u(x-1). \tag{30}$$

Thus, the RV $Y_i$ corresponds to an outcome of **unfair** coin tossing in which 1 appears with probability $p = \frac{e^{\xi}}{1+e^{\xi}}$ and 0 appears with probability $q = 1 - p = \frac{1}{1+e^{\xi}}$. Figure 10.5 (a) and (b) show the distribution functions $F_{B_i}(x)$ and $F_{Y_i}(x)$ of (29) and (30), respectively.

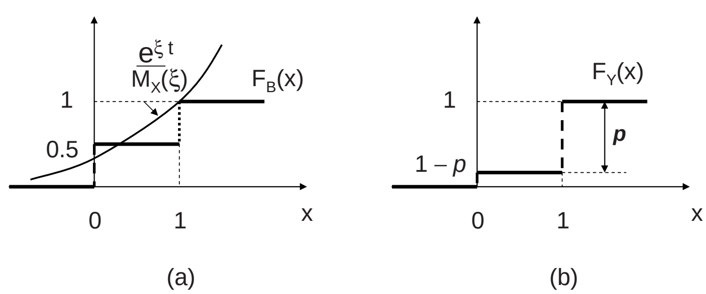



**Figure 10.5** (a) $F_{B_i}(x)$ of fair coin tossing and the exponential tilting function $e^{\xi x}/M_{B_i}(\xi)$; (b) The tilted distribution $F_{Y_i}(x)$: $P[Y_i = 1] = p = e^{\xi}/(1 + e^{\xi})$.

(b) The optimum tilting parameter $\xi^*$ is the same value as found earlier in obtaining the Chernoff bound, i.e.,

$$\xi^* = \ln\left(\frac{\beta}{1-\beta}\right). \tag{31}$$

With this optimum tilting, the unfair coin tossing will yield $Y_i = 1$ with probability $p^* = \frac{e^{\xi^*}}{1+e^{\xi^*}} = \beta$, for each $i = 1, 2, \ldots, n$.

(c) The mean and variance are given by

$$E[T_n] = n\beta, \text{ and } \mathrm{Var}\,[T_n] = n\beta(1-\beta). \tag{32}$$

An alternative derivation is to use the MGF of $T_n$, which can be found from the formula (10.88):

$$M_{T_n}(t) = \frac{M_{S_n}(\xi+t)}{M_{S_n}(\xi)} = \frac{\left(1+e^{\xi+t}\right)^n/2^n}{\left(1+e^{\xi}\right)^n/2^n} = \left(\frac{1+e^{\xi+t}}{1+e^{\xi}}\right)^n. \tag{33}$$

(d) By substituting(31) and

$$e^{-\xi^* b} = \left(\frac{1-\beta}{\beta}\right)^{n\beta}, \quad e^{m_{S_n}(\xi^*)} = \frac{\left(1+e^{\xi^*}\right)^n}{2^n} = 2^{-n}(1-\beta)^{-n}$$

into (10.101), we obtain (10.159). This approximation is more accurate for large $n$ than the Chernoff bound $2^{-n(1-\mathcal{H}(\beta))}$ in (10.72). It indeed lies between the upper and lower bounds obtained based on the refined Stirling formula discussed in Section **??**.

**Additional Notes: Numerical evaluation of the large deviation approximation to the coin tossing problem.**

Returning to the same numerical examples used in Example **??**, we find for $\beta = 0.8$:

$$\ln\left(\frac{\beta}{1-\beta}\right)\sqrt{2\pi n\beta(1-\beta)} = 1.3900\sqrt{n}.$$

Hence

$$P[S_{100} \geq 80] \approx \frac{1}{13.9}2^{-100\times 0.2781} = 3.057\times 10^{-10}(< 4.298\times 10^{-9}) \tag{34}$$

$$P[S_{1000} \geq 800] \approx \frac{1}{43.956}2^{-278.1} = 4.361\times 10^{-86}(< 1.9589\times 10^{-84}), \tag{35}$$

where the numbers in parentheses are the corresponding Chernoff bounds.
Similarly for $\beta = 0.51$ discussed in Problem 10.25, we have

$$\ln\left(\frac{\beta}{1-\beta}\right)\sqrt{2\pi n\beta(1-\beta)} = 0.050129\sqrt{n}.$$

Therefore, we find the large deviation approximations for the tail end of the distribution

$$P[S_{100} \geq 51] \approx (1/0.50129)\times 2^{-0.028856} = 1.9554(> 0.9802)$$
$$P[S_{1000} \geq 510] \approx (1/1.5852)\times 2^{-0.28856} = 0.5165(< 0.8187)$$
$$P[S_{10^4} \geq 5,100] \approx (1/5.0129)\times 2^{-2.8856} = 0.0270(< 0.1353)$$
$$P[S_{10^5} \geq 51,000] \approx (1/15.852)\times 2^{-28.856} = 1.2984\times 10^{-10}(< 2.0584\times 10^{-9})$$
$$P[S_{10^6} \geq 510,000] \approx (1/50.129)\times 2^{-288.56} = 2.7208\times 10^{-89}(< 1.3656\times 10^{-87})$$

where the numbers in parentheses are again the corresponding Chernoff's bounds. The meaningless result for $\beta = 0.51$ and $n = 100$ ($P[S_{100} \geq 51] \approx 1.9554$) is not due to large deviation approximation, but due to the fact the approximation (10.99) is not accurate when $c$ is small. For $\beta = 0.51$ and $n = 100$, we find $c = \ln\left(\frac{\beta}{1-\beta}\right)\sqrt{n\beta(1-\beta)} = 0.2000$. Then $A(\xi^*) \approx$

$e^{-\frac{c^2}{2}}Q(c) = 0.4124$. Then using this value, we have a much better approximation:

$$P[S_{100} \geq 51] \approx 0.4124 \times 2^{-0.028856} = 0.4042(< 0.9802).$$

We should note that an exact evaluation of $A(\xi)$ of (10.91) could have been carried out in this particular case, since we know the exact distribution of $T_n$, which is a binomial distribution

$$F_{T_n}(x) = \sum_{i=0}^{n} \binom{n}{i} \left(\frac{e^{\xi}}{1+e^{\xi}}\right)^i \left(\frac{1}{1+e^{\xi}}\right)^{n-i} u(x-i), \ \ 0 \leq x \leq n, \tag{36}$$

Thus,

$$dF_{T_n}(x) = \sum_{i=0}^{n} \binom{n}{i} \left(\frac{e^{\xi}}{1+e^{\xi}}\right)^i \left(\frac{1}{1+e^{\xi}}\right)^{n-i} \delta(x-i)\, dx, \ \ 0 \leq x \leq n, \tag{37}$$

and

$$\begin{aligned} A(\xi) &= \sum_{i=n\beta}^{n} \binom{n}{i} \left(\frac{e^{\xi}}{1+e^{\xi}}\right)^i \left(\frac{1}{1+e^{\xi}}\right)^{n-i} e^{-(\xi i - n\xi\beta)} \\ &= \frac{e^{n\beta\xi}}{(1+e^{\xi})^n} \sum_{i=n\beta}^{n} \binom{n}{i}, \end{aligned} \tag{38}$$

where we assume that the threshold $b = n\beta$ is an integer. If not, we replace $\beta$ by $\tilde{\beta} = \lceil n\beta \rceil / n$ defined earlier.
Substitution of $A(\xi)$ of (**??**), $e^{-\xi b} = e^{-n\beta\xi}$ and $M_{S_n}(\xi) = \frac{\left(1+e^{\xi}\right)^n}{2^n}$ into (10.90) yields

$$P[S_n \geq n\beta] = 2^{-n} \sum_{i=\lceil n\beta \rceil}^{n} \binom{n}{i}, \tag{39}$$

which is the exact expression for $P[S_n \geq n\beta]$ as given in (25). This result should not be surprising, since we have not introduced any approximation in this manipulation. All we have done is to translate the **fair** coin tossing problem into an equivalent *unfair* coin tossing problem by introducing the tilted distribution.

**10.29 Derivation of (10.123).** The twisted RVs $Y_i$'s are also i.i.d. with common PDF $f_Y(y)$, and their expectations are

$$\begin{aligned} E[Y_i] &= \int y f_Y(y)\, dy = \int y \frac{e^{\xi^* y} f_X(y)}{M_X(\xi^*)}\, dy \\ &= \frac{\int y e^{\xi^* y} f_X(y)\, dy}{M_X(\xi^*)} = \frac{\frac{d}{d\xi^*} \int e^{\xi^* y} f_X(y)\, dy}{M_X(\xi^*)} \\ &= \frac{\frac{d}{d\xi^*} M_X(\xi^*)}{M_X(\xi^*)} = \frac{d}{d\xi^*} \left[\ln M_X(\xi^*)\right] \\ &= m'_X(\xi^*) = \beta. \end{aligned}$$

# 11 Solutions for Chapter 11: Convergence of a Sequence of Random Variables

**Appendix (or Supplementary Material):** This is the proof of Lemma 11.25 in the text that is omitted because of the space.

**Lemma 11.1** [**Conditions for a.s. convergence**]
$X_n \overset{\text{a.s.}}{\to} X$, if and only if, for arbitrary $\epsilon > 0$ and $\delta > 0$, there exists a number $M(\epsilon, \delta)$ such that

$$P\left[\bigcap_{n=m}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \epsilon\}\right] \geq 1 - \delta \tag{1}$$

for all $m \geq M(\epsilon, \delta)$.

*Proof.* Let sets $A$ and $A_n(\epsilon)$ be as defined in (**??**) and (11.6), respectively:

$$A = \{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}, \tag{2}$$

$$A_n(\epsilon) = \{\omega : |X_n(\omega) - X(\omega)| < \epsilon\}. \tag{3}$$

We define a sequence

$$B_m(\epsilon) = \bigcap_{n=m}^{\infty} A_n(\epsilon), \tag{4}$$

which is an increasing sequence with the limit

$$\lim_{m\to\infty} B_m(\epsilon) \triangleq A(\epsilon). \tag{5}$$

The limit $A(\epsilon)$ can be interpreted as

$$A(\epsilon) = \{\omega \in A_n(\epsilon) \text{ for infinitely many values of } n\}. \tag{6}$$

From the definition of almost sure convergence

$$X_n \overset{\text{a.s.}}{\to} X \iff P[A] = 1, \tag{7}$$

where the symbol $\iff$ means "if and only if". Since the events $A$ and $A(\epsilon)$ are related by

$$A = \bigcap_{\epsilon>0} A(\epsilon), \tag{8}$$

we have

$$P[A^c] = P\left[\bigcup_{\epsilon>0} A^c(\epsilon)\right] \le \sum_{\epsilon>0} P[A^c(\epsilon)]. \tag{9}$$

Therefore, it follows that

$$P[A^c] = 0 \iff P[A^c(\epsilon)] = 0 \text{ for any } \epsilon > 0. \tag{10}$$

or

$$P[A] = 1 \iff P[A(\epsilon)] = 1 \text{ for any } \epsilon > 0. \tag{11}$$

Because of (5), we have

$$P[A(\epsilon)] = 1 \iff \lim_{m\to\infty} P[B_m(\epsilon)] = 1 \tag{12}$$

Thus, from (7), (11) and (12), we conclude

$$X_n \overset{\text{a.s.}}{\to} X \iff \lim_{m\to\infty} P[B_m(\epsilon)] = 1 \text{ for any } \epsilon > 0. \tag{13}$$

The right hand side of the above implies, by referring to (4)

$$\lim_{m\to\infty} P\left[\bigcap_{n=m}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \epsilon\}\right] = 1, \text{ for any } \epsilon > 0. \tag{14}$$

In other words, for any $\epsilon > 0$ there exists a number $N(\epsilon, \delta)$ such that for all $m \ge N(\epsilon, \delta)$

$$P\left[\bigcap_{n=m}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \epsilon\}\right] \ge 1 - \delta, \tag{15}$$

which is equivalent to (11.25). □

## 11.1 Preliminaries: Convergence of a Sequence of Numbers or Functions

## 11.2 Types of Convergence for Sequences of Random Variables

**11.1* Example of D. convergence.** The distribution function of $Z_n$ is given by

$$F_{Z_n}(z) = P[Z_n \le z] = P[n(1 - Y_n) \le z] = P\left[Y_n \ge 1 - \frac{z}{n}\right]. \tag{16}$$

Since $Y_n = \max\{X_1, X_2, \dots, X_n\}$,

$$P\left[Y_n \le 1 - \frac{z}{n}\right] = P\left[X_i \le 1 - \frac{z}{n},\ 1 \le i \le n\right] = \left[F_X\left(1 - \frac{z}{n}\right)\right]^n. \tag{17}$$

Note that $0 \le Y_n \le 1$, almost surely, and hence $Z_n \ge 0$, a.s. When $n > z \ge 0$,

$$0 \le 1 - \frac{z}{n} \le 1,$$

and therefore

$$F_X\left(1-\frac{z}{n}\right)=1-\frac{z}{n},\;\; n>z. \tag{18}$$

Hence,

$$\lim_{n\to\infty}\left[F_X\left(1-\frac{z}{n}\right)\right]^n=\lim_{n\to\infty}\left[1-\frac{z}{n}\right]^n=e^{-z}. \tag{19}$$

From (16), (17), and (19), we conclude that

$$\lim_{n\to\infty}F_{Z_n}(z)=1-e^{-z},\;\; z\geq 0,$$

i.e., $Z_n \stackrel{D}{\to} Z$.

**11.2 Bernstein's lemma.**

Let $\epsilon>0$. Then

$$\begin{aligned} P[Z_n\leq z] &= P[X_n+Y_n\leq z] \\ &\leq P[X_n<z+\epsilon,\;|Y_n|\geq\epsilon] \qquad (20)\\ &\leq P[X_n<z+\epsilon]+P[|Y_n|\geq\epsilon] \qquad (21)\\ &\leq P[X_n<z+\epsilon]+\frac{\text{Var}[Y_n]}{\epsilon^2}, \qquad (22)\end{aligned}$$

where inequality (20) holds because

$$X_n+Y_n\leq z\Rightarrow\{X_n<z+\epsilon,\;Y_n\leq-\epsilon\}\text{ or }Y_n\geq\epsilon,$$

and (22) follows from Chebyshev's inequality. Using a similar approach,

$$\begin{aligned} P[Z_n\geq z] &= P[X_n+Y_n\geq z] \\ &\geq P[X_n<z-\epsilon,\;|Y|\geq\epsilon] \\ &\leq P[X_n<z-\epsilon]+\frac{\text{Var}[Y_n]}{\epsilon^2}. \qquad (23)\end{aligned}$$

Taking the limit as $n\to\infty$ on both sides of (22), we have

$$\lim_{n\to}P[Z_n\leq z]\leq\lim_{n\to}P[X_n<z+\epsilon]+\lim_{n\to\infty}\frac{\text{Var}[Y_n]}{\epsilon^2}=F_X(z+\epsilon), \tag{24}$$

where the last equality follows because $X_n \stackrel{D}{\to} X$ and Var$[Y_n]\to 0$. Similarly, taking limits on both sides of (23), we obtain

$$\lim_{n\to\infty}P[Z_n\geq z]\geq F_X(z-\epsilon). \tag{25}$$

From (24) and (25), we have

$$F_X(z-\epsilon)\leq\lim_{n\to\infty}F_{Z_n}(z)\leq F_X(z+\epsilon).$$

Since $\epsilon>0$ is arbitrary, we have

$$\lim_{n\to\infty}F_{Z_n}(z)=F_X(z),$$

i.e., $Z_n \stackrel{D}{\to} X$.

**11.3* Convergence of sample average.**

$$\overline{X}_n - c = \frac{1}{n}\sum_{k=1}^{n}(c + N_k) - c = \frac{1}{n}\sum_{k=1}^{n} N_k.$$

By the weak law of large numbers,

$$\frac{1}{n}\sum_{k=1}^{n} N_k \xrightarrow{P} 0.$$

Therefore,

$$\lim_{n\to\infty} P[|\overline{X}_n - c| \geq \epsilon] = 0,$$

for any $\epsilon > 0$. Hence, $\overline{X}_n \xrightarrow{P} c$.

**11.4 P. convergence and m.s. convergence [319].**

$X_n \xrightarrow{P} X$ implies that there exists an $M(\epsilon, \delta)$ such that for arbitrarily small $\epsilon > 0$ and $\delta > 0$,

$$P[|X_n - X| > \epsilon] < \delta \quad \text{for all } n > M(\epsilon, \delta). \tag{26}$$

Consider the quantity

$$\mathcal{E}_n = E[|X_n - X|^2].$$

We have

$$\mathcal{E}_n = E[|X_n - X|^2 1_{\{|X_n - X| < \epsilon\}}] + E[|X_n - X|^2 1_{\{X_n - X| \geq \epsilon\}}]. \tag{27}$$

Clearly, the first term on the right-hand side can be upper-bounded by $\epsilon^2$.
To upper-bound the second term, we use the fact that the PDF of $X_n$ has finite support for all $n > N$. In particular,

$$P[|X_n| \leq x_0] = 1, \quad \text{for all } n > N. \tag{28}$$

From Theorem 11.2, $X_n \xrightarrow{P} X$ implies $X_n \xrightarrow{D} X$. Hence,

$$P[|X| \leq x_0] = \lim_{n\to\infty} P[|X_n| \leq x_0] = 1. \tag{29}$$

Equations (28) and (29) imply that

$$P[|X - X_n| \leq 2x_0] = 1. \tag{30}$$

Then for $n > \max\{M, N\}$, we have

$$\begin{aligned} & E[|X_n - X|^2 \mathbf{1}_{\{|X_n - X| \geq \epsilon\}}] \\ & \quad = E[|X_n - X|^2 \mathbf{1}_{\{\epsilon \leq |X_n - X| \leq 2x_0\}}] && (31) \\ & \quad \leq 4x_0^2 E[\mathbf{1}_{\{\epsilon \leq |X_n - X| \leq 2x_0\}}] && (32) \\ & \quad = 4x_0^2 P[|X_n - X| \geq \epsilon] < 4x_0^2 \delta, && (33) \end{aligned}$$

where we applied (30) to obtain (31) and we applied (26) in (33). Then we have,

$$\mathcal{E}_n < \epsilon^2 + 4x_0^2\delta \triangleq \delta', \quad \text{for all } n > \max\{M, N\},$$

where $\delta'$ can be made as small as we wish by choosing sufficiently small $\epsilon$ and $\delta$. Hence,

$$\lim_{n\to\infty} E[|X_n - X|^2] = \lim_{n\to\infty} \mathcal{E}_n = 0.$$

**A similar proof:**

Let $F_X(x)$ denote the CDF of the RV $X$. From (29), we have

$$P[|X| \le x_0] = \int_{-x_0}^{x_0} dF_X(x) = 1.$$

Fix $a \in [-x_0, x_0]$. For $n > \max\{N, M\}$, the conditional expectation of $|X_n - X|^2$ given $X = a$ is

$$\begin{aligned}
&E[|X_n - X|^2 \mid X = a] \\
&\le (2x_0)^2 \left[\int_{-\infty}^{a-\epsilon} f_n(x)\,dx + \int_{a+\epsilon}^{\infty} f_n(x)\,dx\right] + \epsilon^2 \int_{a-\epsilon}^{a+\epsilon} f_n(x)\,dx \\
&= 4x_0^2 P[|X_n - X| > \epsilon \mid X = a] + \epsilon^2 P[|X - a| \le \epsilon] \\
&\le 4x_0^2 P[|X_n - X| > \epsilon \mid X = a] + \epsilon^2. \qquad (34)
\end{aligned}$$

Applying the law of iterated expectations from (4.106) and using (34) we have

$$\begin{aligned}
\mathcal{E}_n &= E[E[|X_n - X|^2 \mid X]] \\
&= \int_{-x_0}^{x_0} E[|X_n - X|^2 \mid X = a]\,dF_X(a) \\
&\le 4x_0^2 \int_{-x_0}^{x_0} P[|X_n - X| > \epsilon \mid X = a]\,dF_X(a) + \epsilon^2 \\
&= 4x_0^2 P[|X_n - X| > \epsilon] + \epsilon^2 \\
&< 4x_0^2 \delta + \epsilon^2 \triangleq \delta',
\end{aligned}$$

where we have applied (26) to obtain the last inequality. Since $\delta'$ can be made as small as we wish by choosing sufficiently small $\epsilon$ and $\delta$,

$$\lim_{n\to\infty} E[|X_n - X|^2] = \lim_{n\to\infty} \mathcal{E}_n = 0.$$

**11.5 Proof of Lyapunov's inequality.**

(a) Using $\mu_r = E[|Y|^r]$, the norm (11.40) can be written as

$$\|Y\|_r = \mu_r^{1/r}.$$

Consider a new random variable $X$ defined by

$$X = a|Y|^{\frac{r-s}{2}} + |Y|^{\frac{r+s}{2}},$$

where $a$ is a real parameter. Then

$$E[X^2] = a^2\mu_{r-s} + 2a\mu_r + \mu_{r+s} \ge 0$$

must hold for any $a$. Therefore, the discriminant $D$ of the above quadratic form in $a$ must be non-positive, that is,

$$\frac{D}{4} = \mu_r^2 - \mu_{r-s}\mu_{r+s} \leq 0,$$

from which we have

$$\mu_r^2 \leq \mu_{r-s}\mu_{r+s}, \tag{35}$$

which is a variation of the well-known Cauchy-Schwarz[1] inequality:

$$(E[|XY|])^2 \leq E[|X|^2]E[|Y|^2].$$

(b) From the inequality (35), we have

$$g(r) \leq \frac{1}{2}\left[g(r-s) + g(r+s)\right].$$

Hence $g(r)$ is a convex $\cup$ function that passes through the origin $(0,0)$, and the slope of the tangent at $r$, $g'(r)$, is steeper than $g(r)/r$, i.e.,

$$g'(r) \geq \frac{g(r)}{r}. \tag{36}$$

Therefore, $g(r)$ is a non-decreasing function of $r$ for $r > 0$.

(c) The derivative of $h(r) = \frac{g(r)}{r}$ is non-negative because of (36):

$$h'(r) = \frac{g'(r)r - g(r)}{r^2} \geq 0. \tag{37}$$

Hence, $h(r)$ is also a non-decreasing function.

(d) The function $h(r)$ is related to the norm $\|Y\|_r$ according to

$$h(r) = \frac{\log \mu_r}{r} = \log\left(\mu_r^{1/r}\right) = \log \|Y\|_r.$$

Thus $\|Y\|_r$ is also a non-decreasing function of $r > 0$, that is, (11.41) holds.

**Note:** A somewhat simpler proof can be found by noting that it is sufficient to prove the inequality for the case $s = r - 1$. Then from (35), we have

$$\mu_r^2 \leq \mu_{r-1}\mu_{r+1}, \text{ or } \mu_r^{2r} \leq \mu_{r-1}^r \mu_{r+1}^r, \quad r = 1, 2, \ldots, \tag{38}$$

By setting $r = 1, 2, 3, \ldots, n-1$, we have the following series of inequalities:

$$\mu_1^2 \leq \mu_0\mu_2, \quad \mu_2^4 \leq \mu_1^2\mu_3^2, \quad \mu_3^6 \leq \mu_2^3\mu_4^3, \quad \ldots, \quad \mu_{n-1}^{2(n-1)} \leq \mu_{n-2}^{n-1}\,\mu_n^{n-1}, \tag{39}$$

where $\mu_0 = 1$. By multiplying the first two inequalities, we have

$$\mu_1^2\mu_2^4 \leq \mu_2\mu_1^2\mu_3^2, \text{ or } \mu_2^3 \leq \mu_3^2.$$

By successively multiplying, we obtain

$$\mu_1^2 \leq \mu_2, \quad \mu_2^3 \leq \mu_3^2, \quad \mu_3^4 \leq \mu_4^3, \quad \ldots, \quad \mu_{n-1}^n \leq \mu_n^{n-1}.$$

[1] Karl Hermann Amandus Schwarz (1843-1921) was a German mathematician.

From the last inequality we readily find

$$\mu_{n-1}^{1/n-1} \leq \mu_n^{1/n}, \text{ or } \|Y\|_{n-1} \leq \|Y\|_n, \text{ for } n \geq 2, \tag{40}$$

which implies

$$\|Y\|_s \leq \|Y\|_r, \text{ for } 1 \leq s < r < \infty,$$

**11.6* Properties of $\| Y \|_r$ [131].**

(a) **Hölder's inequality:** From the convexity of the exponential function, we have, from Jensen's inequality, for any real numbers $u$ and $v$, and $\frac{1}{r} + \frac{1}{s} = 1$,

$$\exp\left(\frac{u}{r} + \frac{v}{s}\right) \leq \frac{e^u}{r} + \frac{e^v}{s}. \tag{41}$$

Set

$$u = \ln\left(\frac{|X|}{\|X\|_r}\right)^r, \text{ and } v = \ln\left(\frac{|Y|}{\|Y\|_s}\right)^s.$$

Then the LHS of (41) is

$$\begin{aligned}
\text{LHS} &= \exp\left(\frac{1}{r}\ln\left(\frac{|X|}{\|X\|_r}\right)^r + \frac{1}{s}\ln\left(\frac{|Y|}{\|Y\|_s}\right)^s\right) \\
&= \exp\frac{1}{r}\ln\left(\frac{|X|}{\|X\|_r}\right)^r \cdot \exp\frac{1}{s}\ln\left(\frac{|Y|}{\|Y\|_s}\right)^s \\
&= \frac{|XY|}{\|X\|_r\|Y\|_s}.
\end{aligned}$$

Using

$$e^u = \left(\frac{|X|}{\|X\|_r}\right)^r, \text{ and } e^v = \left(\frac{|Y|}{\|Y\|_s}\right)^s,$$

the RHS of (41) is

$$\text{RHS} = \frac{1}{r}\frac{|X|^r}{\|X\|_r^r} + \frac{1}{s}\frac{|Y|^s}{\|Y\|_s^s}.$$

Thus

$$\frac{|XY|}{\|X\|_r\|Y\|_s} \leq \frac{1}{r}\frac{|X|^r}{\|X\|_r^r} + \frac{1}{s}\frac{|Y|^s}{\|Y\|_s^s}.$$

By taking the expectation, we find

$$\frac{\|XY\|_1}{\|X\|_r\|Y\|_s} \leq \frac{1}{r} + \frac{1}{s} = 1.$$

(b) The proof given in part (a) can carry over to this case. From (41), we have

$$\sum_{i=1}^n \exp\left(\frac{u_i}{r} + \frac{v_i}{s}\right) \leq \sum_{i=1}^n \left(\frac{e_i^u}{r} + \frac{e_i^v}{s}\right). \tag{42}$$

Set

$$u_i = \ln\left(\frac{x_i}{\|\boldsymbol{x}\|_r}\right)^r, \text{ or } e^{u_i} = \frac{x_i^r}{\|\boldsymbol{x}\|_r^r},$$

where

$$\|\boldsymbol{x}\|_r = \left(\sum_{i=1}^n x_i^r\right)^{1/r}, \text{ for } r > 1.$$

Then the LHS of (42) is

$$\text{LHS} = \sum_{i=1}^n \exp\frac{1}{r}\ln\left(\frac{x_i}{\|\boldsymbol{x}\|_r}\right)^r \cdot \exp\frac{1}{s}\ln\left(\frac{y_i}{\|\boldsymbol{x}\|_s}\right)^s = \frac{\sum_{i=1}^n x_i y_i}{\|\boldsymbol{x}\|_r\|\boldsymbol{y}\|_s},$$

and the RHS is

$$\text{RHS} = \frac{\sum_{i=1}^n x_i^r}{r\|\boldsymbol{x}\|_r^r} + \frac{\sum_{i=1}^n y_i^s}{s\|\boldsymbol{y}\|_s^s} = \frac{1}{r} + \frac{1}{s} = 1.$$

**An alternative proof:** We use the hint given in part (b). It is easy to see that $F(x)$ is minimum when $x = 1$. Thus,

$$\frac{x^r}{r} + \frac{x^{-s}}{s} \geq 1.$$

In order to derive

$$uv \leq \frac{u^r}{r} + \frac{v^s}{s}$$

We set $x = u^{1/s}v^{-1/r} = u^r r + sv^{-\frac{s}{r+s}}$ in the above inequality, then we find

$$uv \leq \frac{u^r}{r} + \frac{v^s}{s},$$

and the rest of the proof is similar to the first proof.
The proof for the integral version is similar. Instead of $\sum_{i=1}^n$, use the integral.
(c) **Minkowski's inequality:**

$$\begin{aligned}
E[|X+Y|^r] &= E[|X+Y||X+Y|^{r-1}] \\
&\leq E[(|X|+|Y|)|X+Y|^{r-1}] \\
&= E[|X||X+Y|^{r-1}] + E[|Y||X+Y|^{r-1}] && (43) \\
&\leq (E[|X|^r])^{1/r} \cdot \left(E[|X+Y|^{(r-1)s}]\right)^{1/s} \\
&\quad + (E[|Y|^r])^{1/r} \cdot \left(E[|X+Y|^{(r-1)s}]\right)^{1/s} && (44) \\
&= (||X||_r + ||Y||_r)\left(E[|X+Y|^r]\right)^{\frac{r-1}{r}}, && (45)
\end{aligned}$$

where (44) is obtained by applying Hölder's inequality to each term in (43), with

$$\frac{1}{r} + \frac{1}{s} = 1 \Rightarrow s = \frac{r}{r-1}.$$

Multiplying both sides of the inequality (45) by the factor

$$\frac{||X+Y||_r}{E[|X+Y|^r]},$$

we obtained the desired result:

$$||X+Y||_r \leq ||X||_r + ||Y||_r.$$

**11.7 Convergence in the $r$th mean vs. a.s. convergence.**

(a)

$$E[|X_n - 0|^r| = E[|X_n|^r] = \frac{1}{n}\cdot(n^{1/2r})^r + \frac{n-1}{n}\cdot 0^r = \frac{1}{n^{1/2}} \to 0. \tag{46}$$

Therefore, $X_n \xrightarrow{r} 0$.

(b) Let $M$ be an integer such that

$$M^{1/2r} > \epsilon \Rightarrow M > \epsilon^{2r}.$$

Then for $n \geq M$,

$$P[A_n(\epsilon)] = P[X_n < \epsilon] = P[X_n = 0] = \frac{n-1}{n}.$$

Hence, for $m \geq M$,

$$\begin{aligned} P[B_m(\epsilon)] &= P\left[\bigcap_{n=m}^{\infty} A_n(\epsilon)\right] = \prod_{n=m}^{\infty} P[A_n(\epsilon)] \\ &= \lim_{M\to\infty} \frac{m-1}{m}\cdot\frac{m}{m+1}\cdots\frac{M-1}{M}\frac{M}{M+1} \\ &= \lim_{M\to\infty}\frac{m-1}{M} = 0. \end{aligned}$$

(c) $X_n \xrightarrow{\text{a.s.}} 0$ if and only if $\lim_{m\to\infty} P[B_m(\epsilon)] = 1$. By the result of part (b), $\{X_n\}$ does not converge almost surely to 0.

**11.8 Convergence in the $r$th mean vs. a.s. convergence-continued [319].**

(a)

$$E[|X_n|^r] = \frac{1}{n^2}\cdot e^{nr} + \left(1 - \frac{1}{n^2}\right)\cdot 0 = \frac{e^{nr}}{n^2} \to \infty, \ \ r > 0.$$

Therefore, $X_n$ does not converge in $r$th mean to 0 for $r > 0$.

(b)

$$P[A_n(\epsilon)] = P[X_n < \epsilon] = P[X_n = 0] = 1 - \frac{1}{n^2},$$

provided $e^n < \epsilon$, i.e., $n > \ln(\epsilon)$. Let $M$ be an integer such that $M > \ln(\epsilon)$. For all $m \geq M$,

$$P[B_m(\epsilon)] = P\left[\bigcap_{n=m}^{\infty} A_n(\epsilon)\right] = \prod_{n=m}^{\infty} P[A_n(\epsilon)] \qquad = \prod_{n=m}^{\infty}\left(1 - \frac{1}{n^2}\right).$$

From the first inequality of the hint, we have

$$\frac{1}{2} \le P[B_m(\epsilon)] \le 1.$$

This number approaches unity as $m \to \infty$, i.e.,

$$\lim_{m\to\infty} P[B_m(\epsilon)] = 1.$$

(c) From the result of part (b), we conclude that $X_n \overset{\text{a.s}}{\to} 0$.

## 11.3 Limit theorems

**11.9 Limits in Bernoulli trials.**

(a) The $E_k$'s are independent events, and $P[E_k] = p$ for all $k \ge 1$. Then $A_n = \bigcup_{k=n}^{\infty}$ represents the event that has at least one success in the $n$th and following trials. Thus, the complement $A_k^c$ represents the event that has only failures in the $n$th and following trials. The probability of having $r$ consecutive failures, starting at the $n$th trial, is given by $(1-p)^r = q^r$, regardless of $n$. Hence the probability of having at lease one success is $1 - q^r$, regardless of $n$ again. Thus we have

$$P[A_n] = \lim_{r\to\infty}(1 - q^r) = 1, \text{ for all } n \ge 1,$$

which readily implies that $P[A] = 1$. Thus, infinitely many successes occur with probability 1. This conclusion certainly agrees to our intuition.

(b) We readily find $P[E_k] = p^k$, but the events $E_k$'s are not mutually independent, and the sequence $\{E_k\}$ is a decreasing sequence. Thus,

$$A_n = \bigcup_{k=n}^{\infty} E_k = E_n.$$

Since $A = \bigcap_{n=1}^{\infty} E_n$, we have from Theorem 11.13

$$P[A] = P\left[\bigcap_{n=1}^{\infty} E_n\right] = \lim_{n\to\infty} P[E_n] = \lim_{n\to\infty} p^n = 0.$$

Thus, only finitely many of the events $E_1, E_2, \ldots$ occur with probability 1. This conclusion again agrees to our intuition.

**11.10 Independence of complements of independent events.** If $E_k$'s are independent

$$P[E_1 \cap E_2 \cap \cdots \cap E_n] = P[E_1]P[E_2]\cdots P[E_n].$$

Then from de Morgan's law (2.17), we have

$$E_1^c \cap E_2^c \cap \cdots \cap E_n^c = (E_1 \cup E_2 \cup \cdots E_n)^c = F_n^c,$$

where

$$F_n \triangleq E_1 \cup E_2 \cup \cdots \cup E_n.$$

Thus

$$P[E_1^c \cap E_2^c \cap \cdots \cap E_n^c] = 1 - P[F_n]. \tag{47}$$

We showed in Chapter 2 that for $n = 2$ (see Problem 2.11)

$$P[F_2] = P[E_1 \cup E_2] = 1 - (1 - P[E_1])(1 - P[E_2]) = 1 - P[E_1^c]P[E_2^c]$$

Then by mathematical induction we can show

$$P[F_n] = P[F_{n-1} \cup E_n] = 1 - (1 - P[F_{n-1}])(1 - P[E_n]) = 1 - P[F_{n-1}^c]P[E_n^c] = 1 - \prod_{i=1}^{n} P[E_i^c].$$

Hence

$$P[E_1^c \cap E_2^c \cap \cdots \cap E_n^c] = \prod_{i=1}^{n} P[E_i^c],$$

i.e., the events $E_k^c$'s are mutually independent.

**11.11 Borel-Cantelli lemmas and Bernoulli trials.**

(a) For this case, we have $P[E_k] = p$ for all $k \geq 1$. Thus,

$$\sum_{k=1}^{\infty} P[E_k] = \sum_{k=1}^{\infty} p = \infty, \quad \text{for any} \;\; p > 0.$$

Thus, we conclude from the second lemma, that infinitely many successes occur with probability 1.

(b) For this case, we have

$$\sum_{k=1}^{\infty} P[E_k] = \sum_{k=1}^{\infty} p^k = \frac{p}{1-p} < \infty.$$

Thus, from the first lemma, we can conclude that with probability 1 only finitely many of the events $E_1, E_2, \ldots$ occur.

**11.12 Proof of the CLT.** The normalized average $Z_n$ is given by

$$Z_n = \sum_{k=1}^{n} \tilde{X}_k,$$

where

$$\tilde{X}_k = \frac{X_k - \mu}{\sqrt{n}\sigma}.$$

The moment generating function (MGF) of $Z_n$ is then given by

$$\begin{aligned} M_{Z_n}(t) &= E\left[\exp\left\{t \sum_{k=1}^{n} \left(\frac{X_k - \mu}{\sigma\sqrt{n}}\right)\right\}\right] \\ &= \prod_{k=1}^{n} E\left[\exp\left\{t\left(\frac{X_k - \mu}{\sigma\sqrt{n}}\right)\right\}\right] = [M_{\tilde{X}}(t)]^n, \end{aligned} \tag{48}$$

where $M_{\tilde{X}}(t)$ is the common MGF of $\tilde{X}_k$. Using the Taylor series expansion, we can write $M_{\tilde{X}}(t)$ as follows:

$$M_{\tilde{X}}(t) = 1 + E\left[\frac{X_k - \mu}{\sigma\sqrt{n}}\right] t + \frac{1}{2} E\left[\left(\frac{X_k - \mu}{\sigma\sqrt{n}}\right)^2\right] t^2 + o\left(\frac{t^2}{n^2}\right) \tag{49}$$

$$= 1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n^2}\right). \tag{50}$$

Hence,

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty} \left[1 + \frac{t^2}{2n} + o\left(\frac{t^2}{n^2}\right)\right]^n = e^{\frac{t^2}{2}},$$

which is the MGF of the unit normal distribution. This establishes that $Z_n$ converges in distribution to the unit normal variable.

**11.13 Product of independent RVs**

(a) Taking the natural logarithm of $R_n$, we find

$$S_n = \ln R_n = \sum_{k=1}^{n} \ln Y_k \tag{51}$$

Let $\ln Y_k \triangleq X_k$, then

$$S_n = \sum_{k=1}^{n} X_k.$$

If the independent RVs $X_k$ satisfy the conditions of either Theorem 11.23, or Lyapunov's condition (11.90), or Lindeberge's condtion 11.93, then from the CLT we can argue that the RV $Z_n = \frac{S_n - m_n}{s_n}$, where

$$m_n = E[S_n] = \sum_{k=1}^{n} E[\ln Y_k], \text{ and } s_n = \mathrm{Var}[S_n] = E\left[(\ln Y_k - E[\ln Y_k])^2\right],$$

should converge in distribution to the unit normal variable $U$. Hence for sufficiently large $n$, the distribution of $S_n$ is approximately normally distributed with mean $m_n$ and variance $s_n^2$, i.e.,

$$f_{S_n}(s) \approx \frac{1}{\sqrt{2\pi} s_n} \exp\left\{-\frac{(s - m_n)^2}{2s_n^2}\right\}, \quad -\infty < s < \infty. \tag{52}$$

Then $R_n = e^{S_n}$ is lognormally distributed as discussed in Section 7.4, and its PDF is given by

$$f_{R_n}(r) \approx \frac{1}{\sqrt{2\pi} s_n} \exp\left\{-\frac{(\ln r - m_n)^2}{2s_n^2}\right\}, \quad s > 0, \tag{53}$$

where

(b) In order to find distribution parameters of $X_k$ such as the mean $\mu_k$, variance $\sigma_k$, the third absolute central moment $m_k^3$, or $\sigma_k^{\star 2}$ that are used in expressing the conditions for the CLT to hold, we need to have information about the PDF $f_{Y_k}(y)$ of the RV $Y_k = e^{X_k}$, or at least its moment generating function (MGF) $M_{Y_k}(t) = E\left[e^{tY_k}\right]$. See the discussion of Section 7.4 how to find the mean, variance and higher central moments from the MGF. Needless to say, if the individual $X_k$ is normally distributed (i.e., $Y_k$ is known to be log-normally distributed, the first two moments of $Y_k$ can determine $\mu_k$ and $\sigma_k$ of $X_k$. In this case, however, $S_n$ is known to be normally distributed for any $n$, there is no need to bring up the CLT argument.

# 12 Solutions for Chapter 12: Random Process, Spectral Analysis and Complex Gaussian Process

## 12.1 Introduction

## 12.2 Classification of Random Processes

## 12.3 Stationary Random Process

**12.1* Sinusoidal functions with different frequencies and random amplitudes [175].**

(a)

$$R_X(\tau) = E[X(t+\tau)X(t)] = E\left[\left(\sum_{i=0}^{m}\{A_i\cos\omega_i(t+\tau) + B_i\sin\omega_i(t+\tau)\}\right)\right.$$
$$\left.\cdot\left(\sum_{j=0}^{m}\{A_j\cos\omega_j t + B_j\sin\omega_j t\}\right)\right].$$

Noting that $E[A_iB_j] = 0$ and $E[A_iA_j] = E[B_iB_] = 0$ for $i \neq j$, we find

$$R_X(\tau) = \sum_{i=0}^{m} E[A_i^2\cos\omega_i(t+\tau)\cos\omega_i t + B_i^2\sin\omega_i(t+\tau)\sin\omega_i t]$$
$$= \sum_{i=0}^{m}\sigma_i^2\cos\omega_i\tau = \sigma^2\sum_{i=0}^{m} f_i\cos\omega_i\tau.$$

(b)

$$R_X(\tau) = \sigma^2\int_0^{\pi}\cos\omega\tau\, dF(\omega).$$

(c)

$$R_X(\tau) = \frac{\sigma^2}{\pi}\int_0^{\pi}\cos\omega\tau\, d\omega = \begin{cases}\sigma^2, & \text{if } \tau = 0,\\ 0, & \text{if } \tau \neq 0.\end{cases}$$

For a detailed mathematical treatment see Chapter 9 of Karlin and Taylor [174].

**12.2 Moving average process [175].**

(a)

$$E[Y_n] = \mu(a_0 + a_1 + \cdots + a_{m-1})$$
$$\mathrm{Var}[Y_n] = \sigma^2(a_0^2 + a_1^2 + \cdots + a_{m-1}^2)$$

(b) Write $X_n - \mu \triangleq \tilde{X}_n$. Then $E[\tilde{X}_n] = 0$ and $E[\tilde{X}_m \tilde{X}_n] = \sigma^2 \delta_{m,n}$. The autocovariance between $Y_n$ and $Y_{n+k}$ is

$$\begin{aligned}
&E\left[\left(Y_n - \mu \sum_{i=0}^{m-1} a_i\right)\left(Y_{n+k} - \mu \sum_{j=0}^{m-1} a_j\right)\right] \\
&= E\left[\left(\sum_{i=0}^{m-1} a_i \tilde{X}_{n-i}\right)\left(\sum_{j=0}^{m-1} a_j \tilde{X}_{n+k-j}\right)\right] \\
&= \begin{cases} \sigma^2(a_0 a_k + a_1 a_{k+1} + \cdots + a_{m-1-k} a_{m-1}), & k \le m-1. \\ 0, & k \ge m. \end{cases}
\end{aligned}$$

Since the covariance between $Y_n$ and $Y_{n+k}$ depends only on the time lag $k$ and not on $n$, the process $\{Y_n\}$ is WSS.

(c)

$$R_Y(k) = \begin{cases} \sigma^2\left(1 - \frac{k}{m}\right), & k \le m-1, \\ 0, & k \ge m. \end{cases}$$

(d) The case $m = 1$, $Y_n = a_0 X_n$ corresponds to the uncorrelated random sequence $\{X_n\}$ of Example 12.1. The other extreme $m = \infty$ roughly corresponds to the sequence $\{Z_n\}$ of Example 12.3.

## 12.4 Complex-Valued Gaussian Process

**12.3* Condition for integration in mean-square.**

We can expand the LHS of (12.39) as follows:

$$\begin{aligned}
E[|Y - S_n|^2] &= E[YY^*] - E[YS_n^*] - E[Y^*S_n] + E[S_n S_n^*] \\
&= \int_a^b \int_a^b h(t)E[Z(t)Z^*(s)]h^*(s)\,dt\,ds \\
&- \int_a^b \sum_{i=1}^n h(t)E[Z(t)Z^*(t_i)]h^*(t_i)(t_{i+1} - t_i) \\
&- \int_a^b \sum_{i=1}^n h^*(t)E[Z^*(t)Z(t_i)]h(t_i)(t_{i+1} - t_i) \\
&+ \sum_{i=1}^n \sum_{j=1}^n h(t_i)E[Z(t_i)Z^*(t_j)]h^*(t_j)(t_{i+1} - t_i)(t_{j+1} - t_j).
\end{aligned} \tag{1}$$

Since $E[Z(t)Z^*(s)] = R_{ZZ}(t,s)$, the first term equals $Q$ of (12.40). By taking the limit $n \to \infty$ and $\max\{t_{i=1} - t_i\} \to 0$, the second term becomes

$$\lim_{n\to\infty} \int_a^b \sum_{i=1}^n h(t)E[Z(t)Z^*(t_i)]h^*(t_i)(t_{i+1} - t_i) = \lim_{n\to\infty} \int_a^b \sum_{i=1}^n h(t)R(t,t_i)h^*(t_i)(t_{i+1} - t_i)$$
$$= \int_a^b \int_a^b h(t)R(t,s)h^*(s)\,ds = Q$$

Similarly the third and fourth terms of (1) also equal $Q$. Thus,

$$\lim_{n\to\infty} E[|Y - S_n|^2] = Q - Q - Q + Q = 0.$$

**12.4* Circular symmetry criterion for a complex Gaussian process.**
First, note that $Q_{Ze^{i\theta}Ze^{i\theta}}(s,t) = E[Z(s)e^{i\theta}Z(t)e^{i\theta}] = Q_{ZZ}(s,t)e^{i2\theta}$ Thus the process $Z(t)e^{i\theta}$ satisfies the circular symmetric condition if and only if $Q_{ZZ}(s,t) = 0$ for all $t, s$. Let

$$Z(t)e^{i\theta} = X(t)\cos\theta - Y(t)\sin\theta + i(X(t)\sin\theta + Y(t)\cos\theta)$$
$$\triangleq U(t) + iV(t).$$

If $Z(t) = X(t) + iY(t)$ and $Z(t)e^{i\theta} = U(t) + iV(t)$ have the same distribution, their $2\times 2$-covariance function matrices must be the same. Let the four elements of the matrix be

$$E[X(s)X(t)] \triangleq A(s,t), \quad E[X(s)Y(t)] \triangleq B(s,t)$$
$$E[Y(s)X(t)] = B(s,t), \quad E[Y(s)Y(t)] \triangleq D(s,t).$$

Then, the following relation must hold for any $\theta$.

$$E[U(s)U(t)] = E[(X(s)\cos\theta - Y(s)\sin\theta)(X(t)\cos\theta - Y(t)\sin\theta)\top]$$
$$= \cos^2\theta A(s,t) - \sin\theta\cos\theta(B(s,t) + B(s,t)) + \sin^2\theta D(s,t) = A(s,t), \quad (2)$$
$$E[U(s)V(t)] = \sin\theta\cos\theta(A(s,t) - D(s,t)) - \sin^2\theta B(s,t) + \cos^2\theta B(s,t) = B(s,t). \quad (3)$$
$$E[V(s)U(t)] = \sin\theta\cos\theta(A(s,t) - D(s,t)) - \sin^2\theta B(s,t) + \cos^2\theta B(s,t) = B(s,t), \quad (4)$$
$$E[V(s)V(t)] = \sin^2\theta A(s,t) + \sin\theta\cos\theta(B(s,t) + B(s,t)) + \cos^2\theta D(s,t) = D(s,t). \quad (5)$$

If we set $\theta = \pi/2$ in (2), then $D(s,t) = A(s,t)$. Using this result and setting $\theta = \pi/2$ in (3), we obtain $B(s,t) = -B(s,t)$. Then (12.31) holds.
Conversely if (12.31) holds, then $D(s,t) = A(s,t)$ and $B(s,t) = -B(s,t)$ must hold. The equations (2) through (5) hold for all $\theta$, which implies that the distribution of $Z(t)e^{i\theta}$ is invariant under $\theta$.

**12.5 Brick-wall filter and the sampling function**

$$h(t) = \int_{-\infty}^{\infty} H(f)e^{i2\pi ft}\,df = \int_{-W}^{W} e^{i2\pi ft}\,df = \frac{e^{i2\pi ft} - e^{-i2\pi ft}}{i2\pi t}$$
$$2W\frac{\sin(2\pi Wt)}{2\pi Wt} = 2W\mathrm{sinc}(2Wt),$$

where $\text{sinc}(x) = \frac{\sin \pi x}{\pi x}$. The inverse Fourier transform gives

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-i2\pi f t} \, dt.$$

By setting $f = 0$ in the above, we find

$$1 = \int_{-\infty}^{\infty} 2W \text{sinc}(2Wt) \, dt.$$

By setting $2Wt = x$, we have

$$1 = \int_{-\infty}^{\infty} \text{sinc}(x) \, dx.$$

Since the sinc function is symmetric around $x = 0$, we have

$$\int_{0}^{\infty} \text{sinc}(x) \, dx = 1/2.$$

**12.6 Properties of the Hilbert transform.**

(a) **Hermitian symmetry.**

$$\begin{aligned} F(\omega) &= \int_{-\infty}^{\infty} f(t) e^{-i\omega t} \, dt F(-\omega) = \int_{-\infty}^{\infty} f(t) e^{i\omega t} \, dt \\ &= F^*(\omega) \end{aligned}$$

Since $\hat{f}(t)$ is also a real-valued function, its Fourier transform satisfies the same property as that of $f(t)$.

(b) **Orthogonality of $\hat{g}(t)$ and $g(t)$.** Using the Parseval's formula, we have

$$\begin{aligned} \langle g(t), \hat{g}(t) \rangle &= \int_{-\infty}^{\infty} g(t) \hat{g}^*(t) \, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) \hat{G}^*(\omega) \, d\omega \\ \frac{1}{2\pi} \int_{-\infty}^{\infty} G(\omega) [-i\text{sgn}(\omega) G(\omega)]^* \, d\omega &= \frac{i}{2\pi} \int_{-\infty}^{\infty} |G(\omega)|^2 \text{sgn}(\omega) \, d\omega = 0, \end{aligned}$$

since $|G(\omega)|^2$ is a symmetric (or even) function of $\omega$, hence $|G(\omega)|^2 \text{sgn}(\omega)$ is a skew-symmetric (or odd) function.

(c) **Hilbert transform of a symmetric function.** If $g(t)$ is symmetric around $t = 0$, then

$$\begin{aligned} \hat{g}(-t) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(-t-u)}{u} \, du \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(t+u)}{u} \, du = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{g(t-u')}{u'} \, du' = -\hat{g}(t). \end{aligned}$$

(d)

$$\begin{aligned} R_{\hat{X}X}(\tau) &= E[\hat{X}(t+\tau) X(t)] = E\left[ \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{X(t+\tau-u)}{u} \, du X(t) \right] \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{E[X(t+\tau-u) X(t)]}{u} \, du = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{R_{XX}(\tau-u)}{u} \\ &= \hat{R}_{XX}(\tau), \end{aligned}$$

which is (12.58). Similarly

$$\begin{aligned}
R_{\hat{X}\hat{X}}(\tau) &= E[\hat{X}(t+\tau)\hat{X}(t)] = E\left[\hat{X}(t+\tau)\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{X(t-u)}{u}\,du\right] \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{E[\hat{X}(t+\tau)X(t-u)]}{u}\,du \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{R_{\hat{X}X}(\tau+u)}{u}\,du = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\hat{R}_{XX}(\tau-u')}{u'}\,du'
\end{aligned}$$

$$= R_{XX}(\tau),$$

which is (12.56), where we used the formula (12.53), i.e., $f(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\hat{f}(t-u)}{u}\,du$

$$\begin{aligned}
R_{X\hat{X}}(\tau) &= E[X(t+\tau)\hat{X}(t)] = E[\hat{X}(t)X(t+\tau)] = R_{\hat{X}X}(-\tau) \\
&= -R_{\hat{X}X}(\tau),
\end{aligned}$$

where we used the property (c), since $R_{\hat{X}X}(\tau)$ is a symmetric function of $\tau$.

# 13 Solutions for Chapter 13: Spectral Representation of Random Processes and Time Series

## 13.1 Generalized Fourier Series Expansion

**13.1* Parseval's identity.** Using

$$G^*(f) = \int_{-\infty}^{\infty} g^*(s)e^{2\pi ft}\,ds$$

we have

$$\int_{-\infty}^{\infty} |G(f)|^2\,df = \int_{-\infty}^{\infty} G(f)G^*(f)\,df = \int_{f=-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(t)e^{-i2\pi ft}\,dt \int_{-\infty}^{\infty} g^*(s)e^{2\pi fs}\,ds\right] df$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{f=-\infty}^{\infty} e^{i2\pi(s-t)}\,df\right] g(t)g^*(s)\,dt\,ds$$
$$= \int_{t=-\infty}^{\infty}\left[\int_{s=-\infty}^{\infty} \delta(s-t)g^*(s)\,ds\right] g(t)\,dt$$
$$= \int_{t=-\infty}^{\infty} g^*(t)g(t)\,dt = \int_{-\infty}^{\infty} |g(t)|^2\,dt.$$

**13.2 Periodic WSS random process.**

$$P_m^{(N)} = \frac{1}{N}\left(\sum_{n=1}^{N}(X_n - \overline{X}^{(N)})W^{mn}\right)\left(\sum_{j=1}^{N}(X_j - \overline{X}^{(N)})W^{-mj}\right)$$
$$= \frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{N}(X_n - \overline{X}^{(N)})(X_j - \overline{X}^{(N)})W^{m(n-j)}$$

Let $k = n - j$, or $n = j + k$. Then the summation over $n = 1 \backsimeq N,\ j = 1 \sim N$ can be replaced by summation $j = \max(1, k+1) \sim \min(N-k, N)$ and $k = -N+1 \sim N-1$. Thus,

$$P_m^{(N)} = \frac{1}{N}\sum_{k=-N+1}^{N-1}\ \sum_{\max(1,k+1)}^{\min(N-k,N)}(X_{j+k} - \overline{X}^{(N)})(X_j - \overline{X}^{(N)})$$
$$= \frac{1}{N}\sum_{k=-N+1}^{N-1}(N-|k|)\hat{C}_X^{(N)}[k]W^{mk} = \sum_{k=-N+1}^{N-1}\left(1 - \frac{k}{N}\right)\hat{C}_X^{(N)}[k]W^{mk}.$$

**13.3 Equivalence of random variables.**

If $X \overset{m.s.}{=} Y$, then for $\delta > 0$, we have

$$E[|X - Y|^2] < \delta.$$

Then from Markov's inequality, for any $\epsilon > 0$, we have

$$P[|X - Y| > \epsilon] = P[|X - Y|^2 > \epsilon^2] \leq \frac{E[|X - Y|^2}{\epsilon^2} < \frac{\delta^2}{\epsilon^2} \triangleq \delta',$$

where $\delta' > 0$ can be chosen arbitrarily small by choosing sufficiently small $\delta$.
Conversely, if $X \overset{a.s.}{Y}$, then for any $\delta > 0, \epsilon > 0$, $X$ and $Y$ should satisfy

$$P[|X - Y| > \epsilon] < \delta.$$

Then,

$$E[|X - Y|^2] = \int |x - y|^2 dF_{X,Y}(x, y) = \int_{|x-y|>\epsilon} |x - y|^2 dF_{XY}(xy) + \int_{|x-y|<\epsilon} |x - y|^2 dF_{XY}(x, y)$$
$$\max |X - Y|^2 \delta + \epsilon^2.$$

As long as $\max |X - Y|$ is finite, the above quantity can be made arbitrarily small by proper choice of $\delta$ and $\epsilon$. Then, we have proven that $X \overset{m.s.}{=} Y$.

**13.4* Orthogonality of Fourier expansion coefficients of a periodic WSS process.**
In order to prove the second orthogonality (13.28), we expand the periodic $R(\tau)$ using the Fourier series:

$$R(\tau) = \sum_{k=-\infty}^{\infty} r_k e^{i2\pi f_0 k\tau}, \quad -\infty < \tau < \infty,$$

where

$$r_k = \frac{1}{T} \int_0^T R(\tau) e^{-i2\pi f_0 k\tau}\, d\tau.$$

Then

$$\begin{aligned}
E[X_m^* X_n] &= E\left[\frac{1}{T}\int_0^T e^{i2\pi f_0 mt} X^*(t)\, dt \frac{1}{T}\int_0^T e^{-i2\pi f_0 ns} X(s)\, ds\right] \\
&= \frac{1}{T^2}\int_0^T \int_0^T e^{i2\pi f_0 mt} e^{-i2\pi f_0 ns} E[X(s)X^*(t)]\, ds\, dt \\
&= \frac{1}{T^2}\int_0^T \int_0^T e^{i2\pi f_0 mt} e^{-i2\pi f_0 ns} R_X(s - t)\, ds\, dt \\
&= \frac{1}{T^2}\int_0^T \int_0^T e^{i2\pi f_0 mt} e^{-i2\pi f_0 ns} \left(\sum_{k=-\infty}^{\infty} r_k e^{i2\pi f_0 k(s-t)}\right) ds\, dt \\
&= \sum_{k=-\infty}^{\infty} r_k \left(\frac{1}{T}\int_0^T e^{i2\pi f_0 (m-k)t}\, dt\right)\left(\frac{1}{T}\int_0^T e^{-i2\pi f_0 (n-k)s}\, ds\right) \\
&= \sum_{k=-\infty}^{\infty} r_k \delta_{m,k}\delta_{n,k} = r_n \delta_{m,n},
\end{aligned}$$

where we used (13.27) in the last step.

**13.5 Derivation of (13.59)).**

$$P_m^{(N)} = \frac{1}{N}\left|\sum_{n=0}^{N-1}(x_n - \overline{x}^{(N)})W^{mn}\right|^2$$
$$= \frac{1}{N}\sum_{n=0}^{N-1}\sum_{n'=0}^{N-1}(x_n - \overline{x}^{(N)})(x_{n'} - \overline{x}^{(N)})W^{mn}W^{-mn'}$$

Writing $n = n' + k$, the summation over $n, n'$ can be changed to the summation over $n', k$.

$$P_m^{(N)} = \frac{1}{N}\sum_{k=-N+1}^{N-1}\sum_{n'=0}^{N-1}(x_{n'+k} - \overline{x}^{(N)})(x_{n'} - \overline{x}^{(N)})W^{mk}$$
$$= \frac{1}{N}\sum_{k=-N+1}^{N-1}\left(\sum_{n'=\max\{-k,0\}}^{\min\{N-k-1,N\}}(x_{n'+k} - \overline{x}^{(N)})(x_{n'} - \overline{x}^{(N)})\right)W^{mk}$$
$$= \sum_{k=-N+1}^{N-1}\frac{N-|k|}{N}\frac{1}{N-|k|}\left(\sum_{n'=\max\{-k,0\}}^{\min\{N-k-1,N\}}(x_{n'+k} - \overline{x}^{(N)})(x_{n'} - \overline{x}^{(N)})\right)W^{mk}$$
$$= \sum_{k=-N+1}^{N-1}\left(1 - \frac{|k|}{N}\right)\hat{C}_X^{(N)}[k]W^{mk},$$

where

$$\hat{C}_X^{(N)}[k] = \frac{1}{N-|k|}\left(\sum_{n'=\max\{-k,0\}}^{\min\{N-k-1,N\}}(x_{n'+k} - \overline{x}^{(N)})(x_{n'} - \overline{x}^{(N)})\right), \quad -N+1 \le k \le N-1.$$

**Note:** The summation in (13.60) should be changed to $\sum_{n'=\max\{-k,0\}}^{\min\{N-k-1,N\}}$ as given above.

## 13.2 Generalized Fourier Series Expansion

**13.6 Nonnegative definite matrix.**

$$\boldsymbol{a}^H\boldsymbol{R}\boldsymbol{a} = \boldsymbol{a}^H E[\boldsymbol{X}\boldsymbol{X}^H]\boldsymbol{a} = E[\langle\boldsymbol{a},\boldsymbol{X}\rangle\langle\boldsymbol{X},\boldsymbol{a}\rangle].$$

Since $\langle\boldsymbol{X},\boldsymbol{a}\rangle = \langle\boldsymbol{a},\boldsymbol{X}\rangle^*$, we have

$$\boldsymbol{a}^H\boldsymbol{R}\boldsymbol{a} = |\langle\boldsymbol{a}\boldsymbol{X}\rangle|^2 \ge 0.$$

**13.7 Nonnegativity of spectrum.** Multiply $\boldsymbol{u}^H$ to the both sides of (13.75):

$$\boldsymbol{u}^H\boldsymbol{R}\boldsymbol{u} = \lambda\boldsymbol{u}^H\boldsymbol{u} = \lambda.$$

The left side is nonnegative because $\boldsymbol{R}$ is nonnegative definite as shown in Exercise 13.6. Hence $\lambda \geq 0$. This is true for any eigenvalue $\lambda_n$.

**13.8 Left and right eigenvectors.** By definition

$$\boldsymbol{R}\boldsymbol{u} = \lambda \boldsymbol{u}.$$

Take transpose and complex conjugation of the above equation.

$$\boldsymbol{u}^H \boldsymbol{R}^H = \lambda \boldsymbol{u}^H.$$

But the matrix $\boldsymbol{R}$ is self-adjoint as given in (13.72). Thus,

$$\boldsymbol{u}^H \boldsymbol{R} = \lambda \boldsymbol{u}^H.$$

Thus, $\boldsymbol{u}^H$ is the left-eigenvector.

**13.9* Orthogonality of eigenvectors.** Note that

$$\boldsymbol{u}_i^H \boldsymbol{R} \boldsymbol{u}_j = \lambda_i \boldsymbol{u}_i^H \boldsymbol{u}_j,$$

since $\boldsymbol{u}_i^H$ is a left-eigenvector of $\boldsymbol{R}$. Also

$$\boldsymbol{u}_i^H \boldsymbol{R} \boldsymbol{u}_j = \boldsymbol{u}_i^H \lambda_j \boldsymbol{u}_j,$$

since $\boldsymbol{u}_j$ is a right eigenvector. Taking the difference of the above two equations, we have,

$$(\lambda_i - \lambda_j)\boldsymbol{u}_i^H \boldsymbol{u}_j = 0.$$

Since $\lambda_i \neq \lambda_j$, it follows that $\boldsymbol{u}^H \boldsymbol{u}_j = 0$.

**13.10 Variance of expansion coefficients in the eigenvector expansion.**

$$\begin{aligned} E[\chi_m \chi_n^*] &= E[\boldsymbol{u}_m^H \boldsymbol{X}\boldsymbol{X}^H \boldsymbol{u}_n] = E[\boldsymbol{u}_m^H \boldsymbol{R} \boldsymbol{u}_n] \\ &= E[\boldsymbol{u}_m^H \lambda_n \boldsymbol{u}_n] = \lambda_n \delta_{m,n} \end{aligned}$$

**13.11 Mean square error of eigenvector expansion approximation.**

$$\begin{aligned} \mathcal{E}^2 = E\left[\left|\sum_{n=M+1}^{N} \chi_n \boldsymbol{u}_n\right|^2\right] &= E\left[\left(\sum_{m=M+1}^{N} \chi_m \boldsymbol{u}_m\right)^H \left(\sum_{n=M+1}^{N} \chi_n \boldsymbol{u}_n\right)\right] \\ = \sum_{m=M+1}^{N} \sum_{n=M+1}^{N} \boldsymbol{u}_m^H E[\chi_m^* \chi_n] \boldsymbol{u}_n &= \sum_m \sum_n \boldsymbol{u}^H \lambda_n \delta_{m,n} \boldsymbol{u}_n \\ = \sum_n \lambda_n \boldsymbol{u}_n^H \boldsymbol{u}_n &= \sum_{n=M+1}^{N} \lambda_n \end{aligned}$$

**13.12* Eigenvectors and eigenvalues of a circulant matrix.**

(a) Consider the matrix equation $\boldsymbol{C}\boldsymbol{u} = \lambda \boldsymbol{u}$. Expand this equation and consider the $j$th row:

$$c_{n-j}u_0 + c_{n-j+1}u_1 + \cdots + c_{n-1}u_{j-1} + c_0 u_j + c_1 u_{j+1} + \cdots + c_{n-j-1}u_{n-1} = \lambda u_j,$$

which gives

$$\sum_{k=n-j}^{n-1} c_k u_{k-n+j} + \sum_{k=0}^{n-j-1} c_k u_{k+j} = \lambda u_j.$$

(b) Substituting $u_j = \alpha^j$ into the above, we have

$$\sum_{k=n-j}^{n-1} c_k \alpha^{k-n+j} + \sum_{k=0}^{n-j-1} c_k \alpha^{k+j} = \lambda \alpha^j.$$

By dividing both sides by $\alpha^j$, we have

$$\alpha^{-n} \sum_{k=n-j}^{n-1} c_k \alpha^k + \sum_{k=0}^{n-j-1} c_k \alpha^k = \lambda.$$

(c) If $\alpha^n = 1$, then the last equation becomes

$$\lambda = \sum_{k=0}^{n-1} c_k \alpha^k.$$

Equation $\alpha^n = 1$ has $n$ distinct complex roots:

$$\alpha_m = e^{\frac{i2\pi m}{n}} = W^m, \quad m = 0, 1, 2, \ldots, n-1. \tag{1}$$

Then the $m$th eigenvalue is

$$\lambda_m = \sum_{k=0}^{n-1} c_k W^{km}, \tag{2}$$

and the $m$th eigenvector is

$$\boldsymbol{u}_m = (\alpha_m^0, \alpha_m, \alpha_m^2, \ldots, \alpha_m^{n-1})^\top = (1, W^m, W^{2m}, \ldots, W^{(n-1)m})^\top.$$

From (2), we can write $c_k$ in terms of $\lambda_m$'s, i.e.,

$$c_k = \frac{1}{n} \sum_{m=0}^{n-1} \lambda_m W^{-km},$$

which is the inverse DFT.

**Note:** The more common definition of the DFT and the inverse DFT may be

$$\lambda_m = \sum_{k=0}^{n-1} c_k W^{-km},$$

and

$$c_k = \frac{1}{n} \sum_{m=0}^{n-1} \lambda_m W^{km}.$$

This can be obtained by expressing the $n$ distinct complex roots as

$$\alpha_m = e^{-\frac{i2\pi m}{n}} = W^{-m}.$$

**Alternative proof:**
It is easy to verify that a matrix is circulant, if and only if it can be expressed as the following matrix polynomial:

$$\boldsymbol{C} = c_0\boldsymbol{I} + c_1\boldsymbol{V} + \ldots + c_{n-1}\boldsymbol{V}^{n-1} \tag{3}$$

where $\boldsymbol{I}$ is the identity matrix and

$$\boldsymbol{V} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

is the *cyclic permutation matrix* (also called the *elementary circulant matrix*).
Let $\alpha$ and $\boldsymbol{u}$ denote an eigenvalue and its corresponding eigenvector of $\boldsymbol{V}$, i.e.,

$$\boldsymbol{V}\boldsymbol{u} = \alpha\boldsymbol{u}. \tag{4}$$

Then $\boldsymbol{u}$ is also an eigenvector of $\boldsymbol{C}$, because

$$\boldsymbol{C}\boldsymbol{u} = \sum_{k=0}^{n-1} c_k \boldsymbol{V}^k \boldsymbol{u} = \sum_{i=0}^{n-1} c_k \alpha^k \boldsymbol{u}.$$

Thus, the corresponding eigenvalue of $\boldsymbol{C}$ is

$$\lambda = \sum_{k=0}^{n-1} c_k \alpha^k. \tag{5}$$

So the problem of finding $n$ eigenvectors and eigenvalues of $\boldsymbol{C}$ reduces to that of finding those of the cyclical permutation matrix $\boldsymbol{V}$.
Let $u_i$ represent the $i$th element of the vector $\boldsymbol{u}$, i.e.,

$$\boldsymbol{u} = (u_0, u_1, \ldots, u_{n-1})^\top. \tag{6}$$

Then from (4), we find

$$u_1 = \alpha u_0, u_2 = \alpha u_1 = \alpha^2 u_0, \ldots, u_i = \alpha u_{i-1} = \alpha^i u_0, \ldots, u_{n-1} = \alpha^{n-1} u_0. \tag{7}$$

From (4), we also find

$$\boldsymbol{V}^i \boldsymbol{u} = \alpha^i \boldsymbol{u}, \;\; i = 0, 1, 2, \ldots. \tag{8}$$

Note that the cyclical permutation matrix $\boldsymbol{V}$ satisfies $\boldsymbol{V}^n = \boldsymbol{I}$. By setting $i = n$, we find

$$\alpha^n = 1, \tag{9}$$

to be a necessary and sufficient condition for $\boldsymbol{u}$ to be a non-zero vector. There are $n$ distinct complex roots for $\alpha$, which are given by

$$\alpha_m = \exp\left(\frac{i2\pi m}{n}\right) = W^m, \;\; m = 0, 1, \ldots, n-1,$$

where $W = \exp\left(\frac{i2\pi}{n}\right)$, as defined in (1).

The $m$th eigenvector is found from equation (6) as

$$\boldsymbol{u}_m = (1, W^m, W^{2m}, \ldots, W^{m(n-1)})^\top, \;\; m = 0, 1, \ldots, n-1,$$

where we set $u_{m,0}$, the first component of the $\boldsymbol{u}_m$, to be unity for all $m$. The corresponding eigenvalues are found, from (5), as

$$\lambda_m = \sum_{k=0}^{n-1} c_k W^{mk}, \;\; m = 0, 1, \ldots, n-1,$$

which shows that the eigenvalues are the DFT of $(c_0, c_1, \ldots, c_{n-1})$.

**13.13 Generating function method.** Let $N \to \infty$ in (13.98):

$$\sum_{j=1}^{\infty} R_{i-j} v_{n,j} = \lambda_n v_{n,i}, \;\; i = 1, 2, \ldots.$$

Since the above is the convolution sum with respect to the subscripts $i$ and $j$, it should become a product if we take the generating function (just like in the $Z$-transform).
Multiply $z^i = z^{i-j} z^j$ on both sides and sum from $i = 1$ to $\infty$:

$$\sum_{i=1}^{\infty} \left( \sum_{j=1}^{\infty} R_{i-j} z^{i-j} \right) v_{n,j} z^j = \lambda_n \sum_{i=1}^{\infty} v_{n,i} z^i = \lambda_n v_n(z).$$

By letting $i - j = k$, the LHS becomes

$$\text{LHS} = \sum_{k=-\infty}^{\infty} R_k z^k \sum_{j=1}^{\infty} v_{n,j} z^j = R(z) v(z).$$

Thus we have shown (13.104).

**13.14 Orthogonality of eigenfunctions.** We can prove the orthogonality of eigenfunction in the same way we proved orthogonality of eigenvectors in Problem 13.9. Note

$$\int_0^T \int_0^T v_n^*(t) R_X(t,s) v_m(s)\, dt\, ds = \lambda_n \int_0^T v_n^*(s) v_m(s)\, ds.$$

Similarly

$$\int_0^T \int_0^T v_n^*(t) R_X(t,s) v_m(s)\, dt\, ds = \lambda_m \int_0^T v_n^*(t) v_m(t)\, dt.$$

Then taking the difference of the above, we have

$$0 = (\lambda_n - \lambda_m) \int_0^T v_n^*(s) v_m(s)\, ds.$$

Since $\lambda_n - \lambda_m \neq 0$, it must follows that $\int_0^T v_n^*(s) v_m(s)\, ds = 0$.

**13.15 Derivation of the integral equation (13.136).**

$$R_N(t,u) = \sum_{k=1}^{\infty} \lambda_k v_k(t) v^*(u).$$

and

$$Q(u) = \sum_{j=1}^{\infty} q_j v_j(u).$$

Then,

$$\begin{aligned}\int_0^T R_N(t,u)Q(u)\,du &= \sum_k \sum_j \lambda_k v_k(t) q_j \int_0^T v_k^*(u) v_j(u)\,du \\ &= \sum_k \sum_j \lambda_k q_j \delta_{k,j} v_k(t) = \sum_k \lambda_k q_k v_k(t) \\ &= \sum_k s_k v_k(t) = S(t).\end{aligned}$$

**13.16 Matched filter equivalent to a correlation receiver.**
From (13.137), we have

$$Q^*(t) = h(T-t), \;\; 0 \le t \le T.$$

Hence,

$$\begin{aligned}\int_0^T Q^*(t)X(t)\,dt &= \int_0^T h(T-t)X(t)\,dt = \int_0^T h(s)X(T-s)\,ds \\ &= \int_0^t h(s)X(t-s)\,dt\bigg|_{t=T}.\end{aligned}$$

**13.17* Matched filter and SNR.** We assume that the signal duration interval is $[0,T]$. Otherwise, replace the integration $\int_0^T$ below by $\int_{-\infty}^{\infty}$ throughout.

(a)

$$S_0(t) = \int_0^T h(u)S(t-u)\,du.$$

Thus,

$$P_S = |S_0(T)|^2 = \left|\int_0^T h(u)S(T-u)\,du\right|^2$$

$$N_0(t) = \int_0^T h(u)N(t-u)\,du.$$

Thus,

$$\begin{aligned}P_N = E[|N_0(t)|^2] &= \int_0^T \int_0^T h(u)E[N(t-u)N^*(t-v)]h^*(v)\,dv \\ &= \int_0^T \int_0^T \sigma^2 \delta(v-u)h(u)h^*(v)\,du\,dv = \sigma^2 \int_0^T |h(u)|^2\,du.\end{aligned}$$

(b)

$$\text{SNR} = \frac{|\int_0^T h(u)S(T-u)\,du|^2}{\sigma^2 \int_0^T |h(u)|^2\,du}.$$

Using the Cauchy-Schwartz inequality $|\langle X, Y\rangle|^2 \le |X|^2|Y|^2$, we have

$$\begin{aligned}
\text{SNR} &\le \frac{\int_0^T |S(T-u)|^2\,du \int_0^T |h(u)|^2\,du}{\sigma^2 \int_0^T |h(u)|^2\,du} \\
&= \frac{1}{\sigma^2}\int_0^T |S(T-u)|^2\,du = \frac{E_S}{\sigma^2}
\end{aligned}$$

where the equality holds when

$$h(u) = kS^*(T-u)$$

with some constant $k$. $E_S$ is the signal energy: $E_S = \int_0^T |S(t)|^2\,dt$.
(c) Define $P_N = E[|N_0(T)|^2]$. Then

$$\begin{aligned}
P_N &= \int_0^T \int_0^T h(u)E[N(t-u)N^*(t-v)]h^*(v)\,du\,dv \\
&= \int_0^T \int_0^T R_N(T-u, T-v)h(u)h^*(v)\,du\,dv.
\end{aligned}$$

Hence

$$\text{SNR} = \frac{|\int_0^T h(u)S(T-u)\,du|^2}{\int_0^T \int_0^T R_N(T-u, T-v)h(u)h^*(v)\,du\,dv}$$

Find $h(t)$ that maximizes SNR. To simplify the presentation we use the following vector and matrix representation.

$$h(u) \longrightarrow \boldsymbol{h}, \;\; S(t-u) \longrightarrow \boldsymbol{S}, \;\; R_N(T-u, T-v) \longrightarrow \boldsymbol{R}_N.$$

Then

$$\begin{aligned}
\text{SNR} &= \frac{|\boldsymbol{h}^\top \boldsymbol{S}|^2}{\boldsymbol{h}^\top \boldsymbol{R}_N \boldsymbol{h}^*} = \frac{|(\boldsymbol{R}_N^{1/2}\boldsymbol{h})^\top(\boldsymbol{R}_N^{-1/2}\boldsymbol{S})|^2}{(\boldsymbol{R}^{1/2}\boldsymbol{h})^\top(\boldsymbol{R}_N^{1/2}\boldsymbol{h}^*)} \\
&\le \frac{\|\boldsymbol{R}_N^{1/2}\boldsymbol{h}\|^2\|\boldsymbol{R}_N^{-1/2}\boldsymbol{S}\|^2}{\|\boldsymbol{R}^{1/2}\boldsymbol{h}\|^2} = \|\boldsymbol{R}_N^{-1/2}\boldsymbol{S}\|^2 \\
&= \boldsymbol{S}^\top \boldsymbol{R}_N^{-1}\boldsymbol{S}^*,
\end{aligned}$$

where the equality holds if and only if (by setting an arbitrary scaling constant to be one)

$$\boldsymbol{R}_N^{1/2}\boldsymbol{h} = (\boldsymbol{R}_N^{-1/2}\boldsymbol{S})^*,$$

or

$$\boldsymbol{h} = \boldsymbol{R}_N^{-1}\boldsymbol{S}^*, \;\text{ or }\; \boldsymbol{R}_N\boldsymbol{h} = \boldsymbol{S}^*,$$

Thus, the matched filter $h(u)$ must satisfy the integral equation

$$\int_0^T R_N(t,u)h(u)\,du = S^*(T-t),$$

which is equivalent to the equation for $Q(u)$ of (13.136), with $h(t) = Q^*(T-t)$.
Note: The square root of $\boldsymbol{R}_N$ that appeared in the derivation corresponds to

$$\boldsymbol{R}_N^{1/2} \longrightarrow \sum_{k=1}^{\infty} \sqrt{\lambda_k} v_k(t) v_k(s).$$

Similarly,

$$\boldsymbol{R}_N^{-1/2} \longrightarrow \sum_{k=1}^{\infty} \frac{1}{\sqrt{\lambda_k}} v_k(t) v_k(s).$$

**13.18* Orthogonal expansion of Wiener process (need corrections).**

(a)

$$\sigma^2 \int_0^T \min(t,s)\psi(s)\,ds = \lambda\psi(t).$$

Dividing $[0,T]$ into $[0,t]$ and $(t,T]$,

$$\sigma^2 \left( \int_0^t s\psi(s)\,ds + t\int_t^T \psi(s)\,ds \right) = \lambda\psi(t).$$

Differentiate both sides with respect to $t$:

$$\sigma^2 \left( t\psi(t) + \int_t^T \psi(s)\,ds - t\psi(t) \right) = \lambda\psi'(t).$$

Differentiate again

$$\sigma^2 \left(\psi(t) + t\psi'(t) - \psi(t) - \psi(t) - t\psi'(t)\right) = \lambda\psi^{''}(t).$$

Hence,

$$-\sigma^2\psi(t) = \lambda\psi^{''}(t).$$

(b) If $\lambda < 0$, then

$$\psi^{''}(t) - a^2\psi(t) = 0, \quad \text{where } a^2 = \frac{\sigma^2}{-\lambda}.$$

Then the solutions of this differential equation are known to be

$$\psi(t) = C_1 e^{at} \triangleq \psi_1(t), \text{ and } \psi(t) = C_2 e^{-at} \triangleq \psi_2(t).$$

If we insert $\psi_1(t)$ (by setting $C_1 = 1$ to simplify the matter) into the integral equation, we have

$$a^2 \int_0^T \min(t,s) e^{as}\,ds = -e^{at}.$$

By splitting the integration interval into two parts,

$$a^2 \left( \int_0^t s e^{as} \, ds + t \int_t^T e^{as} \, ds \right) = -e^{at}.$$

Then

$$\begin{aligned} \text{LHS} &= a^2 \left( \int_0^t s \left( \frac{e^{as}}{a} \right)' ds + t \left[ \frac{e^{as}}{a} \right]_t^T \right) \\ &= a^2 \left( \left[ \frac{s e^{as}}{a} \right]_0^t - \int_0^t \frac{e^{as}}{a} \, ds + \frac{t(e^{aT} - e^{at})}{a} \right) \\ &= a^2 \left( \frac{t e^{at}}{a} - \frac{e^{at} - 1}{a^2} + \frac{t e^{aT} - t e^{at}}{a} \right) \\ &= a t e^{at} - e^{at} + 1 + a t e^{aT} - a t e^{at} = -e^{at} + a t e^{aT} - 1. \end{aligned}$$

The LHS equals the RHS $(-e^{at})$ only if $e^{aT} at - 1 = 0$ for all $t$, which does not hold. Hence $\psi_1(t)$ cannot be a solution for any real number $a$. Similarly $\psi_2(t) = e^{-at}$ cannot be a solution of the integral equation. Hence no solution exists.

(c) Let $\frac{\sigma^2}{\lambda} = \omega^2$. Then the solutions of the differential equation (13.247) are

$$\psi(t) = C_1 e^{i\omega t} + C_2 e^{-i\omega t}.$$

Then substituting this into (13.246),

$$\sigma^2 \left[ \int_0^t s \left( C_1 e^{i\omega s} + C_2 e^{-i\omega s} \right) ds + t \int_t^T \left( C_1 e^{i\omega s} + C_2 e^{-i\omega s} \right) ds \right] = \lambda \left( C_1 e^{i\omega t} + C_2 e^{-i\omega t} \right).$$

Dividing both sides by $\lambda$ and performing the integration, we have

$$\begin{aligned} \text{LHS} &= \omega^2 C_1 \left( \frac{t e^{i\omega t}}{i\omega} - \frac{e^{i\omega t} - 1}{(i\omega)^2} + \frac{t e^{i\omega T} + i\omega t e^{i\omega t}}{i\omega} \right) \\ &+ \omega^2 C_2 \left( \frac{t e^{-i\omega t}}{-i\omega} - \frac{e^{-i\omega t} - 1}{(i\omega)^2} + \frac{t e^{-i\omega T} - i\omega t e^{-i\omega t}}{i\omega} \right) \\ &= C_1(-1 - i\omega t e^{i\omega T}) + C_2(-1 + i\omega t e^{-\omega T}) \\ &= -(C_1 + C_2) - i\omega t (C_1 e^{i\omega T} - C_2 e^{-\omega T}). \end{aligned}$$

This equals the RHS $(C_1 e^{i\omega t} + C_2 e^{-i\omega t})$, if and only if

$$C_1 + C_2 = 0, \text{ and } e^{i\omega T} + e^{-i\omega T} = 2 \cos \omega T = 0.$$

Hence,

$$\omega T = \frac{\pi}{2} + n\pi = \frac{(2n+1)\pi}{2}.$$

Thus,

$$\omega_n = \frac{(2n+2)\pi}{2T}, \quad n = 0, \pm 1, \pm 2, \ldots.$$

Thus, the eigenvalues are

$$\lambda_n = \frac{\sigma^2}{\omega_n^2}, \quad n = 0, \pm 1, \pm 2, \ldots.$$

The corresponding eigenfunctions are

$$v_n(t) = C_1 e^{i\omega_n t} - C_1 e^{-i\omega_n t} = i2C_1 \sin \omega_n t.$$

From the normalization requirement $\int_0^T |v_n(t)|^2 \, dt = 1$, we find

$$v_n(t) = \sqrt{\frac{2}{T}} \sin \omega_n t.$$

(d) The K-L expansion coefficients are

$$W_n = \int_0^T \psi_n(t) W(t) \, dt = \sqrt{\frac{2}{T}} \int_0^T \sin \omega_n t W(t) \, dt.$$

From the theory of K-L expansion, we know the set of $\psi_n(t), \quad n = 0, \pm 1, \pm 2, \ldots$ are orthogonal.

$$\begin{aligned} E[W_n^2] &= \frac{2}{T} \int_0^T \int_0^T \sin \omega_n t \omega_n s E[W(t)W(s)] \, dt \, ds \\ &= \frac{2\sigma^2}{T} \int_0^T \int_0^T \min(t,s) \sin \omega_n t \sin \omega_n s \, dt \, ds \end{aligned}$$

Now, we evaluate

$$\begin{aligned} &\int_0^T \min(t,s) \sin \omega_n s \, ds = \int_0^t s \sin \omega_n s \, ds + t \int_t^T \sin \omega_n s \, ds \\ &= \int_0^t s \left( -\frac{\cos \omega_n s}{\omega_n} \right)' ds + t \int_t^T \sin \omega_n s \, ds \\ &= -\left[ \frac{s \cos \omega_n t}{\omega_n} \right]_0^t + \int_0^t \frac{\cos \omega_n s}{\omega_n} \, ds - t \left[ \frac{\cos \omega_n s}{\omega_n} \right]_t^T \\ &= -\frac{t \cos \omega_n t}{\omega_n} + \frac{1}{\omega_2} [\sin \omega_n s]_0^t + t \frac{\cos \omega_n t}{\omega_n} \\ &= \frac{\sin \omega_n t}{\omega^2}. \end{aligned} \tag{10}$$

Thus,

$$E[W_n^2] = \frac{2}{T} \int_0^T \frac{\sin^2 \omega_n t}{\omega_n^2} \, dt = \int_0^T \frac{1 - \cos 2\omega_n t}{2\omega_n^2} = \frac{T}{2\omega_n^2} = \frac{4T^2\sigma^2}{(2n+1)^2\pi^2}.$$

Using the result of (10), we can directly show the orthogonality between $W_n$ and $W_m$ ($m \neq n$), as follows:

$$\begin{aligned} E[W_n W_m] &= \frac{2}{T}\int_0^T \int_0^T \sin\omega_n t \omega_m s E[W(t)W(s)]\,dt\,ds \\ &= \frac{2\sigma^2}{T}\int_0^T \sin\omega_n t \left(\int_0^T \min(t,s)\sin\omega_m s\,ds\right)\,dt \\ &= \frac{2\sigma^2}{T\omega^2}\int_0^T \sin\omega_n t \sin\omega_m t\,dt. \end{aligned}$$

Since

$$\int_0^T \sin\omega_n t \sin\omega_m t\,dt = 0, \quad \text{for } \ m \neq n,$$

we have proved the orthogonality.

(e) Note

$$\omega_{-n} = \frac{(-2n+1)\pi}{2T} = -\frac{(2n-1)\pi}{2T} = -\omega_{n-1}.$$

Hence the set of $\{\omega_n; n \geq 0\}$ is complete. So

$$W(t) = \sqrt{\frac{2}{T}}\sum_{n=-\infty}^{\infty} W_n \sin\omega_n t = \sqrt{\frac{2}{T}}\left(\sum_{n=-\infty}^{-1} W_n \sin\omega_n t + \sum_{n=0}^{\infty} W_n \sin\omega_n t\right).$$

The first term can be written as

$$\begin{aligned} \sum_{n=-\infty}^{-1} W_n \sin\omega_n t &= \sum_{m=1}^{\infty} W_{-m}\sin\omega_{-m} t = -\sum_{m=1}^{\infty} W_{-m}\sin\omega_{m-1} t \\ &= -\sum_{n=0}^{\infty} W_{-n-1}\sin\omega_n t. \end{aligned}$$

Hence,

$$W(t) = \sqrt{\frac{2}{T}}\sum_{n=0}^{\infty}(W_n - W_{-n-1})\sin\omega_n t \triangleq \sqrt{\frac{2}{T}}\sum_{n=0}^{\infty} U_n \sin\omega_n t,$$

where,

$$U_n = (W_n - W_{-n-1}).$$

Hence

$$\begin{aligned} E[U_n^2] &= E[W_n^2] + E[W_{-n-1}^2] \\ &= 4\left(\frac{\sigma T}{(2n+1)\pi}\right)^2 + 4\left(\frac{\sigma T}{(-2n-1)\pi}\right)^2 \\ &= 8\left[\frac{\sigma T}{(2n+1)\pi}\right]^2. \end{aligned}$$

## 13.3 PCA and SVD

**13.19 Orthogonality of the expansion coefficient vectors $\chi_i$'s.**

$$\langle \boldsymbol{\chi}_i, \boldsymbol{\chi}_{i'} \rangle = \boldsymbol{\chi}_i^\top \boldsymbol{\chi}_{i'}^* = \boldsymbol{u}_i^H \mathbf{X}\mathbf{X}^H \boldsymbol{u}_{i'} = \boldsymbol{u}_i^H \mu_{i'} \boldsymbol{u}_{i'} = \mu_{i'} \delta i, i',$$

where we use the fact that $\boldsymbol{u}_{'}$ is the eigenvector associated with the eigenvalue $\mu_{i'}$ of the matrix $\mathbf{X}\mathbf{X}^H$.

**13.20* Sum of squares of the difference.**

(a)

$$\sum_i \sum_j |a_{ij}|^2 = \sum_i \sum_j |b_i c_j|^2 = \sum_i |b_i|^2 \sum_j |c_j|^2 = \|\boldsymbol{b}\|^2 \|\boldsymbol{c}\|^2.$$

(b)

$$a_{ij} = b_i^{(1)} c_j^{(1)} + b_i^{(2)} c_j^{(2)}.$$

Thus

$$|a_{ij}|^2 = (b_i^{(1)} c_j^{(1)} + b_i^{(2)} c_j^{(2)})(b_i^{(1)} c_j^{(1)} + b_i^{(2)} c_j^{(2)})^* = |b_i^{(1)}|^2 |c_j^{(1)}|^2 + |b_i^{(2)}|^2 |c_j^{(2)}|^2,$$

because $\langle \boldsymbol{c}^{(1)}, \boldsymbol{c}^{(2)} \rangle = \langle \boldsymbol{c}^{(2)}, \boldsymbol{c}^{(1)} \rangle = 0$. Thus,

$$\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 = \|\boldsymbol{b}^{(1)}\|^2 \|\boldsymbol{c}^{(1)}\|^2 + \|\boldsymbol{b}^{(2)}\|^2 \|\boldsymbol{c}^{(2)}\|^2.$$

(c) By generalizing the result of part (b), we have that if

$$\boldsymbol{A} = \sum_{i=k+1}^m \boldsymbol{b}^{(i)} \boldsymbol{c}^{(i)\top},$$

Then

$$\|\boldsymbol{A}\|^2 = \sum_{i=k+1}^m \|\boldsymbol{b}^{(i)}\|^2 \|\boldsymbol{c}^{(i)}\|^2.$$

Now let
Let $\boldsymbol{b}^{(i)} = \boldsymbol{u}_i$ and $\boldsymbol{c}^{(i)} = \boldsymbol{\chi}_i, \;\; i = k+1, \ldots, m$. Then using $\|\boldsymbol{u}_i\|^2 = 1$, we have

$$\|\mathbf{X} - \hat{\mathbf{X}}\|^2 = \sum_{i=k+1}^m \|\boldsymbol{\chi}_i\|^2 = \sum_{i=k+1}^m \mu_i,$$

where we used (13.160) to find $\|\boldsymbol{\chi}_i\|^2 = \mu_i$.

**13.21 The covariance matrix based PCA.** Since

$$\overline{x}_1 = \frac{2+1+0}{3} = 1, \;\; \overline{x}_2 = \frac{4+3+0}{3} = \frac{7}{3} = 2.3333, \;\; \text{hence } \overline{\boldsymbol{x}} = \begin{bmatrix} 1 \\ 2.3333 \end{bmatrix}$$

$$\boldsymbol{C} = \frac{1}{n-1}(\mathbf{X}\mathbf{X}^\top - \overline{\boldsymbol{x}}\,\overline{\boldsymbol{x}}^\top).$$

Hence

$$(n-1)\boldsymbol{C} = \mathbf{X}\mathbf{X}^\top - \overline{\boldsymbol{x}}\,\overline{\boldsymbol{x}}^\top = \begin{bmatrix} 5 & 11 \\ 11 & 15 \end{bmatrix} - \begin{bmatrix} 1 & 7/3 \\ 7/3 & 29/9 \end{bmatrix} = \begin{bmatrix} 4 & 26/3 \\ 26/3 & 176/9 \end{bmatrix}$$

The characteristic equation

$$(4-\mu)\left(\frac{176}{9} - \mu\right) - \frac{26^2}{3^2} = 0,$$

where $\mu = n - 1\lambda = 2\lambda$ gives,

$$\mu^2 - 212\mu + 28 = 0,$$

from which we find two eigenvalues

$$\mu_1 = \frac{212 + \sqrt{212^2 - 36 \times 28}}{18} = 23.4227, \;\; \mu_2 = \frac{212 - \sqrt{212^2 - 36 \times 28}}{18} = 0.1328.$$

Note that these eigenvalues are somewhat smaller than those obtained in Example 13.4. The difference is due to $\boldsymbol{C} \leq \boldsymbol{R}$. The corresponding orthonormal eigenvectors are found as

$$\boldsymbol{u}_1 = \begin{bmatrix} 0.4075 \\ 0.9132 \end{bmatrix}, \;\; \boldsymbol{u}_2 = \begin{bmatrix} -0.9132 \\ 0.4075 \end{bmatrix}, \;\; \text{hence } \boldsymbol{U} = \begin{bmatrix} 0.4075 & -0.9132 \\ 0.9132 & 0.4075 \end{bmatrix}.$$

Hence, the PCA expansion coefficient vectors are obtained as

$$\boldsymbol{\chi}_1 = (\mathbf{X} - \overline{\boldsymbol{x}}\,\mathbf{1}^\top)^\top \boldsymbol{u}_1 = \begin{bmatrix} 1 & 1.6667 \\ 0 & 0.6667 \\ -1 & -2.3333 \end{bmatrix} \begin{bmatrix} 0.4075 \\ 0.9132 \end{bmatrix} = \begin{bmatrix} 1.9295 \\ 0.6088 \\ -2.5383 \end{bmatrix},$$

and

$$\boldsymbol{\chi}_2 = (\mathbf{X} - \overline{\boldsymbol{x}}\,\mathbf{1}^\top)^\top \boldsymbol{u}_2 = \begin{bmatrix} 1 & 1.6667 \\ 0 & 0.6667 \\ -1 & -2.3333 \end{bmatrix} \begin{bmatrix} -0.9132 \\ 0.4075 \end{bmatrix} = \begin{bmatrix} -0.2340 \\ 0.2717 \\ -0.0376 \end{bmatrix}.$$

**13.22 Derivation of SVD (13.184).** From (13.182), we have

$$\mathbf{X} = \sum_{i=1}^{m} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^H.$$

Multiply $\boldsymbol{v}_j$ from the right:

$$\mathbf{X}\boldsymbol{v}_j = \sum_{i=1}^{m} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^H \boldsymbol{v}_j = \sum_{i=1}^{m} \sigma_i \boldsymbol{u}_i \delta_{ij} = \sigma_j \boldsymbol{u}_j.$$

Taking the Hermitian of the first equation, we have

$$\mathbf{X}^H = \sum_{i=1}^{m} \sigma_i \boldsymbol{v}_i \boldsymbol{u}_i^H.$$

Multiplying $\boldsymbol{u}_j$ from the right, we have

$$\mathbf{X}^H \boldsymbol{u}_j = \sum_{i=1}^{m} \sigma_i \boldsymbol{v}_i \delta_{i,j} = \sigma_j \boldsymbol{v}_j.$$

**13.23 Derivation of $\boldsymbol{U}$ from $\boldsymbol{V}$.** Because $\boldsymbol{U}$ is a unitary matrix, we can write

$$\mathbf{X} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^H = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{V}^{-1}.$$

Thus,

$$\mathbf{X}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{\Sigma} = \boldsymbol{V}[\boldsymbol{M}^{1/2}\mathbf{0}_{m\times(n-m)}] = [\boldsymbol{V}\boldsymbol{M}^{1/2}\mathbf{0}_{m\times(n-m)}]$$

Hence

$$[\boldsymbol{U}\mathbf{0}_{m\times(n-m)}] = [\mathbf{X}\boldsymbol{V}]_{m\times m}\boldsymbol{M}^{-1/2}.$$

**13.24 Frobenius norm and singular values** The formula can be derived using the same argument we used in Problem 13.20.
Suppose we have the following SVD decomposition of matrix $\boldsymbol{A}$

$$\boldsymbol{A} = \sum_{i}^{\min(m,n)} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^H,$$

So from the solutions of Problem 13.20, we find

$$\|\boldsymbol{A}\|^2 = \sum_{i}^{\min(m,n)} \sigma^2 \|\boldsymbol{u}_i\|^2 \|\boldsymbol{v}_i\|^2.$$

Since both $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are normalized, $\|\boldsymbol{u}_i\| = \|\boldsymbol{v}_i\| = 1$ for all $i$. Hence we obtain (13.187).

## 13.4 ARMA

**13.25 Inclusion of a constant $a_0$ in AR($p$) of (13.192).** Take the expectation of (13.251):

$$E[Y_n] = a_0 + \sum_{i=1}^{p} a_i E[Y_i].$$

From the stationarity assumption $E[Y_n] = E[Y_i] = \mu_Y$, which implies

$$\mu_Y = \frac{a_0}{1 - \sum_{i=1}^{p} a_i},$$

provided $\sum_{i=1}^{p} a_i \neq 1$. After subtracting this expression from both sides of (13.251), and defining $X_n = Y_n - \mu_Y$, we obtain (13.192) and $E[X_n] = 0$.

**13.26* Mean square convergence of (13.196).**
By computing the mean square difference between $X_n$ and $\sum_{j=0}^{k-1} a^j e_{n-j}$, we have

$$E\left[\left(X_n - \sum_{j=0}^{k-1} a^j e_{n-j}\right)^2\right] = E[(a^k X_{n-k})^2] = a^{2k} E[X_{n-k}^2].$$

Since the process is assumed as stationary, $E[X_{n-k}^2]$ is a constant, independent of $k$ and since $|a| < 1$, the RHS decreases to zero with geometric progression. Thus, we have (13.252).

**13.27 Variance and autocorrelation function of the AR(1) time series.**

(a)

$$E[X_n^2] = E\left[\sum_{j=0}^{\infty} a^j e_{n-j} \sum_{k=0}^{\infty} a^k e_{n-k}\right] = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} a^{j+k} E\left[e_{n-j}e_{n-k}\right]$$
$$= \sigma^2 \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} a^{j+k}\delta_{j,k} = \sigma^2 \sum_{j=0}^{\infty} = \frac{\sigma^2}{1-a^2}.$$

(b)

$$R_X[k] = E[X_n X_{n+k}] = E[X_n(aX_{n+k-1} + e_{n+k}] = aE[X_n X_{n+k-1}] = aR_X[k-1].$$

Hence by successive substitution of this recursion, we find

$$R_X[k] = a^2 R_X[k-2] = \cdots = a^k R_X[0] = \frac{\sigma^2 a^k}{1-a^2}.$$

# 14 Solutions for Chapter 14: Point Processes, Renewal Processes and Birth-Death Processes

## 14.1 Poisson Process

**14.1* Alternative derivation of the Poisson process.**

(a) Since the exponential distribution is memoryless, the interval $X_1$ till the first event point $t_1$ is exponentially distributed whether or not $t = 0$ is an event point, and

$$F_{t_1}(t) = F_X(t) = 1 - e^{-\lambda t}, \;\; t \geq 0. \tag{1}$$

(b) Since $t_{n+1} = t_n + X_{n+1}$ and $t_n$ and $X_{n+1}$ are independent, the PDF of $t_{n+1}$ is the convolution of the PDFs of $t_n$ and $X_{n+1}$.
(c) Thus, by setting $n = 1$ in the above, we have

$$f_{t_2}(t) = \int_0^t \lambda e^{-\lambda(t-u)} \lambda e^{-\lambda u}\, du = \lambda^2 t e^{-\lambda t}, \;\; t \geq 0. \tag{2}$$

By repeating the above step, we find

$$f_{t_n}(t) = \frac{(\lambda)^n t^{n-1}}{(n-1)!} e^{-\lambda t}, \;\; t \geq 0, \; n = 1, 2, \ldots. \tag{3}$$

Substitution of this result into (14.69) yields

$$\begin{aligned} P[N(t) = n] &= \int_0^t \left[ f_{t_n}(u) - f_{t_{n+1}}(u) \right] \, du = \frac{\lambda^n}{n!} \int_0^t e^{-\lambda u} (n u^{n-1} - \lambda u^n)\, du \\ &= \frac{(\lambda t)^n}{n!} e^{-\lambda t} \triangleq P(n; \lambda t), \end{aligned} \tag{4}$$

where the last expression was obtained by applying "integration by parts" to the first term in the integral, i.e.,

$$\begin{aligned} \int_0^t n u^{n-1} e^{-\lambda u}\, du &= \int_0^t \left( \frac{du^n}{du} \right) e^{-\lambda u}\, du = u^n e^{-\lambda u} \Big|_0^t + \lambda \int_0^t u^n e^{-\lambda u}\, du \\ &= t^n e^{-\lambda t} + \lambda \int_0^t u^n e^{-\lambda u}\, du. \end{aligned}$$

Thus, we have shown that this renewal process $N(t)$ has a Poisson distribution with mean $\lambda t$, if the lifetime distribution is the exponential distribution (14.99).

**14.2 Bernoulli Process to Poisson process.**

(a)

$$b(i; m, \lambda h + o(h)) = \binom{m}{i} [\lambda h + o(h)]^i [1 - \lambda h + o(h)]^{m-i}. \tag{5}$$

(b)

$$P[N(t) = i] = \frac{(\lambda t)^i}{i!} \lim_{m\to\infty} \frac{m!}{m^i (m-i)!} \lim_{m\to\infty} \left(1 - \frac{\lambda t}{m}\right)^{m-i} = \frac{(\lambda t)^i}{i!} e^{-\lambda t} = P(i; \lambda t),$$

where

$$P(i; a) = \frac{a^i}{i!} e^{-a}, \qquad i = 0, 1, 2, ... \tag{6}$$

**14.3 Superposition of Poisson processes.**

(a) Since $X_j$ is exponentially distributed with parameter $\lambda_j$ we have:

$$P[X_j \geq x] = e^{-\lambda_j}.$$

We have:

$$\begin{aligned} P[Y \geq y] &= P[\min\{X_1, \cdots, X_m\} \geq y] = P[X_1 \leq y, \cdots, X_m \geq y] \\ &= P[X_1 \geq y] \cdot P[X_2 \geq y] \cdots P[X_m \geq y] \\ &= e^{-\lambda_1} \cdots e^{-\lambda_j} = e^{-\lambda}, \end{aligned} \tag{7}$$

where $\lambda = \sum_{i=1}^{m} \lambda_i$. To obtain (7), we have used the fact that $\{X_j\}$ is a set of independent random variables. Therefore, we conclude the $Y$ is an exponentially distributed random variable with parameter $\lambda$.

(b) Given $m$ independent Poisson processes with rates $\lambda_j$, $j = 1, \cdots, m$, the corresponding inter-arrival times $\{X_j\}$ of the processes are independent random variables, exponentially distributed with parameters $\lambda_j$. At an arrival instant of the aggregate process, the time until the next arrival, $Y$, is given by

$$Y = \min\{X_1, \cdots, X_m\}.$$

From part (a), we know that $Y$ is exponentially distributed with parameter $\lambda$. Therefore, the aggregate process is a Poisson process with rate $\lambda$.

**14.4 Consistency check of the Poisson process.**

(a) Let $A(h)$ denote the number of arrivals in an interval $I_h$ of length $h$. Then $A(h)$ is a Poisson RV with parameter $\lambda h$. We have:

$$\begin{aligned}
P[\text{no arrival in } I_h] &= P[A(h) = 0] = e^{-\lambda h} \\
&= 1 - \lambda h + \frac{(\lambda h)^2}{2!} - \frac{(\lambda h)^3}{3!} + \frac{(\lambda h)^4}{4!} + \cdots \\
&= 1 - \lambda h + o(h), \\
P[\text{1 arrival in } I_h] &= P[A(h) = 1] = \lambda h e^{-\lambda h} \\
&= \lambda h(1 - \lambda h + o(h)) = \lambda h + o(h), \\
P[\text{2 or more arrivals in } I_h] &= \sum_{j=2}^{\infty} P[A(h) = j] = \sum_{j=2}^{\infty} \frac{(\lambda h)^j}{j!} \cdot e^{-\lambda h} = o(h).
\end{aligned}$$

(b) Let $X_1$ denote the time of the first arrival after the time origin (say $t = 0$) and $X_2$ denote the inter-arrival time between the first arrival and the second arrival. The RVs $X_1$ and $X_2$ are both exponentially distributed with parameter $\lambda$ and have a common cdf:

$$F_X(x) = 1 - e^{-\lambda x}, \;\; x \geq 0.$$

The event of no arrival in the interval $I_h$ is equivalent to the event $\{X_1 > h\}$. Therefore,

$$P[\text{no arrival in } I_h] = P[X_1 > h] = e^{-\lambda h} = 1 - \lambda h + o(h).$$

The event of two or more arrivals in the interval $I_h$ is equivalent to the event $\{X_1 + X_2 \leq h\}$. Let $Y = X_1 + X_2$. Then,

$$P[\text{2 or more arrivals in } I_h] = P[Y \leq h] = F_Y(h). \tag{8}$$

There are several ways of determining the cdf $F_Y(y)$. Since $X_1$ and $X_2$ are independent, the pdf of $Y$ is given by

$$f_Y(y) = f_{X_1}(y) \circledast f_{X_2}(y).$$

One can further show that in general, the cdf of $Y$ is given by

$$F_Y(y) = F_{X_1}(y) \circledast f_{X_2}(y) = f_{X_1}(y) \circledast F_{X_2}(y). \tag{9}$$

Therefore,

$$\begin{aligned}
F_Y(y) = F_X(y) \circledast f_X(y) &= \int_0^y (1 - e^{-\lambda x}) \cdot \lambda e^{-\lambda(y-x)} dx \\
&= \lambda e^{-\lambda y} \int_0^y (e^{\lambda x} - 1) dx = 1 - e^{-\lambda y} - \lambda y e^{-\lambda y}.
\end{aligned}$$

Returning to (8), we obtain

$$\begin{aligned}
P[\text{2 or more arrivals in } I_h] &= F_Y(h) = 1 - e^{-\lambda h} - \lambda h e^{-\lambda h} \\
&= 1 - (1 - \lambda h + o(h)) - (\lambda h + o(h)) = o(h).
\end{aligned}$$

Finally,

$$\begin{aligned}
P[\text{one arrival in } I_h] &= 1 - P[\text{no arrival in } I_h] - P[\text{two or more arrivals in } I_h] \\
&= 1 - (1 - \lambda h + o(h)) - o(h) = \lambda h + o(h).
\end{aligned}$$

To prove (9), note that

$$
\begin{aligned}
F_Y(y) &= \int_0^y f_{X_1}(x) \circledast f_{X_2}(x)\,dx = \int_0^y \int_0^x f_{X_1}(x-t) f_{X_2}(t)\,dt\,dx \\
&= \int_0^y \int_t^y f_{X_1}(x-t) f_{X_2}(t)\,dx\,dt = \int_0^y \int_0^{y-t} f_{X_1}(\alpha)\,d\alpha\, f_{X_2}(t)\,dt \\
&= \int_0^y F_{X_1}(y-t) f_{X_2}(t)\,dt = F_{X_1}(y) \circledast f_{X_2}(y).
\end{aligned}
$$

**14.5 Decomposition of a Poisson process.**

(a) We are given that $\{X_j\}$ is a sequence of i.i.d. random variables, exponentially distributed with parameter $\lambda$. Then for fixed $n$,

$$S_n = X_1 + \cdots + X_n$$

and has an Erlang-$n$ distribution. The cdf of $S_n$ is given by

$$
\begin{aligned}
F_{S_n}(x) &= P[S_n \le x] = 1 - P[\text{fewer than } n \text{ arrivals in an interval of length } x] \\
&= 1 - \sum_{j=0}^{n-1} P[A(x) = j] = 1 - e^{-\lambda x} \sum_{j=0}^{n-1} \frac{(\lambda x)^j}{j!}.
\end{aligned}
$$

Therefore,

$$P[S_n > x] = \sum_{j=0}^{n-1} P[A(x) = j] = 1 - e^{-\lambda x} \sum_{j=0}^{n-1} \frac{(\lambda x)^j}{j!}.$$

Hence,

$$
\begin{aligned}
P[S_N > x] &= \sum_{n=1}^{\infty} P[S_N > s|N = n]P[N = n] = \sum_{n=1}^{\infty} e^{-\lambda x} \sum_{j=0}^{n-1} \frac{(\lambda x)^j}{j!} \cdot (1-r)^{n-1} r \\
&= re^{-\lambda x} \sum_{n=0}^{\infty} \sum_{j=0}^{n} \frac{(\lambda x)^j}{j!} \cdot (1-r)^n = re^{-\lambda x} \sum_{j=0}^{\infty} \sum_{n=j}^{\infty} \frac{(\lambda x)^j}{j!} \cdot (1-r)^n \\
&= re^{-\lambda x} \sum_{j=0}^{\infty} \sum_{m=0}^{\infty} \frac{(\lambda x)^j}{j!} \cdot (1-r)^{m+j} = re^{-\lambda x} \sum_{j=0}^{\infty} \frac{[(\lambda(1-r)x]^j}{j!} \cdot \sum_{m=0}^{\infty} (1-r)^m \\
&= re^{-\lambda x} \cdot e^{\lambda(1-r)x} \left(\frac{1}{r}\right) = e^{-\lambda r x}
\end{aligned}
$$

which shows that $S_N$ has an exponential distribution with parameter $\lambda r$.

(b) In decomposing the Poisson stream into $m$ substreams, each arrival is assigned independently to the $k$th substream with probability $r_k$, where $\sum_{k=1}^{m} r_k = 1$. Consider an arrival that is assigned to the $k$th substream. The number of subsequent arrivals of the original Poisson stream until the next arrival that is assigned to the $k$th substream is a random variable $N$ with distribution

$$P[N = n] = (1 - r_k)^{n-1} r_k, \quad n = 0, 1, \cdots .$$

Therefore, the inter-arrival time between arrivals assigned to the $k$th substream is a random variable

$$S_N = X_1 + \cdots X_N,$$

where $X_i$ are inter-arrival times of the original Poisson process. Hence, the $X_i$ are i.i.d. and exponentially distributed with parameter $\lambda$. By the result from part (a), $S_N$ is exponentially distributed with parameter $r_k\lambda$.

**14.6 Alternate decomposition of a Poisson stream.** Let $X_i$ represent the interarrival time between the $i$th and the $(i+1)$-st arrival. For substream 1, the time between the first and the second arrival is given by

$$Y = X_1 + X_2 + \cdots X_m.$$

The event $\{Y \le y\}$ is equivalent to the event that there are fewer than $m$ arrivals of the original Poisson stream in an interval of length $y$, i.e.,

$$\begin{aligned} F_Y(y) \triangleq P[Y \le y] &= 1 - P[\text{fewer than } m \text{ arrivals in an interval of length } y] \\ &= 1 - \sum_{j=0}^{m-1} P[A(y) = j] = 1 - \sum_{j=0}^{m-1} \frac{(\lambda y)^j}{j!} e^{-\lambda y}, \end{aligned}$$

which is an Erlang-$m$ distribution with mean $m/\lambda$.

**14.7 Derivation of the Poisson distribution.**

(a) Equation (4.62) can be written as:

$$\frac{d}{dt} \ln(P_0(t)) = -\lambda,$$

which is a simple, separable first-order differential equation. Integrating both sides and solving for $P_0(t)$ yields $P_0(t) = Ke^{-\lambda t}$, where the constant $K$ is determined by the initial condition $P_0(0) = 1$. Hence, $K = 1$ and

$$P_0(t) = e^{-\lambda t}. \tag{10}$$

Substituting (10) into (4.61) with $n = 1$, we obtain

$$P_1'(t) + \lambda P_1(t) = \lambda e^{-\lambda t}. \tag{11}$$

Equation (11) is a first-order differential equation that can be reduced to a separable form by multiplying both sides by an integrating factor. More generally, let us re-write (11) as follows:

$$P_1'(t) + R(t)P_1(t) = Q(t), \tag{12}$$

where in this case, $R(t) = \lambda$ and $Q(t) = \lambda e^{-\lambda t}$. The integrating factor can be obtained by supposing that the left-hand side of (12) to be the derivative of a product $\phi(t)P_1(t)$, given by

$$\phi(t)P_1'(t) + \phi'(t)P_1(t). \tag{13}$$

Multiplying the left-hand side of (12) by $\phi(t)$, we have

$$\phi(t)P_1'(t) + \phi(t)R(t)P_1(t) = \phi(t)Q(t). \tag{14}$$

Equating the left-hand side of (14) with (13), we see that they can be made equal by choosing $\phi(t)$ such that

$$\phi'(t) = \phi(t)R(t). \tag{15}$$

This is a simple separable equation that has the solution

$$\phi(t) = e^{\int R(t)dt}, \tag{16}$$

which is the integrating factor we seek.
After multiplying (11) by the integrating factor $\phi(t)$, we obtain:

$$\frac{d}{dt}\left[e^{\int R(t)dt}P_1(t)\right] = Q(t)e^{\int R(t)dt}. \tag{17}$$

The left-hand side is an exact derivative that can be integrated directly. In particular, we have

$$\frac{d}{dt}[e^{\lambda t}P_1(t)] = \lambda. \tag{18}$$

Hence, we obtain (using the fact that $P_1(0) = 1$):

$$P_1(t) = \lambda t e^{-\lambda t}. \tag{19}$$

To proceed by induction, we postulate the result

$$P_n(t) = \frac{(\lambda t)^n}{n!}e^{-\lambda t} \tag{20}$$

and show that the result holds for $P_{n+1}(t)$. From (4.61), we have:

$$P'_{n+1}(t) + \lambda P_{n+1}(t) = \lambda P_n(t) = \lambda\frac{(\lambda t)^n}{n!}e^{-\lambda t}.$$

Multiplying both sides by the integrating factor $\phi(t) = e^{\lambda t}$, we have:

$$\frac{d}{dt}[e^{\lambda t}P_{n+1}(t)] = \lambda\frac{(\lambda t)^n}{n!}.$$

Integrating both sides and using the fact that $P_{n+1}(0) = 0$, we obtain the required result:

$$P_{n+1}(t) = \frac{(\lambda t)^{(n+1)}}{(n+1)!}e^{-\lambda t}.$$

Thus, by induction, we have shown the validity of (20) for all $n \geq 0$.
(b) Taking the Laplace transform of the system of differential equations (4.61) and (4.62), we have:

$$sP_n^*(s) - P_n(0) = -\lambda P_n^*(s) + \lambda P_{n-1}^*(s), \ \ n \geq 1 \tag{21}$$

$$sP_0^*(s) - P_0(0) = -\lambda P_0^*(s). \tag{22}$$

From (22) and the fact that $P_0(0) = 1$, we obtain $P_0^*(s) = \frac{1}{s+\lambda}$. Noting that $P_n(0) = 0$ in (21) we have, in particular for $n = 1$,

$$(s+\lambda)P_1^*(s) = \lambda P_0^*(s).$$

Therefore, $P_1^*(s) = \frac{\lambda}{(s+\lambda)^2}$ . Using induction, it is straightforward to show that

$$P_n^*(s) = \frac{\lambda^n}{(s+\lambda)^{n+1}} \tag{23}$$

Inverting (23), we obtain the desired result:

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t},$$

where we have used the following Laplace transform properties:

$$\mathcal{L}^{-1}\{f^*(s+a)\} = f(t)e^{-at} \tag{24}$$

$$\mathcal{L}^{-1}\left\{\frac{1}{s^{n+1}}\right\} = \frac{t^n}{n!}. \tag{25}$$

**14.8* Uniformity and statistical independence of Poisson arrivals.** TBD

(a) We wish to prove that the joint PDF of $U_1, \ldots, U_n$ conditioned on $\{N(T) = n\}$ is given by

$$f_{U_1 \cdots U_n}(u_1, \ldots, u_n | N(T) = n) = \frac{1}{T^n}, \tag{14.103}$$

where $U_1, \ldots, U_n$ are the unordered arrival times of a Poisson process in the interval $(0, T]$. Let $u_1, \ldots, u_n$ be distinct values in the interval $(0, T)$. Without loss of generality, assume that $0 < u_1 < u_2 < \cdots < u_n < T$. Define intervals $I_j \in (u_j, u_j + h_j]$, where $h_j > 0$, $j = 1, \ldots, n$, such that the intervals are disjoint and each $I_j$ is contained in $(0, T]$. Let $\mathcal{E}_n$ denote the event that $n$ arrivals fall in the intervals $I_j$, $j = 1, \ldots, n$, with exactly one arrival in each interval. The interval $(0, T]$ can be partitioned into $2n + 1$ intervals as follows:

$$(0, T] = (0, u_1] \cup I_1 \cup (u_1 + h_1, u_2] \cup \cdots \cup I_n \cup (u_n + h_n, T]. \tag{26}$$

When the event $\mathcal{E}_n$ occurs, exactly one arrival occurs in each interval $I_j$, $j = 1, \ldots, n$, with probability $\lambda h e^{-\lambda h}$, and no arrival occurs in each of the other subintervals in the partition (26). Therefore,

$$\begin{aligned} P[\mathcal{E}_n] &= (e^{-\lambda u_1})(\lambda h_1 e^{-\lambda h_1})(e^{-\lambda(u_2 - u_1 - h_1)}) \cdots (\lambda h_n e^{-\lambda h_n})(e^{-\lambda(T - u_n - h_n)}) \\ &= \lambda^n e^{-\lambda T} \prod_{j=1}^{n} h_j. \end{aligned} \tag{27}$$

We can also write

$$\begin{aligned} P[\mathcal{E}_n] &\overset{(a)}{=} \sum_{\sigma} P[U_1 \in I_{\sigma(1)}, U_2 \in I_{\sigma(2)}, \ldots, U_n \in I_{\sigma(n)}, N(T) = n] \\ &\overset{(b)}{=} n!\, P[U_1 \in I_1, U_2 \in I_2, \ldots, U_n \in I_n], \end{aligned} \tag{28}$$

where the summation on the right-hand side of (a) is over all permutations $\sigma$ on the set $\{1, \ldots, n\}$. Step (b) follows because the $n$ arrivals are unordered. From (27) and (28), we obtain

$$P[U_1 \in I_1, \ldots, U_n \in I_n] = \frac{\lambda^n}{n!} e^{-\lambda T} \prod_{j=1}^{n} h_j. \tag{29}$$

Hence,

$$\begin{aligned} &P[U_1 \in I_1, \ldots, U_n \in I_n | N(T) = n] \\ &\quad = \frac{P[U_1 \in I_1, \ldots, U_n \in I_n]}{P[N(T) = n]} = \frac{\frac{\lambda^n}{n!} e^{-\lambda T} \prod_{j=1}^{n} h_j}{\frac{(\lambda T)^n}{n!} e^{-\lambda T}} = \frac{\prod_{j=1}^{n} h_j}{T^n}. \end{aligned} \tag{30}$$

We remark that (30) holds for any ordering of the $u_i$'s, although we assumed $u_1 < \cdots < u_n$ for convenience in obtaining (27). The joint density of the unordered arrivals, $U_1, \ldots, U_n$, conditioned on the event $\{N(T) = n\}$, then follows from (cf. (4.92))

$$f_{U_1 \cdots U_n}(u_1, \ldots, u_n | N(t) = n) = \lim_{\substack{h_j \to 0 \\ j=1,\ldots,n}} \frac{P[U_1 \in I_1, \ldots, U_n \in I_n | N(T) = n]}{\prod_{j=1}^{n} h_j}, \tag{31}$$

which results in (14.103).
(b) The result (14.103) suggests the following procedure to generate Poisson arrivals in an interval of $(0, T]$:

a. Draw the number of arrivals $n$ from a Poisson distribution with parameter $\lambda T$.
b. For $i = 1, \ldots, n$, draw the value of the unordered arrival time $U_i$ from a uniform distribution on $(0, T]$, independently of the others.

## 14.2 Birth-Death (BD) Process

**14.9** **Pure birth process.** When $\lambda(n) = \lambda$ and $\mu(n) = 0$ for all $n \geq 0$, the differential-difference equations of the BD process become:

$$P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t), \;\; n = 1, 2, \cdots, \tag{32}$$

$$P_0'(t) = -\lambda P_0(t). \tag{33}$$

Using the same procedure as in Problem **??**, these equations can be solved to obtain the Poisson distribution:

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \;\; n = 0, 1, \cdots.$$

**14.10** **Time-dependent solution.**
When $\mu(n) = 0$ for all $n \geq 0$, but state-dependent birth rates $\lambda(n)$ are permitted, the differential-difference equations of the BD process become:

$$P_n'(t) = -\lambda(n) P_n(t) + \lambda(n-1) P_{n-1}(t), \;\; n = 1, 2, \cdots, \tag{34}$$

$$P_0'(t) = -\lambda P_0(t). \tag{35}$$

If we multiply both sides of (34) by the integrating factor $e^{\lambda(n)t}$, we obtain:

$$\frac{d}{dt}[e^{\lambda(n)t}P_n(t)] = \lambda(n-1)P_{n-1}(t)e^{\lambda(n)t}. \tag{36}$$

After integrating both sides from 0 to $t$ and re-arranging, we obtain:

$$P_n(t) = e^{-\lambda(n)t}\left[\lambda(n-1)\int_0^t P_{n-1}(x)e^{\lambda(n)x}dx + K\right], \tag{37}$$

where $K$ is a constant determined by the initial condition $P_n(0) = K$.

**14.11 Pure death process.** When $\lambda(n) = 0$ and $\mu(n) = \mu$ for all $n \geq 0$, the differential-difference equations of the BD process become:

$$P'_{N_0}(t) = -\mu P_{N_0}(t), \tag{38}$$
$$P'_n(t) = -\mu P_n(t) + \mu P_{n+1}(t), \;\; n = 1, \cdots, N_0 - 1 \tag{39}$$
$$P'_0(t) = \mu P_1(t). \tag{40}$$

Solving (38), we obtain

$$P_{N_0}(t) = e^{-\mu t}. \tag{41}$$

Similar to the approach in Problem 4.7, we can obtain from (39) and (40), the following result:

$$P_n(t) = \mu e^{-\mu t}\int_0^t e^{\mu x}P_{n+1}(x)dx, \;\; n = 1, \cdots, N_0 - 1. \tag{42}$$

Applying (42) successively for $n = 1, 2, \cdots, N_0 - 1$, we obtain:

$$P_n(t) = \frac{(\mu t)^{N_0-n}}{(N_0-n)!}e^{-\mu t}, \;\; n = 1, \cdots N_0.$$

For each $t$ we have:

$$P_0(t) + \sum_{n=1}^{N_0} P_n(t) = 1.$$

Thus, we find that

$$P_0(t) = 1 - \sum_{n=1}^{N_0}\frac{(\mu t)^{N_0-n}}{(N_0-n)!}e^{-\mu t} = 1 - \sum_{n=0}^{N_0-1}\frac{(\mu t)^n}{n!}e^{-\mu t}.$$

**14.12 The time-dependent PGF.** Multiply both sides of (32) and (33) by $z^n$ and sum from $n = 0$ to $\infty$ to obtain:

$$\sum_{n=0}^{\infty} P'_n(t)z^n = -\lambda\sum_{n=0}^{\infty} P_n(t)z^n + \lambda\sum_{n=1}^{\infty} P_{n-1}(t)z^n, \tag{43}$$

which is equivalent to

$$\frac{\partial}{\partial t}G(z,t) = -\lambda G(z,t) + \lambda z G(z,t) \tag{44}$$
$$= -\lambda(1-z)G(z,t). \tag{45}$$

Equation (45) is a simple, separable first order differential equation whose solution is:

$$G(z,t) = e^{-\lambda(1-z)t} = \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}z^n. \tag{46}$$

Hence,

$$P_n(t) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}.$$

**14.13* Time-dependent solution for a certain BD process.** When $\lambda(n) = \lambda$ and $\mu(n) = n\mu$ for $n \geq 0$, the differential-difference equations for the BD process become:

$$P_n'(t) = -(\lambda + n\mu)P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t), \quad n = 1, 2, \cdots, \tag{47}$$

$$P_0'(t) = \lambda P_0(t) + \mu P_1(t). \tag{48}$$

Multiply both sides of (47) and (48) by $z^n$ and sum from $n = 0$ to $\infty$ to obtain:

$$\begin{aligned}\sum_{n=0}^{\infty} P_n'(t)z^n = {}& -\lambda\sum_{n=0}^{\infty} P_n(t)z^n - \mu\sum_{n=1}^{\infty} nP_n(t)z^n \\ & + \lambda\sum_{n=1}^{\infty} P_{n-1}(t)z^n + \mu\sum_{n=0}^{\infty}(n+1)P_{n+1}(t)z^n,\end{aligned} \tag{49}$$

which can be written as:

$$\frac{\partial}{\partial t}G(z,t) = -\lambda G(z,t) - \mu z\frac{\partial}{\partial z}G(z,t) + \lambda z G(z,t) + \mu\frac{\partial}{\partial z}G(z,t). \tag{50}$$

Re-arranging terms we have the following partial differential equation in $G(z,t)$:

$$\left[\frac{\partial}{\partial t} + \mu(z-1)\frac{\partial}{\partial z}\right]G(z,t) = \lambda(z-1)G(z,t). \tag{51}$$

It remains to verify that the solution

$$G(z,t) = \exp\left\{\frac{\lambda}{\mu}(1 - e^{-\mu t})(z-1)\right\} \tag{52}$$

satisfies (51). Alternatively, we may obtain the form of the solution (52) as follows. Based on (51), we suppose that $G(z,t)$ has the form: $G(z,t) = \exp(f(z,t))$. In this case, (52) reduces to the following partial differential equation:

$$\left[\frac{\partial}{\partial t} + \mu(z-1)\frac{\partial}{\partial z}\right]f(z,t) = \lambda(z-1). \tag{53}$$

Based on (53), we suppose that $f(z,t)$ is separable as follows: $f(z,t) = \lambda(z-1)f(t)$. Then (53) simplifies to:

$$f'(t) + \mu f(t) = 1. \tag{54}$$

This is a first-order differential equation that can be solved by multiplying both sides by the integrating factor $e^{\mu t}$, resulting in:

$$\frac{d}{dt}[e^{\mu t}f(t)] = e^{\mu t}. \tag{55}$$

The solution of the above equation is:

$$f(t) = \frac{1}{\mu} + Ke^{-\mu t},$$

where the constant $K$ is determined from:

$$G(0,0) = P_0(0) = 1.$$

But $G(0,0) = 1$ implies that $f(0) = 0$, which determines $K$ as $-1/\mu$. Thus,

$$f(t) = \frac{1}{\mu}\left(1 - e^{-\mu t}\right),$$

and

$$G(z,t) = \exp[\lambda(z-1)f(t)].$$

## 14.3 Renewal Process

**14.14* Derivation of (14.72).**

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} m_X, \tag{56}$$

as $n \to \infty$. Noting that

$$\frac{1}{N(t)}\sum_{i=1}^{N(t)} X_i = \frac{t_{N(t)}}{N(t)}, \tag{57}$$

we deduce from (56) that

$$\frac{N(t)}{t_{N(t)}} \xrightarrow{\text{a.s.}} \frac{1}{m_X}, \tag{58}$$

as $t \to \infty$. The left-hand side of (58) can be written as

$$\frac{N(t)}{t}\frac{t}{t_{N(t)}}. \tag{59}$$

Since $\frac{t}{t_{N(t)}} \xrightarrow{\text{a.s.}} 1$ as $t \to \infty$, we can establish from (58) and (59) that (14.72) holds.

**14.15 The Poisson process as a renewal process.**

(a) Since

$$P[N(t) = n] = \frac{(\lambda t)^n}{n!}e^{-\lambda t},$$

we have

$$M(t) = E[N(t)] = \sum_{n=1}^{\infty} \frac{n(\lambda t)^n}{n!}e^{-\lambda t} = e^{-\lambda t}\lambda t \sum_{n=1}^{\infty} \frac{(\lambda t)^{n-1}}{(n-1)!} = e^{-\lambda t}\lambda t e^{\lambda t} = \lambda t, \text{ for } t \geq 0.$$

Hence $m(t) = M'(t) = \lambda$ for $t \geq 0$.

(b) One approach is to use the formula (14.66) and find $f_{t_2}(x), f_{t_3}(x), \ldots$, but a simpler way is to use the relation (14.68), i.e.,

$$F_{t_k}(t) = P[N(t) \geq k] = \sum_{n=k}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t},$$

which, by differentiation, leads to

$$f_{t_k}(t) = \frac{\lambda(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t},$$

which is the $k$-stage Erlang distribution with mean $k/\lambda$.

(c)

$$\begin{aligned}
\text{LHS} &= M(t) = \lambda t \\
\text{RHS} &= 1 - e^{-\lambda t} + \lambda^2 \int_0^t (t-x)e^{-\lambda x}\, dx \\
&= 1 - e^{-\lambda t} + \lambda^2 t \int_0^t e^{-\lambda x}\, dx - \lambda^2 \int_0^t x e^{-\lambda x}\, dx.
\end{aligned}$$

Substituting the following results into the above

$$\begin{aligned}
\int_0^t e^{-\lambda x}\, dx &= \frac{1 - e^{-\lambda t}}{\lambda} \\
\int_0^t x e^{-\lambda x}\, dx &= -\frac{t e^{-\lambda t}}{\lambda} + \frac{1 - e^{-\lambda t}}{\lambda^2},
\end{aligned}$$

we find

$$\text{RHS} = 1 - e^{-\lambda t} + \lambda t(1 - e^{-\lambda t}) + \lambda t e^{-\lambda t} - (1 - e^{-\lambda t}) = \lambda t.$$

Hence the renewal equation holds, as expected.

(d) By substituting $f_X^*(s) = \frac{\lambda}{s+\lambda}$ into (14.79), we have

$$m^*(s) = \frac{\frac{\lambda}{s+\lambda}}{1 - \frac{\lambda}{s+\lambda}} = \frac{\lambda}{s}.$$

By applying the inverse Laplace transform, we find

$$m(t) = \mathcal{L}\left\{\frac{\lambda}{s}\right\} = \lambda u(t),$$

where $u(t)$ is the unit step function.

(e) Substituting $F_X(t) = 1 - e^{-\lambda t}$ into (14.85), we find $f_R(r) = \lambda e^{\lambda r} = f_X(r)$, as expected for the exponential distribution of $X$.

(f) Hence, its Laplace transform is the same as $f_X^*(s)$, i.e., $f_R^*(s) = \frac{\lambda}{s+\lambda}$. Then $f_R^{*(1)} = -\lambda(s+\lambda)^{-2}$ Then, we find

$$m_R = -f_R^*(0) = \lambda^{-1}.$$

Alternatively, we can use the formula (14.89):

$$m_r = \frac{E[X^2]}{2m_X} = \frac{\sigma_X^2 + m_X^2}{2m_X} = \frac{2m_X^2}{2m_X} = m_X = \lambda^{-1}.$$

**14.16 Residual lifetime of an Erlang distributed lifetime.**

(a) The mean of this $k$-stage Erlang distribution is $m_X = k/\lambda$. The PDF of the residual lifetime $R$ is given, from (14.85), as

$$f_R(r) = \frac{1}{m_X}\left(1 - F_X(r)\right) = \frac{1}{m_X} e^{-\lambda r} \sum_{j=0}^{k-1} \frac{(\lambda r)^j}{j!}.$$

(b) The Laplace transform of the above PDF is

$$f_R^*(s) = \frac{1}{m_X}\left(\frac{1}{s} - \frac{f_X^*(s)}{s}\right).$$

Since $f_X(t)$ is the $k$-fold convolution of the exponential PDF of the mean $1/\lambda$, its Laplace transform is

$$f_X^*(s) = \left(\frac{\lambda}{s+\lambda}\right)^k.$$

Thus,

$$f_R^*(s) = \frac{1}{m_X} \frac{1 - \left(\frac{\lambda}{s+\lambda}\right)^k}{s}.$$

(c) We have

$$\begin{aligned} f_X^*(s) &= \lambda^k (s+\lambda)^{-k} \\ f_X^{*'}(s) &= -k\lambda^k (s+\lambda)^{-k-1} \\ f_X^{*(2)}(s) &= k(k+1)\lambda^k (s+\lambda)^{-k-2} \end{aligned}$$

Hence

$$f_X^{*(2)}(0) = \frac{k(k+1)}{\lambda^2}.$$

Thus, from (14.88), we find

$$m_R = \frac{f_X^{*(2)}(0)}{2m_X} = \frac{k+1}{2\lambda} = \frac{k+1}{2k} m_X \leq m_X,$$

where the equality holds when $k = 1$, for which the Erlang distribution reduces to the exponential distribution, and $m_R = m_X$, as expected.

**14.17 Residual lifetime of a uniformly distributed lifetime.**

(a) Let $X^*$ be the lifetime of the interval that the random observer hits. Then its PDF is given by

$$f_{X^*}(x) = \frac{x f_X(x)}{m_X},$$

as argued in the text. For the uniform RV $X$ over $[0, L]$, we have $m_X = L/2$ and $f_X(x) = 1/L$, for $0 \le x \le L$. Given $X^* = x$, the age $Y$ is uniformly distributed in $[0, x]$, i.e.,

$$f_Y(y|X^* = x) = \begin{cases} \frac{1}{x}, & 0 \le y \le x \\ 0, & y > x. \end{cases}$$

Thus the unconditional PDF of $Y$ is

$$f_Y(y) = \int_0^\infty f_Y(y|X^* = x) f_{X^*}(x)\, dx = \frac{\int_y^\infty f_X(x)\, dx}{m_X} = \frac{1 - F_X(y)}{m_X}.$$

Substituting $F_X(y) = y/L$, we have

$$f_Y(y) = \frac{2}{L}\left(1 - \frac{y}{L}\right), \;\; 0 \le y \le L.$$

Hence,

$$F_Y(y) = \frac{y}{L}\left(2 - \frac{y}{L}\right), \;\; 0 \le y \le L.$$

(b) The age and the residual life are symmetric as discussed in the text. So the distribution of $R$ is the same as that of $Y$, i.e.,

$$F_R(r) = \frac{r}{L}\left(2 - \frac{r}{L}\right), \;\; 0 \le r \le L.$$

**14.18 Moments of residual lifetime and age.** The PDF of $R$ is given by (14.85):

$$f_R(r) = \frac{1 - F_X(r)}{m_X}.$$

Thus,

$$\begin{aligned} E[R^n] &= \frac{1}{m_X}\int_0^\infty r^n (1 - F_X(r))\, dr = \frac{1}{m_X}\int_0^\infty \left(\frac{r^{n+1}}{n+1}\right)' (1 - F_X(r))\, dr \\ &= \frac{1}{m_X}\left[\frac{r^{n+1}(1 - F_X(r))}{n+1}\right]_0^\infty + \frac{1}{m_X}\int_0^\infty \frac{r^{n+1}}{n+1} f_X(r)\, dr = \frac{E\left[X^{n+1}\right]}{(n+1)E[X]}. \end{aligned}$$

# 15 Solutions for Chapter 15: Discrete-Time Markov Chains

## 15.1 Markov Processes and Markov Chains

**15.1* Homogeneous Markov chain.**

(a) Straightforward.

(b)

$$\boldsymbol{p}^\top(1) = \boldsymbol{p}^\top(0)\boldsymbol{P} = (1\,0\,0)\begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 0 & 2/3 \\ 0 & 1/5 & 4/5 \end{bmatrix} = (1/2\;1/2\;0)$$

$$\boldsymbol{p}^\top(2) = \boldsymbol{p}^\top(1)\boldsymbol{P} = (1/2\;1/2\;0)\begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 0 & 2/3 \\ 0 & 1/5 & 4/5 \end{bmatrix} = (5/12\;1/4\;1/3)$$

$$\boldsymbol{p}^\top(3) = \boldsymbol{p}^\top(2)\boldsymbol{P} = (5/12\;1/4\;1/3)\begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 0 & 2/3 \\ 0 & 1/5 & 4/5 \end{bmatrix} = (7/24\;11/40\;13/30)$$

(c)

$$\boldsymbol{g}\top(z) = \boldsymbol{p}^\top(0)[\boldsymbol{I} - \boldsymbol{P}z]^{-1}.$$

$$\begin{aligned}\det|\boldsymbol{I} - \boldsymbol{P}z| &= \det\begin{bmatrix} 1-\frac{z}{2} & -\frac{z}{2} & 0 \\ -\frac{z}{3} & 1 & -\frac{2}{3}z \\ 0 & -\frac{z}{5} & 1-\frac{4}{5}z \end{bmatrix} \\ &= 1 - \frac{13z}{10} + \frac{z^2}{10} + \frac{z^3}{5} = (1-z)\left(1 - \frac{3z}{10} - \frac{z^2}{5}\right) \triangleq \Delta(z).\end{aligned}$$

Hence,

$$[\boldsymbol{I} - \boldsymbol{P}z]^{-1} = \frac{1}{\Delta(z)}\begin{bmatrix} 1-\frac{4z}{5}-\frac{2z^2}{15} & \frac{z}{2}\left(1-\frac{4z}{5}\right) & \frac{z^2}{3} \\ \frac{z}{3}\left(1-\frac{4z}{5}\right) & \left(1-\frac{z}{2}\right)\left(1-\frac{4z}{5}\right) & \frac{z^2}{3}\left(1-\frac{z}{2}\right) \\ \frac{z^2}{15} & \frac{z}{5}\left(1-\frac{z}{2}\right) & 1-\frac{z}{2}-\frac{z^2}{6} \end{bmatrix}$$

Then by substituting $\boldsymbol{p}^\top(0) = (1\;0\;)$ and the last expression into (15.25), we have

$$\boldsymbol{g}^\top(z) = (g_1(z), g_2(z), g_3(z)),$$

where

$$g_1(z) = \frac{1}{\Delta(z)}\left(1 - \frac{4z}{5} - \frac{2z^2}{15}\right),$$
$$g_2(z) = \frac{z}{2\Delta(z)}\left(1 - \frac{4z}{5}\right),$$
$$g_3(z) = \frac{z^2}{3\Delta(z)}.$$

Hence

$$\lim_{t\to\infty} g_1(z) = \lim_{z\to 1}(1-z)g_1(z) = \frac{2}{15},$$
$$\lim_{t\to\infty} g_2(z) = \frac{1}{5},$$
$$\lim_{t\to\infty} g_1(z) = \frac{2}{3}$$

**15.2 State probabilities.**

(a) States 1 and 2 are transient states, and state 3 is an absorbing state.

(b)

$$\det[\boldsymbol{I} - z\boldsymbol{P}] = \det\begin{bmatrix} 1-\frac{2z}{3} & -\frac{z}{3} & 0 \\ -\frac{z}{2} & 1 & -\frac{z}{2} \\ 0 & 0 & 1-z \end{bmatrix} = \left(1 - \frac{z^2}{3}\right)(1-z) - \frac{z^2 91 - z)}{6}$$
$$= (1-z)\left(1 - \frac{2z}{3} - \frac{z^2}{6}\right) \triangleq \Delta(z).$$

Hence,

$$z_0 = 1,\ z_1 = -4 + 2\sqrt{10},\ z_2 = -4 - 2\sqrt{10}.$$

(c)

$$[\boldsymbol{I} - z\boldsymbol{Q}]^{-1} = \begin{bmatrix} 1-z & \frac{z(1-z)}{3} & \frac{z^2}{6} \\ \frac{z(1-z)}{2} & \left(1-\frac{2z}{3}\right)(1-z) & \frac{z}{2}\left(1-\frac{2z}{3}\right) \\ 0 & 0 & 1-\frac{2z}{3}-\frac{z^2}{6} \end{bmatrix}.$$

From (15.25), we have

$$g_1(z) = \frac{1-z}{\Delta(z)} = \frac{1}{1-\frac{2z}{3}-\frac{z^2}{6}},$$
$$g_2(z) = \frac{z(1-z)}{3\Delta(z)} = \frac{\frac{z}{3}}{1-\frac{2z}{3}-\frac{z^2}{6}},$$
$$g_3(z) = \frac{z^2}{6\Delta(z)} = \frac{\frac{z^2}{6}}{(1-z)\left(1-\frac{2z}{3}-\frac{z^2}{6}\right)}.$$

By writing

$$\left(1-\frac{2z}{3}-\frac{z^2}{6}\right)=(1-\alpha z)(1-\beta z),$$

where

$$\alpha=\frac{1+\sqrt{5/2}}{3},\ \ \beta=\frac{1-\sqrt{5/2}}{3},$$

we obtain the partial fraction expansion

$$g_1(z)=\frac{1}{\left(1-\frac{\beta}{\alpha}\right)(1-\alpha z)}+\frac{1}{\left(1-\frac{\alpha}{\beta}\right)(1-betaz)}.$$

Hence,

$$p_1(n)=\frac{\alpha^n}{1-\frac{\beta}{\alpha}}+\frac{\beta^n}{1-\frac{\alpha}{\beta}}=\frac{3\left(\alpha^{n+1}-\beta^{n+1}\right)}{\sqrt{10}},\ \ n=0,1,2,\ldots.$$

Similarly,

$$g_2(z)=\frac{1}{3(\alpha-\beta)(1-\alpha z)}+\frac{1}{3(\beta-\alpha)(1-\beta z)}.$$

Hence,

$$p_2(n)=\frac{\alpha^n-\beta^n}{\sqrt{10}},\ \ n=0,1,2,\ldots.$$

Similarly,

$$g_3(z)=\frac{1}{1-z}+\frac{1}{6\alpha^2\left(1-\frac{1}{\alpha}\right)\left(1-\frac{\beta}{\alpha}\right)(1-\alpha z)}+\frac{1}{6\beta^2\left(1-\frac{1}{\beta}\right)\left(1-\frac{\alpha}{\beta}\right)(1-\beta z)}.$$

Hence,

$$p_3(n)=1+\frac{1}{2\sqrt{10}}\left(\frac{\alpha^n}{\alpha-1}-\frac{\beta^n}{\beta-1}\right),\ \ n=0,1,2,\ldots.$$

**15.3 Simple queueing problem.** Let

$$P_{ij}=P[X_{n+1}=j|X_n=i],\ \ i,j\in\mathcal{S}=\{0,1,2,\ldots\}.$$

Then

$$P_{0j}=\begin{cases}1-\alpha, & j=0,\\ \alpha, & j=1\\ 0 & j\geq 2.\end{cases}$$

$$P_{ij}=\begin{cases}\beta(1-\alpha), & i\geq 1,\ j=i-1,\\ 1-\alpha(1-\beta)-\beta(1-\alpha), & i\geq 1,\ j=i,\\ \alpha(1-\beta), & 1\geq 1,\ \ j=i+1,\\ 0, & i\geq 1,\ |j-i|\geq 2.\end{cases}$$

## 15.2 Computation of State Probabilities

**15.4*** **Transitive property.** If $i \leftrightarrow j$, then there exists at least one $m$ such that $P_{ij}^{(m)} > 0$, which allows $i$ to reach $j$. Similarly, $j \leftrightarrow k$ means there exists $n$ such that $P_{jk}^{(n)} > 0$. Then

$$P_{ik}^{(m+n)} \geq P_{ij}^{(m)} P_{jk}^{(n)} > 0,$$

hence, $i \to k$. A symmetrical argument shows that $k \to i$. Thus, $i \leftrightarrow k$. Thus, we have proven the transitive property of the communication property $\leftrightarrow$.

**15.5*** **Stationary distribution.**

(a)

$$\det[\boldsymbol{P} + \boldsymbol{E} - \boldsymbol{I}] = \det \begin{bmatrix} 0 & 2 & 1 \\ \frac{5}{4} & \frac{1}{4} & \frac{3}{2} \\ 1 & \frac{3}{2} & \frac{1}{2} \end{bmatrix} = \frac{27}{8}.$$

and

$$[\boldsymbol{P} + \boldsymbol{E} - \boldsymbol{I}]^{-1} = \frac{8}{27} \begin{bmatrix} -\frac{17}{8} & \frac{1}{2} & \frac{11}{4} \\ \frac{7}{8} & -1 & \frac{5}{4} \\ \frac{13}{8} & 2 & -\frac{5}{2} \end{bmatrix}.$$

Therefore,

$$\boldsymbol{\pi}^\top = \left( \frac{1}{9}, \frac{4}{9}, \frac{4}{9} \right).$$

(b)

$$\det[\boldsymbol{P} + \boldsymbol{E} - \boldsymbol{I}] = \det \begin{bmatrix} \frac{1}{2} & \frac{3}{2} & 1 \\ \frac{4}{3} & 0 & \frac{5}{3} \\ 1 & \frac{6}{5} & \frac{4}{5} \end{bmatrix} = \frac{3}{2}.$$

and

$$[\boldsymbol{P} + \boldsymbol{E} - \boldsymbol{I}]^{-1} = \frac{2}{3} \begin{bmatrix} -2 & 0 & \frac{5}{2} \\ \frac{3}{5} & -\frac{3}{5} & \frac{1}{2} \\ \frac{8}{5} & \frac{9}{10} & -2 \end{bmatrix}.$$

Therefore,

$$\boldsymbol{\pi}^\top = \left( \frac{2}{15}, \frac{1}{5}, \frac{2}{3} \right).$$

**15.6** **Characteristic equation.**

(a) Straightforward

(b) – A periodic chain with the period$= 2$.

– Symmetric in the sense states 1 and 4 and states 2 and 3 are interchangeable.

(c)

$$\det[\boldsymbol{I} - z\boldsymbol{P}] = \det \begin{bmatrix} 1 & -z & 0 & 0 \\ -\frac{z}{2} & 1 & -\frac{z}{2} & 0 \\ 0 & -\frac{z}{2} & 1 & -\frac{z}{2} \\ 0 & 0 & -z & 1 \end{bmatrix}$$

$$= \det \begin{bmatrix} 1 & -\frac{z}{2} & 0 \\ -\frac{z}{2} & 1 & -\frac{z}{2} \\ 0 & -z & 1 \end{bmatrix} + \frac{z}{2}\det \begin{bmatrix} -z & 0 & 0 \\ -\frac{z}{2} & 1 & -\frac{z}{2} \\ 0 & -z & 1 \end{bmatrix}$$

$$= 1 - \frac{z^2}{2} - \frac{Z^4}{4} + \frac{z}{2}\left(-z + \frac{z^3}{2}\right) = (1 - z^2)\left(1 - \frac{z^2}{4}\right)$$

$$= (1 - z)(1 + z)\left(1 - \frac{z}{2}\right)\left(1 + \frac{z}{2}\right).$$

Hence

$$z = \pm 1, \pm 2$$

.

**15.7 Roots of a characteristic equation.**

(a) If $z_k$'s are the roots, $p_i(n),\ i \in \mathcal{S}$ has the form

$$p_i(n) = \sum_k a_k z_k^{-n},\ i \in \mathcal{S},\ n = 0, 1, 2, \ldots.$$

Since $p_i(n) \leq 1$ for all $n$, it must follow that $|z_k^{-1}| \leq 1$, or $|z_k| \geq 1$ for all $k$.

(b) All roots such that $|z_k^{-1}| < 1$ decay to zero as $n \to \infty$. So, there must be one root whose magnitude is one, i.e., $|z_0^{-1}| = 1$. It is clear that $z_0$ must be unity, i.e., $z_0 = 1$, otherwise some probability $p_i(n)$ would become negative, or a complex number.

If there are multiple roots equal to unity, the chain can be separated into independent ergodic subchains.

(c) The chain contains a subchain that is periodic with period $k$. If the other factors of the characteristic equation are function of $z^k$, then the entire chain is periodic.

**15.8 Transition probability matrix of a certain structure.**

(a) The diagram is straightforward. States 1 and 2 are absorbing states. States 3 and 4 are transient states and are periodic with period 2.

(b) The submatrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is obvious, since states 1 and 2 are absorbing states. The submatrix $\begin{bmatrix} 0 & \frac{1}{2^n} \\ \frac{1}{2^n} & 0 \end{bmatrix}$ for states 3 and 4 is true for $n$ is odd. For $n =$ even, it should be instead $\begin{bmatrix} \frac{1}{2^n} & 0 \\ 0 & \frac{1}{2^n} \end{bmatrix}$. Another way of proving $\boldsymbol{P}^n$ is by mathematical induction.

(c)

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{Q}_1^n & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_2^n & \boldsymbol{0} \\ * & * & \boldsymbol{C}^n \end{bmatrix}.$$

The result of (a) correspond to a special case in which $\boldsymbol{C} = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{bmatrix}$.

**15.9 First-passage time matrix.**

(a)

$$\begin{aligned} \mu_{ij} &= P_{ij} + \sum_k P_{ik}(\mu_{kj} + 1) - P_{ij}(\mu_{jj} + 1) \\ &= \sum_k P_{ik} + \sum_k P_{ik} \left( \mu_{kj} - P_{ik} \delta_{kj} \mu_{jj} \right). \end{aligned}$$

Hence,

$$\boldsymbol{M} = \boldsymbol{E} + \boldsymbol{P} \left( \boldsymbol{M} - \boldsymbol{M}_{\mathrm{dg}} \right).$$

(b)

$$\boldsymbol{\pi}^\top \boldsymbol{M} = \boldsymbol{\pi}^\top \boldsymbol{E} + \boldsymbol{\pi}^\top \boldsymbol{P} \left( \boldsymbol{M} - \boldsymbol{M}_{\mathrm{dg}} \right) = \mathbf{1}^\top + \boldsymbol{\pi}^\top \left( \boldsymbol{M} - \boldsymbol{M}_{\mathrm{dg}} \right).$$

Hence,

$$\mathbf{1}^\top = (\pi_1 \mu_{11}, \pi_2 \mu_{22}, \ldots, \pi_N \mu_{NN}).$$

Therefore,

$$\pi_i \mu_i = 1 \ \text{ for } \ 1 \le i \le N.$$

(c) For the Markov chain with

$$\boldsymbol{P} = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

we obtained

$$\boldsymbol{\pi}^\top = \left( \frac{1}{9}, \frac{4}{9}, \frac{4}{9} \right).$$

Hence, $\boldsymbol{M}_{rmdg} = \begin{bmatrix} 9 & 0 & 0 \\ 0 & \frac{9}{4} & 0 \\ 0 & 0 & \frac{9}{4} \end{bmatrix}$. The equation of (a) becomes

$$\begin{bmatrix} 0 & \mu_{12} & \mu_{13} \\ \mu_{21} & \frac{9}{4} & \mu_{23} \\ \mu_{31} & \mu_{32} & \frac{9}{4} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & \mu_{12} & \mu_{13} \\ \mu_{21} & 0 & \mu_{23} \\ \mu_{31} & \mu_{32} & 0 \end{bmatrix}$$

From this we can solve $\mu_{ij}$'s to find

$$\boldsymbol{M} = [\mu_{ij}] = \begin{bmatrix} 9 & 1 & \frac{7}{2} \\ 8 & \frac{9}{4} & \frac{5}{2} \\ 10 & 2 & \frac{9}{4} \end{bmatrix}.$$

For the Markov chain with

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{1}{5} & \frac{4}{5} \end{bmatrix}.$$

we obtained $\boldsymbol{\pi}^\top = \left(\frac{2}{15}, \frac{1}{5}, \frac{2}{3}\right)$. Hence

$$\boldsymbol{M}_{\mathrm{dg}} = \begin{bmatrix} \frac{15}{2} & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & \frac{3}{2} \end{bmatrix}.$$

The equation we must solve is

$$\begin{bmatrix} \frac{15}{2} & \mu_{12} & \mu_{13} \\ \mu_{21} & 5 & \mu_{23} \\ \mu_{31} & \mu_{32} & \frac{3}{2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{1}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} 0 & \mu_{12} & \mu_{13} \\ \mu_{21} & 0 & \mu_{23} \\ \mu_{31} & \mu_{32} & 0 \end{bmatrix}.$$

Hence

$$\boldsymbol{M} = [\mu_{ij}] = \begin{bmatrix} \frac{15}{2} & 2 & \frac{9}{2} \\ 13 & 5 & \frac{5}{2} \\ 18 & 5 & \frac{3}{2} \end{bmatrix}.$$

**15.10 Computation formula for the stationary distribution.**

By definition an ergodic Markov chain is aperiodic and irreducible, and every state is positive-recurrent. Then there exists a **unique** stationary distribution $\boldsymbol{\pi}$ satisfying

$$\boldsymbol{\pi}^\top \boldsymbol{P} = \boldsymbol{\pi}^\top, \tag{1}$$

and

$$\boldsymbol{\pi}^\top \mathbf{1} = 1. \tag{2}$$

Let $\boldsymbol{x}_0$ be a solution to the following **homogeneous equation**

$$\boldsymbol{x}^\top (\boldsymbol{P} + k\boldsymbol{E} - \boldsymbol{I}) = \mathbf{0}. \tag{3}$$

Then

$$\boldsymbol{x}_0^\top (\boldsymbol{P} - \boldsymbol{I}) = -k\boldsymbol{x}_0^\top \boldsymbol{E} = -ka\mathbf{1}^\top, \tag{4}$$

where

$$a = \boldsymbol{x}_0^\top \mathbf{1}, \tag{5}$$

Multiplying the above by $\mathbf{1}$, and using $\boldsymbol{P}\mathbf{1} = \mathbf{1}$, we have

$$0 = \boldsymbol{x}_0^\top (\boldsymbol{P} - \boldsymbol{I})\mathbf{1} = -kaM.$$

Therefore $a = 0$, and $\boldsymbol{x}_0$ is a solution to (1).

Then from the uniqueness of the stationary distribution of a finite ergodic Markov chain (see (15.102), it follows that

$$\boldsymbol{x}_0 = c\boldsymbol{\pi},$$

for some constant $c$. But if we multiply the transpose of the above by $\mathbf{1}$ from the right, we find

$$\boldsymbol{x}_0^\top \mathbf{1} = c\boldsymbol{\pi}^\top \mathbf{1}.$$

From (5), the LHS is $a = 0$. The RHS is $c$, since $\boldsymbol{\pi}^\top \mathbf{1} = 1$ from the definition of $\boldsymbol{\pi}^\top \mathbf{1} = 1$. Hence,

$$\boldsymbol{x}_0 = \mathbf{0}.$$

It then follows that (3) admits only trivial solutions, and hence $\boldsymbol{P} + k\boldsymbol{E} - \boldsymbol{I}$ is invertible. It is straightforward to show that $\boldsymbol{\pi}$ satisfies the following **non-homogeneous equation**:

$$\boldsymbol{x}^\top (\boldsymbol{P} + k\boldsymbol{E} - \boldsymbol{I}) = k\mathbf{1}^\top, \tag{6}$$

because

$$\boldsymbol{\pi}^\top (\boldsymbol{P} + k\boldsymbol{E} - \boldsymbol{I}) = \boldsymbol{\pi}^\top + k(\boldsymbol{\pi}^\top \mathbf{1})\mathbf{1}^\top - \boldsymbol{\pi}^\top = k\mathbf{1}^\top.$$

Therefore, $\boldsymbol{\pi}$ is given by (15.109).

**15.11 Number of returns.** Assume $n_0 = 0$ without loss of generality, i.e, the system (the Markov chain) is in state $i$ at time $n = 0$. Suppose that the system returns to state $i$ for the first time at $n_1$ and for a second time at $n_2$. Let event $E_k$ be defined as

$$E_k \triangleq \{n_k < \infty\}.$$

Then $\{E_k\}$ is a decreasing event and $E_k$ represents that there are at least $k$ returns to state $i$. Note that

$$P[E_1] = f_{ii}.$$

where $f_{ii}$ is the probability that the chain ever returns to state $i$, as defined in (15.60). After each return to state $i$, the system starts afresh all over again. Thus,

$$P[E_2] = P[E_2 \cap E_1] = P[E_2|E_1]P[E_1] = f_{ii}^2.$$

In general the sequence

$$P[E_k] = f_{ii}^k.$$

If state $i$ is a transient state, then $f_{ii} < 1$. Thus

$$\sum_{k=1}^{\infty} P[E_k] = \sum_{k=1}^{\infty} f_{ii}^k = \frac{f_{ii}}{1 - f_{ii}} < \infty.$$

Let us define

$$A = \bigcap_{n=1}^{\infty} \left( \bigcup_{k=n}^{i} nftyE_k \right),$$

which represents the event that "infinitely many of events $E_1, E - 2, \ldots$ occur. Then from the first Borel-Cantelli Lemma,

$$P[A] = 0.$$

Thus with probability 1 only finitely many of the events $E_k$ do occur. In other words, the chain returns to state $i$ only finitely often.
If the state $i$ is recurrent, then $f_{ii} = 1$, hence

$$P[E_k] = 1 \text{ for every } k.$$

Let $N$ be the number of times the chain $X_n$ returns to state $i$ as $n \to \infty$. Since the event $E_k$ and the event $\{N \geq k\}$ are equivalent, it follows that

$$P[N \geq k] = 1, \text{ for every } k,$$

or

$$P[N = \infty] = 1.$$

Thus, the chain returns to the returns to a recurrent state infinitely often.

# 16 Solutions for Chapter 16: Semi-Markov Processes and Continuous-Time Markov Chains

## 16.1 Semi-Markov Process

**16.1 Alternating renewal process.**

(a) Let $\mathcal{S} = \{0, 1\}$, where state 1 represents the "up" state, and state 0 is the "down" sate. Then the machine alternates between the two states with TPM

$$\boldsymbol{P} = \tilde{\boldsymbol{P}} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

and the sojourn time distributions are $F_{10}(s) = P[S_{10} \le s] = F_U(s)$, and $F_{01}(s) = P[S_{01} \le s] = F_D(s)$. Thus, it defines a renewal process,

(b) The counting process $N(t)$ increments by one each time either a state transition $1 \to 0$ or $0 \to 1$ occurs, whereas $I(t)$ flip-flops between states 1 and 0. So it is clear that

$$I(t) = N(t) \ \ (mod 2) = \begin{cases} 1, & \text{if } N(t) \text{ is odd} \\ 0, & \text{if } N(t) \text{ is even.} \end{cases}$$

(c) Since the stationary distribution of the EMC with $\tilde{\boldsymbol{P}}$ given above is $\tilde{\pi}_0 = \tilde{\pi}_1 = 0.5$. Thus, from (16.10, we find

$$\pi_1 = \frac{\tilde{\pi}_1 E[S_1]}{\tilde{\pi}_0 E[S_0] + \tilde{\pi}_1 E[S_1]} = \frac{E[U]}{E[D] + E[U]}.$$

**16.2* Conditional independence of sojourn times.**

Using a basic property of conditional probability (see (2.59) in Section 2.4.1), we have

$$\begin{aligned} &P[\tau_1 \le u_1, \tau_2 \le u_2, \dots, \tau_n \le u_n | X_0, X_1, \dots] \\ &\quad = P[\tau_1 \le u_1 | \tau_2 \le u_2, \dots, \tau_n \le u_n, X_0, X_1, \dots] \\ &\quad\quad \cdot P[\tau_2 \le u_2 | \tau_3 \le u_2, \dots, \tau_n \le u_n, X_0, X_1, \dots] \\ &\quad\quad \cdot P[\tau_3 \le u_3 | \tau_4 \le u_4, \dots, \tau_n \le u_n, X_0, X_1, \dots] \\ &\quad\quad \cdots \\ &\quad\quad \cdot P[\tau_n \le u_n | X_0, X_1, \dots]. \end{aligned} \tag{1}$$

Since $\tau_j$ depends only on $X_{j-1}$ and $X_j$, we have for $1 \le j \le n-1$:

$$\begin{aligned} &P[\tau_j \le u_j | \tau_{j+1} \le u_{j+1}, \dots, \tau_n \le u_n, X_0, X_1, \dots] \\ &\quad = P[\tau_j \le u_j | X_{j-1}, X_j] = F_{X_{j-1} X_j}(u_j). \end{aligned} \tag{2}$$

Applying (2) in (1), we obtain the desired result

$$
\begin{aligned}
P[\tau_1 \le u_1, &\tau_2 \le u_2, \ldots, \tau_n \le u_n | X_0, X_1, \ldots] \\
&= F_{X_0 X_1}(u_1) F_{X_1 X_2}(u_2) \cdots F_{X_{n-1} X_n}(u_n).
\end{aligned}
$$

**16.3* Semi-Markovian kernel.**
Suppose we are given the semi-Markovian kernel $\boldsymbol{Q}(t) = [Q_{ij}(t)]$, $i, j \in \mathcal{S}$. We can obtain $\boldsymbol{P} = [P_{ij}]$ as follows:

$$
\begin{aligned}
P_{ij} &= P[X_{n+1} = j | X_n = i] \\
&= \lim_{t \to \infty} P[X_{n+1} = j, t_{n+1} - t_n \le t | X_n = i] = \lim_{t \to \infty} Q_{ij}(t).
\end{aligned}
$$

Then we can obtain $\boldsymbol{F}(t) = F_{ij}(t)$ as follows:

$$
\begin{aligned}
F_{ij}(t) &= P[t_{n+1} - t \le t | X_n = i, X_{n+1} = j] \\
&= \frac{P[X_{n+1}, t_{n+1} - t_n \le t | X_n = i]}{P[X_{n+1} = j | X_n = i]} \\
&= \frac{Q_{ij}(t)}{P_{ij}} \qquad (3) \\
&= \frac{Q_{ij}(t)}{\lim_{t \to \infty} Q_{ij}(t)}.
\end{aligned}
$$

Conversely, given $\boldsymbol{P}$ and $\boldsymbol{F}(t)$, the semi-Markovian kernel can be obtained from (3) as follows:

$$
Q_{ij}(t) = F_{ij}(t) P_{ij}. \qquad (4)
$$

**16.4 Markov renewal process and renewal process (cf. [57]).**

(a) TBD

(b) **Note:** Equation (16.81) is incorrect in the textbook, and should instead be written as follows:

$$
Q_{ik}^{n+1}(t) = \sum_{j \in \mathcal{S}} \int_0^t Q_{jk}^n(t-u) dQ_{ij}(u), \quad n = 0, 1, 2, \ldots. \qquad (16.81)
$$

Since $t_0 = 0$,

$$
Q_{ij}^0(t) = P[X_0 = j, t_0 \le t | X = i] = P[X_0 = j | X_0 = i] = \delta_{ij}.
$$

Given that $X_0 = i$, the event $\{X_{n+1} = k, t_{n+1} \le t\}$ occurs when the to $X_1$ happens at some time $u \le t$ into some state $j \in \mathcal{S}$, followed by $n$ further transitions ending up with $X_{n+1} = k$

in time $\leq t - u$. Thus, we have

$$\begin{aligned} Q_{ik}^{n+1}(t) &= P[X_{n+1} = k, t_{n+1} \leq t | X_n = i] \\ &= \sum_{j \in \mathcal{S}} \int_0^t P[X_{n+1} = k, t_{n+1} - t_1 \leq t - u | X_1 = j] \, dQ_{ij}(u) \\ &= \sum_{j \in \mathcal{S}} \int_0^t P[X_n = k, t_n \leq t - u | X_0 = j] \, dQ_{ij}(u) \\ &= \sum_{j \in \mathcal{S}} \int_0^t Q_{jk}^n(t-u) dQ_{ij}(u). \end{aligned}$$

(c) Consider the counting process $\{N_j(t)\}$ defined by

$$N_j(t) = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = j, t_n \leq t\}}.$$

This process counts the number of times that $X_n = j$ with $t_n \leq t$. Then

$$\begin{aligned} M_{ij}(t) &= E[N_j(t) | X_0 = i] \\ &= \sum_{n=0}^{\infty} E[\mathbf{1}_{\{X_n = j, t_n \leq t\}} | X_0 = i] \\ &= \sum_{n=0}^{\infty} P[X_n = j, t_n \leq t | X_0 = i] = \sum_{n=0}^{\infty} Q_{ij}^n(t). \end{aligned}$$

## 16.2 Continuous-time Markov Chain (CTMC)

**16.5* Markovian property of an SMP.**

We will show that a CTMC $X(t)$ is equivalent to an SMP with sojourn time distributions $F_{ij}(t)$ given by

$$F_{ij}(t) = 1 - e^{-\nu_i t}, \quad t \geq 0, \quad i, j \in \mathcal{S}. \tag{16.17}$$

For simplicity, assume that none of the states of $X(t)$ is an absorbing state. Suppose that the CTMC $X(t)$ enters state $i$ at time 0. Let $S_i$ denote the sojourn time of $X(t)$ in state $i$ starting at time 0 before it makes a jump to another state $j \neq i$. For $s, t \geq 0$, we have

$$\begin{aligned} P[S_i > s + t \mid S_i > s] &= P[X(\tau) = i;\ 0 \leq \tau \leq s + t \mid X(\tau) = i;\ 0 \leq \tau \leq s] \\ &= P[X(\tau) = i;\ s \leq \tau \leq s + t \mid X(\tau) = i;\ 0 \leq \tau \leq s] \\ &= P[X(\tau) = i;\ s \leq \tau \leq s + t \mid X(s) = i] \qquad (5) \\ &= P[X(\tau) = i,\ 0 \leq \tau \leq t \mid X(0) = i] \qquad (6) \\ &= P[S_i > t], \qquad (7) \end{aligned}$$

where (5) is due to the Markov property (see Definition 15.1) and (6) is due to the assumed stationarity (or time-homogeneity) of the process $X(t)$. This implies that the random variable

$S_i$ is memoryless and must then be an exponential random variable[1], say with rate $\nu_i$. In other words, the sojourn time distributions of $X(t)$ are given by (16.17).

Let $t_0 = 0$ and let $t_1, t_2, \ldots$ denote the jump times of $X(t)$. It remains to show that the process $\{X_n\}$ defined by $X(t_n)$, $n = 0, 1, 2, \ldots$ is a DTMC. For $n \geq 1$, we have

$$\begin{aligned} P[X_n = j \mid X_0, X_1, \ldots, X_{n-1}] & \\ = P[X(t_n) = j \mid X(t_0), X(t_n), \ldots, X(t_{n-1})] & \quad (8) \\ = P[X(t_n) = j \mid X(t_{n-1})] & \\ = P[X_n = j \mid X_{n-1}] = P_{X_{n-1},i}. & \quad (9) \end{aligned}$$

In the above derivation, (8) formally resembles the Markov property given in Definition 15.1. A key difference, however, is that the times $t_1, t_2, \ldots$ are random variables, not constants. Nevertheless, (8) does in fact hold in the case of a CTMC and is called the *strong Markov property*. The strong Markov property holds when $t_1, t_2, \ldots$ are *stopping times* for $X(t)$. [2]

**16.6* CTMC as an SMP.**

Let $X(t)$ be a CTMC characterized by an infinitesimal generator matrix $\boldsymbol{Q} = [Q_{ij}]$. As shown in Problem 16.5, $X(t)$ is equivalent to an SMP with sojourn time distributions given by (16.17). Let $\{X_n\}$ denote the embedded Markov chain (EMC) of $X(t)$ (see (16.2)) and let $\boldsymbol{P} = [P_{ij}]$ denote its transition probability matrix (TPM).

Suppose that the CTMC $X(t)$ enters state $i$ at time 0. We shall first assume that state $i$ is not an absorbing state. In this case, $P_{ii} = 0$. The CTMC $X(t)$ remains in state $i$ for a sojourn time $S_i$ and then transitions to another state $j \neq i$. As shown in Problem 16.5, $S_i$ is exponentially distributed with parameter $\nu_i > 0$. Therefore, $P[S_i < h] = 1 - e^{-\nu_i h}$, $h \geq 0$. For sufficiently small $h$,

$$P[S_i < h] = 1 - P_{ii}(h).$$

Hence,

$$\lim_{h\to 0} \frac{P[S_i < h]}{h} = \lim_{h\to} \frac{1 - P_{ii}(h)}{h} = -Q_{ii}, \tag{10}$$

[1] Let $g(t) = P[S_i > t]$. Then (7) implies that $g(t+s) = g(t)g(s)$ for all $s, t \geq 0$. It is well-known that the unique solution to this functional equation has the form $g(t) = e^{\alpha t}$ for some constant $\alpha$.

[2] A random variable $T$ taking values in $[0, +\infty]$ is called a stopping time for a process $X(t)$ if for every $t$, $0 \leq t < \infty$, the occurrence or non-occurrence of the event $\{T \leq t\}$ is completely determined from $\{X(u),\ u \leq t\}$. For a stopping time $T$ and a CTMC $X(t)$, the following strong Markov property holds:

$$P[X(T+s) = j \mid X(i),\ u \leq T] = P[X(s) = j \mid X(0)] = P_{X(0),j}(s).$$

For further details, the reader is referred to, e.g., Cinlar [57], Section 8.1.

Since the left-hand side of (10) is given by $\nu_i$, we have $\nu_i = -Q_{ii}$. The transition probability $P_{ij} = P[X(S_i) = j \mid X(0) = i]$ can be expressed as

$$\begin{aligned}
P_{ij} &= \lim_{h\to 0} P[X(S_i+h) = j \mid X(t) = i,\ 0 \le t < S_i; X(S_i+h) \ne i] \\
&= \lim_{h\to 0} \frac{P[X(S_i+h) = j \mid X(S_i-) = i]}{P[X(S_i+h) \ne i \mid X(S_i-) = i]} \qquad (11)\\
&= \lim_{h\to 0} \frac{P_{ij}(h)}{1-P_{ii}(h)} = \lim_{h\to 0} \frac{\frac{P_{ij}(h)}{h}}{\frac{1-P_{ii}(h)}{h}} = \frac{Q_{ij}}{-Q_{ii}}, \qquad (12)
\end{aligned}$$

where we have applied the strong Markov property to obtain (11) and (16.23) and (16.24) to obtain the last equality in (12). If state $i$ is absorbing, $\nu_i = Q_{ii} = 0$ and $P_{ii} = 1$.
In summary, the CTMC $X(t)$ with generator $\boldsymbol{Q}$ can be characterized as an SMP with sojourn time distributions

$$F_{ij} = 1 - e^{Q_{ii}t}, \quad t \ge 0,\ i, j \in \mathcal{S}, \tag{13}$$

and transition probabilities given by

$$P_{ij} = \begin{cases} \frac{Q_{ij}}{-Q_{ii}}, & \text{if } i \text{ is not absorbing} \\ \delta_{ij}, & \text{if } i \text{ is absorbing.} \end{cases} \tag{14}$$

The SMP representation of a CTMC provides a convenient approach to simulate a sample path of a CTMC given an initial state $X(0) = x_0$. If state $x_0$ is not an absorbing state ($Q_{x_0x_0} \ne 0$), the dwell time $\tau_1$ in state $i$ as an exponentially distributed random variable with parameter $\nu_{x_0} = -Q_{x_0x_0}$. The next state $x_1$ is then determined according to the the probability distribution $\{P_{x_0j}\},\ j \in \mathcal{S}$, given by (23). In case $x_0$ is an absorbing state, the CTMC remains forever in this state, so dwell time $\tau_1 = +\infty$ and the procedure terminates. The procedure is repeated, if necessary, from state $x_1$ to produce a dwell time $\tau_2$, etc. The resulting sequence $\{(x_0, \tau_1), (x_1, \tau_2), \ldots\}$ specifies the sample path of the CTMC.

**Alternative solution:**

From Exercise 16.3, the semi-Markovian kernel of an SMP can be written as

$$Q_{ij}(t) = P[X_1 = j, \tau_1 \le t | X_0 = i] = F_{ij}(t)P_{ij} = (1 - e^{-\nu_i t})P_{ij}, \tag{15}$$

where we applied (16.17) to obtain the last equality.
The transition probability matrix function (TPMF) for a CTMC $X(t)$ is given by $\boldsymbol{P}(t) = [P_{ij}(t)]$ where (cf. (16.18))

$$P_{ij}(t) = P[X(t) = j | X(0) = i] = P[X(t) = j | X_0 = i], \quad i, j \in \mathcal{S},\ \ 0 \le t < \infty. \tag{16}$$

We shall show that the transition probability function $P_{ij}(t)$ can be related to the semi-Markovian kernel $Q_{ij}(t)$ as follows:

$$P_{ij}(t) = \delta_{ij}\left[1 - \sum_{k\in\mathcal{S}} Q_{ik}(t)\right] + \sum_{k\in\mathcal{S}} \int_0^t P_{kj}(t-s)\, dQ_{ik}(s), \tag{17}$$

where $\delta_{ij} = 0$ is the Kronecker delta. This equation can be interpreted as follows: First suppose that $i \neq j$. Given that $X(0) = X_0 = i$ at time $t_0 = 0$, in order for the event $\{X(t) = j\}$ to happen, $X(t)$ takes its first jump from state $i$ to some state $k$ at a time $s$, $0 < s \leq t$ and then given that $X(s) = k$, $X(t)$ ends up in state $j$ at time $t$. Now if $i = j$, there is an additional possibility that $X(t)$ does not take its first jump until after time $t$. Equation (17) can be derived more formally as follows:

$$P_{ij}(t) = P[X(t) = j, T_1 > t \mid X_0 = i] + P[X(t) = j, T_1 \leq t \mid X_0 = i]. \tag{18}$$

For the first term on the right, we have

$$\begin{aligned} &P[X(t) = j, T_1 > t \mid X_0 = i] \\ &\quad = P[T_1 > t \mid X_0 = i] \cdot P[X(t) = j \mid T_1 > t, X_0 = i] \\ &\quad = \left[1 - \sum_{k \in \mathcal{S}} Q_{ik}(t)\right] \cdot \delta_{ij}. \end{aligned} \tag{19}$$

For the second term, we have

$$\begin{aligned} &P[X(t) = j, T_1 \leq t \mid X_0 = i] \\ &\quad = E[P[X(t) = j, T_1 \leq t \mid X_0 = i, X_1, T_1] \mid X_0 = i] \\ &\quad = E[\mathbf{1}_{\{T_1 \leq t\}} \cdot P[X(t) = j \mid X_1, T_1, X_0 = i] \mid X_0 = i] \\ &\quad = E[\mathbf{1}_{\{T_1 \leq t\}} \cdot P[X(t - T_1) = j \mid X_1, X_0 = i] \mid X_0 = i] \\ &\quad = E[\mathbf{1}_{\{T_1 \leq t\}} P_{X_1, j}(t - T_1) \mid X_0 = i] \\ &\quad = \sum_{k \in \mathcal{S}} \int_0^t P_{kj}(t - s) \, dQ_{ik}(s). \end{aligned} \tag{20}$$

Substituting (19) and (20) into (18), we obtain (17).
Applying (16.23),

$$\begin{aligned} Q_{ij} &= \frac{dP_{ij}}{dt}\Big|_{t=0} \\ &= -\delta_{ij} \sum_{k \in \mathcal{S}} Q'_{ik}(0) + \sum_{k \in \mathcal{S}} P_{kj}(0) Q'_{ik}(0) \\ &= -\delta_{ij} \sum_{k \in \mathcal{S}} \nu_i P_{ik} + \sum_{k \in \mathcal{S}} \delta_{kj} \nu_i P_{ik}. \end{aligned}$$

For $i \neq j$, we have

$$Q_{ij} = \nu_i P_{ij}, \tag{21}$$

whereas

$$Q_{ii} = -\nu_i \sum_{k \neq i} P_{ik}. \tag{22}$$

If $i$ is an absorbing state, then $P_{ii} = 1$ and from (21) and (22) we have $Q_{ij} = 0$ for all $j \in \mathcal{S}$. In this case, $\nu_i = 0$. If $i$ is not an absorbing state, then $P_{ii} = 0$ and from (22) we have $Q_{ii} = -\nu_i$.

In this case, $\nu_i > 0$ and in particular, we have

$$\nu_i = -Q_{ii}, \quad P_{ij} = \frac{Q_{ij}}{\nu_i} = \frac{Q_{ij}}{-Q_{ii}}. \tag{23}$$

**16.7 Semi-Markovian kernel of a CTMC.**
A CTMC $X(t)$ can be viewed as an SMP for which the sojourn time distribution $F_{ij}(t)$ is exponentially distributed with some parameter $\nu_i > 0$ (see (16.17)):

$$F_{ij}(t) = 1 - e^{-\nu_i t}, \quad t \geq 0, \quad i, j \in \mathcal{S}. \tag{16.17}$$

Regarding $X(t)$ as an SMP (see Definition 16.1), let $\{X_n\}$ denote the embedded Markov chain (EMC) of $X(t)$ (see (16.2)) and let $\boldsymbol{P} = [P_{ij}]$ denote its transition probability matrix (TPM). The semi-Markovian kernel of an SMP can be written as

$$\begin{aligned} Q_{ij}(t) &= P[X_{n+1} = j, \tau_{n+1} \leq t \mid X_n = i] \\ &= P[\tau_{n+1} \leq t \mid X_n = i, X_{n+1} = j] \cdot P[X_{n+1} = j \mid X_n = i] \\ &= F_{ij}(t) P_{ij}. \end{aligned}$$

From the solution to Problem 16.6, $F_{ij}(t)$ is given by (13) and $P_{ij}$ is given by (14). Hence, we have

$$Q_{ij}(t) = \begin{cases} (1 - e^{Q_{ii} t}) \frac{Q_{ij}}{-Q_{ii}}, & \text{if } i \text{ is not absorbing,} \\ 0, & \text{if } i \text{ is absorbing .} \end{cases} \tag{24}$$

**16.8 Nonuniform Markov chain.**
For the special case of $\boldsymbol{Q}$, it is not uniform, since $Q_{i,0} \to \infty$ as $i \to \infty$. Then

$$Q_{0,0}^{(2)} = \left[\boldsymbol{Q}^2\right]_{0,0} = 0 + 1 \cdot 1^{-2} + 2 \cdot 2^{-2} + \ldots + k \cdot k^{-2} + \cdots = \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

Thus, it is apparent the series (16.39) does not converge.

**16.9 Invariant (or stationary) distribution of a CTMC.** The probability distribution at time $t$ $\boldsymbol{\pi}(t) = (p_i(t),\ i \in \mathcal{S})^\top$ is given by (16.40). Differentiating both sides and setting it to zero, we find

$$\frac{d\boldsymbol{\pi}^\top(t)}{dt} = \boldsymbol{\pi}^\top(0) \frac{d\boldsymbol{P}(t)}{dt} = \boldsymbol{\pi}^\top(0) \boldsymbol{Q} \boldsymbol{P}(t) = \mathbf{0}^\top.$$

Thus,

$$\boldsymbol{\pi}^\top(0) \boldsymbol{Q} = \mathbf{0}^\top,$$

which is (16.41).

**16.10* Balance equations.**
From (16.42), we have

$$\sum_{j \neq i} \pi_j Q_{ji} + \pi_i Q_{ii} = 0, \quad \text{for all } i \in \mathcal{S}.$$

From (16.24)

$$Q_{ii} = -\sum_{j \neq i} Q_{ij}.$$

By substituting this into the above equation, we arrive at (16.43).

**16.11 Markov Modulated Poisson Process.**

(a) Although the process $N(t)$ is not by itself a Markov process, the bivariate process $Z(t) = (N(t), X(t))$ is a Markov process with state space $\{0, 1, 2, \ldots\} \times$. To see this, note that for any $j \in \mathcal{S}$ and $n \in \{0, 1, 2, \ldots\}$,

$$\begin{aligned}
&P[Z(t) = (n, j) \mid Z(\tau), 0 \leq \tau \leq s] \\
&\quad = P[N(t) = n, X(t) = j \mid N(\tau), X(\tau),\ 0 \leq \tau \leq s] \\
&\quad = P[X(t) = j, N(t) = n \mid N(s), X(s)] \\
&\quad = P[Z(t) = (j, n)) \mid Z(s)].
\end{aligned}$$

Let $Z_n = Z(t_n) = (X(t_n), N_(t_n))$. Since $X_n = X(t_n)$ and $N_n = N(t_n) = n$, we have $Z_n = (X_n, n)$. Using the strong Markov property of $Z(t)$ (see the solution to Problem 16.5) we have

$$\begin{aligned}
&P[X_{n+1} = j, t_{n+1} - t_n \leq t \mid X_0, \ldots X_n; t_0, \ldots, t_n] \\
&\quad = P[X_{n+1} = j, t_{n+1} - t_n \leq t \mid Z_0, \ldots, Z_n] \\
&\quad = P[X_{n+1} = j, t_{n+1} - t_n \leq t \mid Z_n] \\
&\quad = P[X_{n+1} = j, t_{n+1} - t_n \leq t \mid X_n].
\end{aligned}$$

Hence, $(X_n, t_n)$ is a Markov renewal process.

(b) Equation (16.83) in the text should be corrected as follows:

$$D_{ij}(t) = \lambda_i e^{-\lambda_i t} e^{Q_{ii} t} \delta_{ij} + \int_0^t e^{-\lambda_i s} \sum_{k \neq i} Q_{ik} D_{kj}(t - s)\, ds, \quad i, j \in \mathcal{S}. \tag{16.83}$$

Equation (16.83) can be derived in a manner similar to the derivation of (17) in the alternative solution to Problem 16.6. Let $J_1$ denote the time of the first jump of the process $X(t)$. We have

$$\begin{aligned}
D_{ij}(t) &= P[X_1 = j, t_1 \in [t, t + dt) \mid X_0 = i] \\
&= P[X_1 = j, t_1 \in [t, t + dt), J_1 > t \mid X_0 = i] \\
&\quad + P[X_1 = j, t_1 \in [t, t + dt), J_1 \leq t \mid X_0 = i].
\end{aligned} \tag{25}$$

The first term can be evaluated as follows:

$$\begin{aligned}
&P[X_1 = j, t_1 \in [t, t + dt), J_1 > t \mid X_0 = i] \\
&\quad = P[X_1 = j, t_1 \in [t, t + dt) \mid J_1 > t, X_0 = i] \cdot P[J_1 > t \mid X_0 = i] \\
&\quad = \delta_{ij} \lambda_i e^{-\lambda_i t} \cdot e^{Q_{ii} t}.
\end{aligned} \tag{26}$$

The second term can be evaluated as follows:

$$\begin{aligned}
&P[X_1 = j, t_1 \in [t, t+dt), J_1 \le t \mid X_0 = i] \\
&\quad = E[P[X_1 = j, t_1 \in [t, t+dt), J_1 > t \mid X_0 = i, J_1, X(J_1)] \mid X_0 = i] \\
&\quad = E[\mathbf{1}_{\{J_1 \le t\}} \cdot P[X_1 = j, t_1 \in [t - J_1, t - J_1 + dt) \mid X_0 = X(J_1)] \mid X_0 = i] \\
&\quad = E[\mathbf{1}_{\{J_1 \le t\}} \cdot D_{X(J_1)j}(t - J_1) \mid X_0 = i] \\
&\quad = \int_0^t \sum_{k \in \mathcal{S}} D_{kj}(t-s) e^{-\lambda_i s} Q_{ik} e^{Q_{ii} s}\, ds
\end{aligned} \tag{27}$$

Substituting (26) and (27) into (25) yields (16.83) given above.
(c) Differentiating both sides of (16.83) with respect to $t$, we have

$$\begin{aligned}
\frac{dD_{ij}(t)}{dt} &= \delta_{ij}\lambda_i(Q_{ii} - \lambda_i)e^{(Q_{ii}-\lambda_i)t} + e^{(Q_{ii}-\lambda_i)t}\left[e^{-(Q_{ii}-\lambda_i)t}\sum_{k\neq i} Q_{ik} D_{kj}(t)\right] \\
&\quad + (Q_{ii} - \lambda_i)e^{(Q_{ii}-\lambda_i)t}\int_0^t e^{-(Q_{ii}-\lambda_i)s}\sum_{k\neq i} Q_{ik} D_{kj}(s)\, ds \\
&= \sum_{k\neq i} Q_{ik} D_{kj}(t) + (Q_{ii} - \lambda_i)\left[\delta_{ij}\lambda_i e^{(Q_{ii}-\lambda_i)t} + \int_0^t e^{-(Q_{ii}-\lambda_i)s}\sum_{k\neq i} Q_{ik} D_{kj}(s)\, ds\right] \\
&= \sum_{k\neq i} Q_{ik} D_{kj}(t) + (Q_{ii} - \lambda_i) D_{ij}(t).
\end{aligned} \tag{28}$$

The last equation above implies

$$\frac{d\boldsymbol{D}(t)}{dt} = (\boldsymbol{Q} - \boldsymbol{\Lambda})\boldsymbol{D}(t). \tag{29}$$

The first-order differential equation (29), together with the initial condition $\boldsymbol{D}(0) = \boldsymbol{\Lambda}$ lead to the desired result

$$\boldsymbol{D}(t) = \exp[(\boldsymbol{Q} - \boldsymbol{\Lambda})t]\boldsymbol{\Lambda}. \tag{16.84}$$

(c) In this case, we have

$$\boldsymbol{\Lambda} = \begin{bmatrix} 0 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \tag{30}$$

and from (16.84), we see that $\boldsymbol{D}(t)$ has the form

$$\boldsymbol{D}(t) = \begin{bmatrix} 0 & D_{11}(t) \\ 0 & D_{12}(t) \end{bmatrix}. \tag{31}$$

This implies that for all of the renewal points $t_n$, except possibly the first renewal point $t_0$, we have $X_n = 2$. By Problem 16.4 (a), the renewal points form a (possibly delayed) renewal process.

## 16.3 Reversible Markov chains

**16.12* Converse of reversed balance equation for DTMC.** TBD
We have an ergodic DTMC $\{X_n\}$ with TPM $\boldsymbol{P}$. Let $\tilde{\boldsymbol{P}}$ be a TPM and $\boldsymbol{\pi} = [\pi_i]$, $i \in \mathcal{S}$ be a probability distribution, such that the reversed balance equations hold:

$$\pi_i \tilde{P}_{ij} = \pi_j P_{ji}, \quad i, j \in \mathcal{S}. \tag{16.57}$$

Summing both sides of (16.57) over $j \in \mathcal{S}$ and using the fact that each row of $\tilde{P}$ must sum to one, we have

$$\pi_i = \sum_{j \in \mathcal{S}} \pi_j P_{ji},$$

i.e., $\boldsymbol{\pi}^\top = \boldsymbol{\pi}^\top \boldsymbol{P}$. Since $\{X_n\}$ is ergodic, $\boldsymbol{\pi}$ is the unique stationary distribution of $\{X_n\}$. From (16.60), we have

$$P[\tilde{X}_n = x_0 \mid \tilde{X}_{n-1} = x_1] = \frac{\pi_{x_0} P_{x_0 x_1}}{\pi_{x_1}}$$

Applying (16.57) to the RHS, we find that $P[\tilde{X}_n = x_0 \mid \tilde{X}_{n-1} = x_1] = \tilde{P}_{x_1 x_0}$. Hence, $\tilde{\boldsymbol{P}}$ is the TPM of the reversed process $\{\tilde{X}_n\}$. To show that $\boldsymbol{\pi}$ is the stationary distribution of $\{\tilde{X}_n\}$, we sum both sides of (16.57) over $i \in \mathcal{S}$, which leads to the conclusion $\boldsymbol{\pi}^\top = \boldsymbol{\pi}^\top \tilde{\boldsymbol{P}}$. Therefore, $\boldsymbol{\pi}$ is the unique stationary distribution of $\{\tilde{X}_n\}$.

**16.13 Kolmogorov's criterion for DTMC.**
Let $\{X_n\}$ be an ergodic DTMC. Suppose $\{X_n\}$ is reversible. By Theorem 16.6, its TPM $\boldsymbol{P}$ and stationary distribution $\boldsymbol{\pi}$ satisfy the detailed balance equations (16.63). Let $\{x_1, x_2, \ldots, x_n\}$ be any sequence of states in the state space $\mathcal{S}$. By the detailed balance equations we have

$$\begin{aligned}
\pi_{x_1} P_{x_1 x_2} &= \pi_{x_2} P_{x_2 x_1} \\
\pi_{x_2} P_{x_2 x_3} &= \pi_{x_3} P_{x_3 x_2} \\
&\cdots\cdots\cdots \\
\pi_{x_{n-1}} P_{x_{n-1} x_n} &= \pi_{x_n} P_{x_{n-1} x_n} \\
\pi_{x_n} P_{x_n x_1} &= \pi_{x_1} P_{x_1 x_n}.
\end{aligned}$$

Multiplying the above equations together and canceling out the product $\pi_{x_1} \pi_{x_2} \cdots \pi_{x_n}$, which is positive, since $\{X_n\}$ is ergodic, we obtain Kolmogorov's criterion

$$P_{x_1 x_2} P_{x_2 x_3} \cdots P_{x_{n-1} x_n} P_{x_n x_1} = P_{x_1 x_n} P_{x_n x_{n-1}} \cdots P_{x_3 x_2} P_{x_2 x_1}. \tag{32}$$

Conversely, suppose that (32) for any sequence of states $\{x_1, x_2, \ldots, x_n\}$. Summing both sides of (32) over $x_3, x_4, \ldots, x_n$, we have

$$\sum_{x_3, x_4, \ldots, x_n} P_{x_1 x_2} P_{x_2 x_3} \cdots P_{x_{n-1} x_n} P_{x_n x_1} = \sum_{x_3, x_4, \ldots, x_n} P_{x_1 x_n} P_{x_n x_{n-1}} \cdots P_{x_3 x_2} P_{x_2 x_1}. \tag{33}$$

The LHS can be written as

$$\text{LHS} = P_{x_1x_2}[\boldsymbol{P}^2]_{x_2x_4} \sum_{x_4,\ldots,x_n} P_{x_4x_5}P_{x_5x_6}\cdots P_{x_nx_1} \tag{34}$$

$$= P_{x_1x_2}[\boldsymbol{P}^3]_{x_2x_5} \sum_{x_5,\ldots,x_n} P_{x_5x_6}\cdots P_{x_nx_1} \tag{35}$$

$$= \cdots = P_{x_1x_2}[\boldsymbol{P}^{n-1}]_{x_2x_1}. \tag{36}$$

Similarly, the RHS of (33) is given by

$$\text{RHS} = [\boldsymbol{P}^{n-1}]_{x_1x_2}P_{x_2x_1}.$$

Hence, we have

$$P_{x_1x_2}[\boldsymbol{P}^{n-1}]_{x_2x_1} = [\boldsymbol{P}^{n-1}]_{x_1x_2}P_{x_2x_1}. \tag{37}$$

Now $\boldsymbol{P}^n$ is the $n$-step transition probability matrix of $\{X_n\}$, i.e., $\boldsymbol{P}^n = [P_{ij}^{(n)}]$, $i, j \in \mathcal{S}$. Hence, (37) can be written as

$$P_{x_1x_2}P_{x_2x_1}^{(n-1)} = P_{x_1x_2}^{(n-1)}P_{x_2x_1}. \tag{38}$$

Taking the limit as $n \to \infty$ on both sides and applying Theorem 15.9, we obtain the detailed balance equation

$$\pi_{x_1}P_{x_2x_1} = \pi_{x_2}P_{x_2x_1}.$$

By Theorem 16.6, $\{X_n\}$ is reversible. For an alternative approach to arrive at the detailed balance equations, see the solution to Problem 16.17.

**16.14*** (a) The LHS of (16.85) can be written as

$$\begin{aligned}
\text{LHS} &= \frac{P[\tilde{X}(t_m) = x_0, \tilde{X}(t_{m-1}) = x_1, \tilde{X}(t_{m-2}) = x_2, \ldots, \tilde{X}(t_0) = x_m]}{P[\tilde{X}(t_{m-1}) = x_1, \tilde{X}(t_{m-2}) = x_2, \ldots, \tilde{X}(t_0) = x_m]} \\
&= \frac{P[X(-t_m) = x_0, X(-t_{m-1}) = x_1, X(-t_{m-2}) = x_2, \ldots, X(-t_0) = x_m]}{P[X(-t_{m-1}) = x_1, X(-t_{m-2}) = x_2, \ldots, X(-t_0) = x_m]} \\
&= \frac{\pi_{x_0}P_{x_0x_1}(t_m - t_{m-1})P_{x_1x_2}(t_{m-1} - t_{m-2})\cdots P_{x_{m-1}x_m}(t_1 - 1_0)}{\pi_{x_1}P_{x_1x_2}(t_{m-1} - t_{m-2})\cdots P_{x_{m-1}x_m}(t_1 - 1_0)} \\
&= \frac{\pi_{x_0}P_{x_0x_1}(t_m - t_{m-1})}{\pi_{x_1}},
\end{aligned} \tag{39}$$

which is the RHS of (16.85). Since the RHS of (16.85) does not depend on $x_2, x_2, \ldots, x_m$, the second equality in (16.85) holds. Let $\tilde{\boldsymbol{P}}(t)$ denote the transition probability functions of $\{X(-t)\}$. Then the second equality in (16.85) implies (16.86):

$$\tilde{P}_{ij}(t) = \frac{\pi_j P_{ji}(t)}{\pi_i}, \quad i, j \in \mathcal{S}. \tag{16.86}$$

(b) Differentiating both sides of (16.86) by $t$, we have

$$\frac{d\tilde{P}_{ij}(t)}{dt} = \frac{\pi_j}{\pi_i}\frac{dP_{ji}(t)}{dt}, \quad i, j \in \mathcal{S}.$$

Setting $t = 0$ on both sides and re-arranging terms, we obtain the reversed balance equations (16.64) for the CTMC:

$$\pi_i \tilde{Q}_{ij} = \pi_j Q_{ij}, \quad i, j \in \mathcal{S}. \tag{16.64}$$

**16.15** We have an ergodic CTMC $\{X(t)\}$ with generator $\boldsymbol{Q}$. Let $\tilde{\boldsymbol{Q}}$ be a generator and $\boldsymbol{\pi} = [\pi_i]$, $i \in \mathcal{S}$ be a probability distribution, such that the CTMC reversed balance equations hold:

$$\pi_i \tilde{Q}_{ij} = \pi_j Q_{ji}, \quad i, j \in \mathcal{S}. \tag{16.64}$$

Summing both sides of (16.57) over $j \in \mathcal{S}$ and using the fact that each row of $\tilde{Q}$ must sum to zero, we have

$$0 = \sum_{j \in \mathcal{S}} \pi_j Q_{ji},$$

i.e., $\boldsymbol{\pi}^\top \boldsymbol{Q} = \boldsymbol{0}$. Since $\{X(t)\}$ is ergodic, $\boldsymbol{\pi}$ is the unique stationary distribution of $\{X(t)\}$. From Problem 16.14 (b), the transition rate of the reversed process $\{\tilde{X}(t)\}$ from state $i$ to $j$ is given by

$$\frac{\pi_j Q_{ij}}{\pi_i} = \tilde{Q}_{ij},$$

where the last equality is obtained by applying (16.64). Hence, $\tilde{\boldsymbol{Q}} = [\tilde{Q}_{ij}]$ is the generator matrix for $\{\tilde{X}(t)\}$. To show that $\boldsymbol{\pi}$ is the stationary distribution of $\{\tilde{X}(t)\}$, we sum both sides of (16.64) over $i \in \mathcal{S}$, which leads to the conclusion $\boldsymbol{\pi}^\top \tilde{\boldsymbol{Q}} = \boldsymbol{0}$. Therefore, $\boldsymbol{\pi}$ is the unique stationary distribution of $\{\tilde{X}(t)\}$.

**16.16*** Let $\{X(t)\}$ be an ergodic CTMC with generator $\boldsymbol{Q}$ and let $\{\tilde{X}(t)\}$ be its reversed process with generator $\tilde{\boldsymbol{Q}}$. The CTMC $\{X(t)\}$ is reversible if and only if

$$Q_{ij} = \tilde{Q}_{ij}, \quad i, j \in \mathcal{S}. \tag{40}$$

Applying (40) in the reversed balance equations (16.64), leads to the conclusion that $\{X(t)\}$ is reversible if and only if

$$\pi_i Q_{ij} = \pi_j Q_{ji}, \quad i, j \in \mathcal{S}. \tag{16.65}$$

**16.17**

Let $\{X(t)\}$ be an ergodic CTMC. Suppose $\{X_n\}$ is reversible. By Theorem 16.6, its generator $\boldsymbol{Q}$ and stationary distribution $\boldsymbol{\pi}$ satisfy the detailed balance equations (16.65). Let $\{x_1, x_2, \ldots, x_n\}$ be any sequence of states in the state space $\mathcal{S}$. By the detailed balance equations we have

$$\begin{aligned}
\pi_{x_1} Q_{x_1 x_2} &= \pi_{x_2} Q_{x_2 x_1} \\
\pi_{x_2} Q_{x_2 x_3} &= \pi_{x_3} Q_{x_3 x_2} \\
&\cdots\cdots\cdots \\
\pi_{x_{n-1}} Q_{x_{n-1} x_n} &= \pi_{x_n} Q_{x_{n-1} x_n} \\
\pi_{x_n} Q_{x_n x_1} &= \pi_{x_1} Q_{x_1 x_n}.
\end{aligned}$$

Multiplying the above equations together and canceling out the product $\pi_{x_1} \pi_{x_2} \cdots \pi_{x_n}$, which is positive, since $\{X(t)\}$ is ergodic, we obtain Kolmogorov's criterion for the CTMC:

$$Q_{x_1 x_2} Q_{x_2 x_3} \cdots Q_{x_{n-1} x_n} Q_{x_n x_1} = Q_{x_1 x_n} Q_{x_n x_{n-1}} \cdots Q_{x_3 x_2} Q_{x_2 x_1}. \tag{16.66}$$

Conversely, suppose that (16.66) holds for any sequence of states $\{x_1, x_2, \ldots, x_n\}$. We will show that for $i, j \in \mathcal{S}$, the detailed balance equations (16.65) hold. Our approach follows the proof given in the book by Kelly [177, Theorem 1.7]. Fix a reference state $x_0 \in \mathcal{S}$. Since $\{X(t)\}$ is irreducible, there is a sequence of states $\{i, x_n, x_{n-1}, \ldots, x_0\}$ such that $Q_{ix_n} Q_{x_n x_{n-1}} \cdots Q_{x_1 x_0} > 0$. We then define

$$\pi_i = C \frac{Q_{x_0 x_1} Q_{x_1 x_2} \cdots Q_{x_{n-1} x_n} Q_{x_n i}}{Q_{i x_n} Q_{x_n x_{n-1}} \cdots Q_{x_1 x_0}}. \tag{41}$$

By virtue of (16.66), this definition does not depend on the particular sequence of states $\{x_n, x_{n-1}, \ldots, x_0\}$. Next we show that $\pi_i$, $i \in \mathcal{S}$, satisfies the detailed balance equations (16.65). If $Q_{ij} = Q_{ji} = 0$, (16.66) is satisfied trivially. From (41), we have for $Q_{ji} > 0$

$$\pi_j = C \frac{Q_{x_0 x_1} Q_{x_1 x_2} \cdots Q_{x_{n-1} x_n} Q_{x_n i} Q_{ij}}{Q_{ji} Q_{i x_n} Q_{x_n x_{n-1}} \cdots Q_{x_1 x_0}} \tag{42}$$

$$= C \frac{\pi_i Q_{ij}}{Q_{ji}}, \tag{43}$$

where we applied (41) a second time to obtain the last equality. Hence, the detailed balance equations (16.65) hold. This in turn implies that the global balance equations given by $\boldsymbol{\pi Q} = 0$ with $\boldsymbol{\pi} = [\pi]$, $i \in \mathcal{S}$, are satisfied. Since the CTMC $\{X(t)\}$ is ergodic, $\pi$ is the unique stationary distribution, with an appropriate choice of the normalization constant $C$. Then by Theorem 16.10, the CTMC $\{X(t)\}$ is reversible.

**16.18** The generator $\boldsymbol{Q} = [Q_{ij}]$ of the BD process $N(t)$, given in (14.47), satisfies

$$Q_{ij} = \begin{cases} -\lambda_0, & i = j = 0, \\ \lambda_0, & i = 0, j = 1, \\ \mu_i, & i = j + 1, \\ -\lambda_i - \mu_i, & i = j, \\ \lambda_i, & i = j - 1, \\ 0, & |i - j| > 2. \end{cases} \tag{44}$$

Hence, when $|i - j| > 2$, $Q_{ij} = Q_{ji} = 0$, so detailed balance equations (16.65) are satisfied trivially. When $i = j$, we also see immediately that (16.65) is satisfied. When $j = i + 1$, the detailed balance equations become

$$\pi_i Q_{i,i+1} = \pi_{i+1} Q_{i+1,i}, \quad i \geq 0,$$

which is equivalent to

$$\lambda_i \pi_i = \mu_{i+1} \pi_{i+1}, \quad i \geq 0 \tag{45}$$

Therefore, it suffices to show that (45) holds.

To that end, consider the global balance equation $\boldsymbol{\pi}^\top \boldsymbol{Q} = \boldsymbol{0}$. This system of equations can be written as:

$$-\lambda_0 \pi_0 + \mu_1 \pi_1 = 0, \tag{46}$$

$$-\lambda_0 \pi_0 - (\lambda_1 + \mu_1) \pi_1 + \mu_2 \pi_2 = 0, \tag{47}$$

$$-\lambda_{i-1} \pi_{i-1} - (\lambda_i + \mu_i) \pi_i + \mu_{i+1} \pi_{i+1} = 0, \quad i \geq 2. \tag{48}$$

Equation (46) establishes that (45) holds when $i = 0$. Adding (46) to (47), we obtain

$$\lambda_1 \pi_1 = \mu_2 \pi_2, \tag{49}$$

which establishes (45) when $i = 1$. Proceeding by induction, assume that (45) holds for $i = n$, i.e.,

$$\lambda_n \pi_n = \mu_{n+1} \pi_{n+1}. \tag{50}$$

From (48) with $i = n + 1$, we have

$$\lambda_n \pi_n - (\lambda_{n+1} + \mu_{n+1})\pi_{n+1} + \mu_{n+2}\pi_{n+2} = 0. \tag{51}$$

Adding (50) to (51) and rearranging terms, we obtain

$$\lambda_{n+1} \pi_{n+1} = \mu_{n+2} \pi_{n+2}. \tag{52}$$

By the induction principle, we have established (45). Hence, the BD process $N(t)$ is reversible.

## 16.4 An application: phylogenetic tree and its Markov chain representation

**16.19** It is straightforward to verify that $\boldsymbol{Q}$ given in (16.77) and $\boldsymbol{\pi}$ given in (16.78) satisfy the global balance equation $\boldsymbol{\pi}^\top \boldsymbol{Q} = \boldsymbol{0}^\top$. Hence, $\boldsymbol{\pi}$ given in (16.78) is the stationary probability vector of the F81 model.

**16.20** Using a symbolic mathematical software package (e.g., MATLAB's Symbolic Toolbox, Maple, Mathematica, etc.), the generator $\boldsymbol{Q}$ in the F81 model given in (16.77) may be diagonalized as follows:

$$\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1}, \tag{53}$$

where $\boldsymbol{D} = \mathrm{diag}\{-1, -1, -1, 0\}$,

$$\boldsymbol{V} = \begin{bmatrix} \frac{-\pi_{\mathrm{C}}}{\pi_{\mathrm{A}}} & \frac{-\pi_{\mathrm{G}}}{\pi_{\mathrm{A}}} & \frac{-\pi_{\mathrm{T}}}{\pi_{\mathrm{A}}} & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{V}^{-1} = \begin{bmatrix} -\pi_{\mathrm{A}} & 1-\pi_{\mathrm{C}} & -\pi_{\mathrm{G}} & -\pi_{\mathrm{T}} \\ -\pi_{\mathrm{A}} & -\pi_{\mathrm{C}} & 1-\pi_{\mathrm{G}} & -\pi_{\mathrm{T}} \\ -\pi_{\mathrm{A}} & -\pi_{\mathrm{C}} & -\pi_{\mathrm{G}} & 1-\pi_{\mathrm{T}} \\ \pi_{\mathrm{A}} & \pi_{\mathrm{C}} & \pi_{\mathrm{G}} & \pi_{\mathrm{T}} \end{bmatrix}. \tag{54}$$

Using (16.72) and $\mu(e) = \rho(e)\tau(e)$, we have

$$\begin{aligned} \boldsymbol{P}(e) &= e^{\boldsymbol{Q}\mu(e)} = \boldsymbol{V}e^{\boldsymbol{D}\mu(e)}\boldsymbol{V}^{-1} \\ &= \begin{bmatrix} \pi_{\mathrm{A}} + (1-\pi_{\mathrm{A}})e^{-\mu(e)} & \pi_{\mathrm{C}}(1-e^{-\mu(e)}) & \pi_{\mathrm{G}}(1-e^{-\mu(e)}) & \pi_{\mathrm{T}}(1-e^{-\mu(e)}) \\ \pi_{\mathrm{A}}(1-e^{-\mu(e)}) & \pi_{\mathrm{C}} + (1-\pi_{\mathrm{C}})e^{-\mu(e)} & \pi_{\mathrm{G}}(1-e^{-\mu(e)}) & \pi_{\mathrm{T}}(1-e^{-\mu(e)}) \\ \pi_{\mathrm{A}}(1-e^{-\mu(e)}) & \pi_{\mathrm{C}}(1-e^{-\mu(e)}) & \pi_{\mathrm{G}} + (1-\pi_{\mathrm{G}})e^{-\mu(e)} & \pi_{\mathrm{T}}(1-e^{-\mu(e)}) \\ \pi_{\mathrm{A}}(1-e^{-\mu(e)}) & \pi_{\mathrm{C}}(1-e^{-\mu(e)}) & \pi_{\mathrm{G}}(1-e^{-\mu(e)}) & \pi_{\mathrm{T}} + (1-\pi_{\mathrm{T}})e^{-\mu(e)} \end{bmatrix}. \end{aligned} \tag{55}$$

From (55), it is apparent that (16.87) holds.

**16.21*** (a) It is easy to verify that the $(i, j)$ element of the matrix $\mathbf{\Pi}\boldsymbol{Q}$ is given by

$$[\mathbf{\Pi}\boldsymbol{Q}]_{ij} = \pi_i Q_{ij}, \quad i, j \in \mathcal{S}, \tag{56}$$

and that the $(i, j)$ element of the matrix $(\mathbf{\Pi}\boldsymbol{Q})^\top$ is given by

$$[(\mathbf{\Pi}\boldsymbol{Q})^\top]_{ij} = \pi_j Q_{ji}, \quad i, j \in \mathcal{S}. \tag{57}$$

It is then clear that the detailed balance equations (16.73) hold if and only if

$$[\mathbf{\Pi}\boldsymbol{Q}]_{ij} = [(\mathbf{\Pi}\boldsymbol{Q})^\top]_{ij},$$

i.e., if and only if $\mathbf{\Pi}\boldsymbol{Q}$ is a symmetric matrix.

(b) Given a DTMC with TPM $\boldsymbol{P}$ and stationary probability vector $\boldsymbol{\pi}$, we again define the matrix $\mathbf{\Pi} = \mathrm{diag}[\pi_i,\ i \in \mathcal{S}]$. Next, we verify that the $(i, j)$ element of the matrix $\mathbf{\Pi}\boldsymbol{P}$ is given by

$$[\mathbf{\Pi}\boldsymbol{P}]_{ij} = \pi_i P_{ij}, \quad i, j \in \mathcal{S}, \tag{58}$$

and that the $(i, j)$ element of the matrix $(\mathbf{\Pi}\boldsymbol{P})^\top$ is given by

$$[(\mathbf{\Pi}\boldsymbol{P})^\top]_{ij} = \pi_j P_{ji}, \quad i, j \in \mathcal{S}. \tag{59}$$

It is then clear that the detailed balance equations (16.63) hold if and only if

$$[\mathbf{\Pi}\boldsymbol{P}]_{ij} = [(\mathbf{\Pi}\boldsymbol{P})^\top]_{ij},$$

i.e., if and only if $\mathbf{\Pi}\boldsymbol{P}$ is a symmetric matrix.

(c) Let $\boldsymbol{P}(\tau) = [P_{ij}(\tau)]$, $i, j \in \mathcal{S}$, denote the matrix of transition probability functions of the given CTMC. By an argument similar to that given in parts (a) and (b), it suffices to show that the matrix $\mathbf{\Pi}\boldsymbol{P}(\tau)$ is symmetric for any $\tau > 0$. We have that $\boldsymbol{P}(\tau) = e^{\boldsymbol{Q}\tau}$. Hence, it suffices to show that

$$\mathbf{\Pi} e^{\boldsymbol{Q}\tau} = [\mathbf{\Pi} e^{\boldsymbol{Q}\tau}]^\top = e^{\boldsymbol{Q}^\top \tau}\mathbf{\Pi}. \tag{60}$$

By part (a), the CTMC is reversible if and only if

$$\mathbf{\Pi}\boldsymbol{Q} = \boldsymbol{Q}^\top \mathbf{\Pi}. \tag{61}$$

Now suppose that

$$\mathbf{\Pi}\boldsymbol{Q}^n = (\boldsymbol{Q}^\top)^n \mathbf{\Pi} \tag{62}$$

for $n \geq 1$. Then

$$\begin{aligned} \mathbf{\Pi}\boldsymbol{Q}^{n+1} &= (\mathbf{\Pi}\boldsymbol{Q}^n)\boldsymbol{Q} = (\boldsymbol{Q}^\top)^n \mathbf{\Pi}\boldsymbol{Q} \\ &= (\boldsymbol{Q}^\top)^n \boldsymbol{Q}^\top \mathbf{\Pi} = (\boldsymbol{Q}^\top)^{n+1}\mathbf{\Pi}. \end{aligned}$$

By the induction principle, (62) holds for all $n \geq 1$. Now we have

$$\begin{aligned}\mathbf{\Pi} e^{\boldsymbol{Q}\tau} &= \mathbf{\Pi} \sum_{k=0}^{\infty} \frac{(\boldsymbol{Q}\tau)^k}{k!} = \sum_{k=0}^{\infty} \frac{\mathbf{\Pi}\boldsymbol{Q}^k \tau^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{(\boldsymbol{Q}^\top)^k \tau^k \mathbf{\Pi}}{k!} = e^{\boldsymbol{Q}^\top \tau}\mathbf{\Pi},\end{aligned}$$

which establishes (60).

**16.22** The matrix $\mathbf{\Pi}\boldsymbol{Q}$ is given by

$$\mathbf{\Pi}\boldsymbol{Q} = \begin{bmatrix} -\pi_{\mathrm{A}}(\pi_{\mathrm{C}}+\pi_{\mathrm{G}}+\pi_{\mathrm{T}}) & \pi_{\mathrm{A}}\pi_{\mathrm{C}} & \pi_{\mathrm{A}}\pi_{\mathrm{G}} & \pi_{\mathrm{A}}\pi_{\mathrm{T}} \\ \pi_{\mathrm{A}}\pi_{\mathrm{C}} & -\pi_{\mathrm{C}}(\pi_{\mathrm{A}}+\pi_{\mathrm{G}}+\pi_{\mathrm{T}}) & \pi_{\mathrm{C}}\pi_{\mathrm{G}} & \pi_{\mathrm{C}}\pi_{\mathrm{T}} \\ \pi_{\mathrm{A}}\pi_{\mathrm{G}} & \pi_{\mathrm{C}}\pi_{\mathrm{G}} & -\pi_{\mathrm{G}}(\pi_{\mathrm{A}}+\pi_{\mathrm{C}}+\pi_{\mathrm{T}}) & \pi_{\mathrm{G}}\pi_{\mathrm{T}} \\ \pi_{\mathrm{A}}\pi_{\mathrm{T}} & \pi_{\mathrm{C}}\pi_{\mathrm{T}} & \pi_{\mathrm{G}}\pi_{\mathrm{T}} & -\pi_{\mathrm{T}}(\pi_{\mathrm{A}}+\pi_{\mathrm{C}}+\pi_{\mathrm{G}}) \end{bmatrix}$$

Since this matrix is symmetric, Problem 16.21 (a) shows that the detailed balance equations hold and hence the F81 model is reversible.

**16.23*** (a) Applying the given values into (16.77), we have

$$\boldsymbol{Q} = \begin{bmatrix} -0.9 & 0.2 & 0.2 & 0.5 \\ 0.1 & -0.8 & 0.2 & 0.5 \\ 0.1 & 0.2 & -0.8 & 0.5 \\ 0.1 & 0.2 & 0.2 & -0.5 \end{bmatrix}.$$

Using (16.72) with $\rho(e) = 1$ and $\tau(e) = 1$, we compute $\boldsymbol{P}(e)$ with four decimal places of precision:

$$\begin{aligned}\boldsymbol{P}(e) &= e^{\boldsymbol{Q}} \\ &= \begin{bmatrix} 0.4311 & 0.1264 & 0.1264 & 0.3161 \\ 0.0632 & 0.4943 & 0.1264 & 0.3161 \\ 0.0632 & 0.1264 & 0.4943 & 0.3161 \\ 0.0632 & 0.1264 & 0.1264 & 0.6839 \end{bmatrix}.\end{aligned}$$

(b) We first obtain the stationary distribution $\boldsymbol{\pi}$, which is the unique solution to

$$\boldsymbol{\pi}^\top \boldsymbol{Q} = 0, \quad \boldsymbol{\pi}^\top \mathbf{1} = 1.$$

Let $\boldsymbol{E}$ denote the matrix of all ones. Then stationary distribution can be computed as follows (cf. (15.107)):

$$\boldsymbol{\pi}^\top = \mathbf{1}^\top (\boldsymbol{Q} + \boldsymbol{E})^{-1} = [0.1, 0.2, 0.2, 0.5].$$

Let $\mathbf{\Pi} = \mathrm{diag}\{0.1, 0.2, 0.2, 0.5\}$. Applying (16.75), the mean substitutions that occur on an edge $e \in \mathcal{E}$ is given by

$$\kappa(e) = \mathrm{Tr}\{\mathbf{\Pi}\boldsymbol{Q}\} = 0.66.$$

(c) Let $X_v$ denote the random variable associated with node $v \in \{0, 1, 2, 3, 4\}$ in the phylogenetic tree of Figure 16.4. Let $\boldsymbol{X} = (X_0, X_1, X_2, X_3, X_4)$ denote the corresponding vector and let $\boldsymbol{x} = (x_0, x_1, x_2, x_3, x_4)$. The joint distribution of $\boldsymbol{X}$ is given by

$$\begin{aligned} P[\boldsymbol{X} = \boldsymbol{x}] = P[X_0 = x_0]P[X_1 = x_1 \mid X_0 = 0]P[X_2 = x_2 \mid X_1 = x_1] \\ P[X_3 = x_3 \mid X_1 = x_1]P[X_4 = x_4 \mid X_0 = x_0] \\ = \pi_{x_0} P_{x_0 x_1} P_{x_1 x_2} P_{x_1 x_3} P_{x_0 x_4}, \end{aligned}$$

where $P_{ij}$ is the $(i, j)$ element of the TPM $\boldsymbol{P}(e)$ given in part (a). The likelihood of the character $\chi_3$ is given by

$$\begin{aligned} P_{\chi_3} &= P[X_2 = \mathrm{C}, X_3 = \mathrm{T}, X_4 = \mathrm{A}] \\ &= \sum_{x_0, x_1} P[\boldsymbol{X} = (x_0, x_1, \mathrm{C}, \mathrm{T}, \mathrm{A})] \\ &= \sum_{x_0} \pi_{x_0} P_{x_0 \mathrm{A}} \sum_{x_1} P_{x_1 \mathrm{C}} P_{x_1 \mathrm{T}} P_{x_0 x_1} \\ &= \sum_{x_0} \pi_{x_0} P_{x_0 \mathrm{A}} f_{x_0}, \end{aligned} \tag{63}$$

where we define

$$f_{x_0} = \sum_{x_1} P_{x_1 \mathrm{C}} P_{x_1 \mathrm{T}} P_{x_0 x_1}.$$

We compute to four decimal places,

$$f_{\mathrm{A}} = 0.0694, f_{\mathrm{C}} = 0.1121, f_{\mathrm{G}} = 0.0694, f_{\mathrm{T}} = 0.0865. \tag{64}$$

Substituting into (63), we obtain $P_{\chi_3} = 0.0080$.

# 17 Solutions for Chapter 17: Random Walk, Brownian Motion and Diffusion Process

## 17.1 Random Walk

**17.1 Alternative derivation of (17.7).**

(a) The event $X_n = k$ is reachable from either $X_{n-1} = k-1$ with a positive step $S_n = +1$ or from $X_{n-1} = k+1$ with a negative step $S_n = -1$, and these events are independent. Hence

$$P[X_n = k] = P[X_{n-1} = k-1]P[S_n = +1] + P[X_{n-1} = k+1]P[S_n = -1],$$

which gives (17.8). The initial position is $X_0 = 0$ with probability one, which gives (17.9).
(b) As for (17.9),

$$\text{LHS} = P[X_0 = k] = \binom{0}{\frac{k}{2}} \left(\frac{p}{q}\right)^{k/2}.$$

Since $\binom{0}{0} = 1$ and $\binom{0}{k} = 0$ for all $k \neq 0$, and $\left(\frac{p}{q}\right)^{k/2} = 1$ for $k = 0$, we see the above expression equals $\delta_{k,0}$.
As for (17.8),

$$\text{LHS} = P[X_{n+1} = k] = \binom{n+1}{\frac{n+k+1}{2}} p^{\frac{n+k+1}{2}} q^{\frac{n-k+1}{2}}.$$

and

$$\begin{aligned}
\text{RHS} &= pP[X_n = k-1] + qP[X_n = k+1] \\
&= p\binom{n}{\frac{n+k-1}{2}} + q\binom{n}{\frac{n+k+1}{2}} p^{\frac{p+k+1}{2}} q^{\frac{n-k-1}{2}} \\
&= \left[\binom{n}{\frac{n+k-1}{2}} + \binom{n}{\frac{n+k+1}{2}}\right] p^{\frac{p+k+1}{2}} q^{\frac{n-k-1}{2}}.
\end{aligned}$$

Thus, it suffices to show

$$\binom{n+1}{\frac{n+k+1}{2}} = \binom{n}{\frac{n+k-1}{2}} + \binom{n}{\frac{n+k+1}{2}},$$

which readily follows the following formula, known as Pascal's triangle:

$$\binom{n+1}{i+1} = \binom{n}{i} + \binom{n}{i+1}.$$

**17.2* Properties of the simple random walk.**

(i) Spatial homogeneity: Both LHS and RHS of (17.3)equal $P[\sum_{i=1^n} S_i = k - a]$, because $X_n - X_0 = k - a = (k + b) - (k + a)$.

(ii) Temporal homogeneity: LHS of (17.4) is equal to $P\left[\sum_{i=1}^{n} S_i\right]$ and the RHS is equal to $P\left[\sum_{i=m+1}^{m+n} S_i\right]$. Both involve the sum of $n$ i.i.d. RVs $S_i$s, their probability distributions must be identical.

(iii) Independent increment: We can write $X_{n_i} - X_{m_i} = \sum_{j\in(m_i,n_i]} S_j$. If the set of intervals $(m_i, n_i]$'s are mutually disjoint, then all the $S_j$ terms contributing to the increments $X_{n_i} - X_{m_i}$'s are mutually independent.

(iv) If we know $X_n$, then the probability distribution of $X_{n+m}$ depends only on the steps $S_{n+1}, S_{n+2}, \ldots, S_{n+m}$, and the values of $X_0, X_1, \ldots, X_{n-1}$ are not relevant.

**17.3 Gambler ruin problem.** The system of equations (17.22) should now be modified as

$$r_i = pr_{i+1} + qr_{i-1}, \;\; i = 1, 2, \ldots, c - 1,$$

with the boundary condition $r_0 = 1$ and $r_c = 0$. Multiply the above equation by $z^i$ and sum over $i = 1$ to $i = c - 1$:

$$\sum_{i=1}^{c-1} r_1 z^i = p \sum_{i=1}^{c-1} r_{i+1} z^i + q \sum_{i=1}^{c-1} r_{i-1} z^i.$$

By defining the generating function

$$R(z) = \sum_{i=0}^{c} r_i z^i = \sum_{i=0}^{c-1} r_i z^i,$$

we have

$$R(z) - 1 = pz^{-1}[R(z) - r_1 z - 1] + qz[R(z) - r_{c-1} z^{c-1}], \tag{1}$$

which we can rewrite as

$$R(z)(1 - pz^{-1} - qz) = 1 - (r_1 z + 1)pz^{-1} - qr_{c-1} z^c.$$

Multiplying the above by $z$, and setting $q = 1 - p$, we have

$$[R(z)(1 - z)(qz - p) = z - (r_1 z + 1)p - qr_{c-1} z^{c+1}. \tag{2}$$

By setting $z = 1$ and $z = \frac{p}{q}$, we have

$$\begin{aligned} 0 &= 1 - (r_1 + 1)p - qr_{c-1} \\ 0 &= \frac{p}{q} - \left(r_1 \frac{p}{q} + 1\right) p - qr_{c-1} \left(\frac{p}{q}\right)^{c+1}, \end{aligned}$$

from which we have

$$\begin{aligned} pr_1 + qr_{c-1} &= q \\ pr_1 + \left(\frac{p}{q}\right)^c qr_{c-1} &= p. \end{aligned}$$

Then when $p \neq q$, we find

$$r_{c-1} = \frac{\gamma^c - \gamma^{c-1}}{\gamma^c - 1}, \text{ where } \gamma = \frac{q}{p},$$
$$r_1 = \frac{\gamma^c - \gamma}{\gamma^c - 1}.$$

Substituting these results into (2), we can obtain

$$R(z) = \frac{1}{\gamma^c - 1} \frac{(1-\gamma z)\gamma^c(1-z^c) - (1-z)(1-\gamma^c z^c)}{(1-z)(1-\gamma z)},$$

from which we find

$$R(z) = \frac{1}{\gamma^c - 1}\left[\frac{\gamma^c(1-z^c)}{1-z} - \frac{1-\gamma^c z^c}{1-\gamma z}\right] = \sum_{i=0}^{c-1} \frac{\gamma^c - \gamma^i}{\gamma^c - 1} z^i.$$

Thus,

$$r_i = \frac{\gamma^c - \gamma^i}{\gamma^c - 1}, \quad i = 0, 1, 2, \ldots, c. \tag{3}$$

For $p = q$, i.e., $\gamma = 1$, we find the solution by letting $\gamma \to 1$ in the above and using l'Hôpital's (sometimes spelled as l'Hospital's) rule, obtaining

$$r_i = \frac{c-i}{c} = 1 - \frac{i}{c}.$$

A somewhat quicker and simpler approach is as follows. Once we have found the characteristic roots $z = 1$ and $z = \frac{p}{q} = \gamma^{-1}$ from (2), we can assume the following solution form, :

$$r_i = A + B\gamma^i,$$

which clearly satisfies the system equations for $1 \leq i \leq c-1$. The constants $A$ and $B$ can be uniquely determined from the boundary conditions at $i = 0$ and $i = c$, i.e.,

$$1 = A + B, \text{ and } 0 = A + B\gamma^c,$$

from which

$$A = -B = \frac{1}{\gamma^c - 1}, \ \gamma = \frac{q}{p}.$$

Thus,we obtain (3). When $p = q$, $z = 1$ is a double root. Then the solution form should be

$$r_i = A' + B'i,$$

From the boundary condition we find

$$A' = 1, \ B' = -\frac{1}{c}.$$

**17.4 Expected duration of the game.** States $i : 1 \leq i \leq c-1$ are recurrent states and states 0 and $c$ are absorbing states. Recall the formula for the mean passage-time $\mu_{ij}$ of (15.62)

$$\mu_{ij} = P_{ij} + \sum_{k \neq j} P_{ik}(\mu_{kj} + 1), \tag{4}$$

The game ends when the system enters the absorbing state 0 (i.e, player $A$ loses), or when the system enters the other absorbing state $c$ (namely, player $A$ wins). These events will occur with probability $f_{i,0} = r_i$ and $f_{i,c} = 1 - r_i$, respectively, as we obtained in the previous Problem. Thus,

$$\mu_i = \mu_{i,0} r_i + \mu_{i,c}(1 - r_i).$$

By setting $j = 0$ or $j = c$ in (4), we have

$$\mu_{i,0} = q\delta_{i,1} + p\mu_{i+1,0} + q\mu_{i-1,0} + 1, \;\; 1 \leq i \leq c-1, \tag{5}$$

$$\mu_{i,c} = p\delta_{i,c-1} + p\mu_{i+1,c} + q\mu_{i-1,c} + 1, \;\; 1 \leq i \leq c-1, \tag{6}$$

with the boundary conditions

$$\mu_{0,0} = 0, \; \mu_{c,0} = \infty, \; \mu_{0,c} = \infty, \;\; \text{and} \;\; \mu_{c,c} = 0.$$

By multiplying (5) by $r_i$ and (6) by $(1 - r_i)$ and summing them up, we find

$$\mu_i = q\delta_{i,1} r_1 + p\delta_{i,c-1}(1 - r_{c-1}) + p\mu_{i+1} + q\mu_{i-1} + 1, \;\; 1 \leq i \leq c-1, \tag{7}$$

with the boundary conditions

$$\mu_0 = 0, \;\; \text{and} \;\; \mu_c = 0. \tag{8}$$

Define the generating function $M(z)$ of the expected duration of the game $\mu_i$ from the state $i$:

$$M(z) = \sum_{i=0}^{c} \mu_i z^i = \sum_{i=1}^{c-1} \mu_i z^i.$$

By multiplying (7) by $z^i$ and summing it over $i = 1$ to $i = c-1$, we have

$$M(z) = qr_1 z + p(1 - r_{c-1})z^{c-1} + pz^{-1}(M(z) - \mu_1 z) + qz(M(z) - \mu_{c-1}z^{c-1}) + \frac{z - z^c}{1 - z}, \tag{9}$$

from which we obtain

$$M(z)(1 - pz^{-1} - qz) = qr_1 z + p(1 - r_{c-1})z^{c-1} - p\mu_1 - q\mu_{c-1}z^c + \frac{z - z^c}{1 - z},$$

or

$$-qz^{-1}(z-1)\left(z - \frac{p}{q}\right) M(z) = qr_1 z + p(1 - r_{c-1})z^{c-1} - p\mu_1 - q\mu_{c-1}z^c + \frac{z - z^c}{1 - z}. \tag{10}$$

By setting $z = 1$, we have

$$0 = qr_1 + p(1 - r_{c-1}) - p\mu_1 - q\mu_{c-1} + c - 1. \tag{11}$$

Similarly, by setting $z = \frac{p}{q}$, we find

$$0 = pr_1 + p(1 - r_{c-1})\left(\frac{p}{q}\right)^{c-1} - p\mu_1 - p\mu_{c-1}\left(\frac{p}{q}\right)^{c-1} + c - 1. \tag{12}$$

**Case 1: When $\gamma = \frac{q}{p} = 1, i.e., p = q = \frac{1}{2}$.** In this case, we have from (17.32)

$$r_1 = \frac{c-1}{c}, \text{ and } r_{c-1} = \frac{1}{c}.$$

Equation (10) (multiplied by $1/q(=2)$) becomes

$$-z^{-1}(z-1)^2 M(z) = \frac{c-1}{c} z + \frac{c-1}{c} z^{c-1} - (\mu_1 + \mu_{c-1} z^c) + \frac{2(z - z^c)}{1-z}, \tag{13}$$

Both (11) and (12) (multiplied by two) reduce to the same equation:

$$0 = 2\frac{c-1}{c} - (\mu_1 + \mu_{c-1}) + 2(c-1). \tag{14}$$

By differentiating (13), we have

$$\begin{aligned} &- \frac{[2(z-1)M(z) + (z-1)^2 M'(z)]z - (z-1)^2 M(z)}{z^2} \\ &= \frac{c-1}{c} + \frac{(c-1)^2}{c} z^{c-2} - \mu_{c-1} c z^{c-1} + \frac{2[1 - cz^{c-1} + (c-1)z^c]}{(1-z)^2}. \end{aligned}$$

Multiplying the above by $c$ and setting $z = 1$, we obtain

$$0 = \frac{c-1}{c} - \frac{(c-1)c^2}{c} - \mu_{c-1} c + c(c-1), \tag{15}$$

from which we have

$$\mu_{c-1} = 2(c-1),$$

and substitution of this into (14) leads to

$$\begin{aligned} \mu_1 &= \frac{2(c-1)}{c} + 2(c-1) - \mu_{c-1} = \frac{2(c-1)(c+1)}{c} - 2(c-1) \\ &= \frac{2(c-1)}{c} \end{aligned}$$

Then (9) becomes

$$M(z) = \frac{(c-1)z}{2c} + \frac{(c-1)z^{c-1}}{2c} + \frac{z^{-1}}{2}(M(z) - \mu_1 z) + \frac{z}{2}(M(z) - \mu_{c-1} z^{c-1}) + \frac{z - z^c}{1-z}, \tag{16}$$

or

$$-\frac{(z-1)^2}{2z} M(z) = -(c-1)z^c + \frac{(c-1)z^{c-1}}{2c} + \frac{(c-1)z}{2c} - \frac{c-1}{c} + \frac{z-z^c}{1-z},$$

from which we find

$$\begin{aligned} M(z) &= \frac{1}{c(z-1)^2}\left[2c(c-1)z^{c+1} - (c-1)z^c - (c-1)z^2 + 2(c-1)z - \frac{2cz^2(1-z^{c-1})}{1-z}\right] \\ &= \frac{1}{c(1-z)^3}\left[-2c(c-1)z^{c+2} + (2c^2 + c - 1)z^{c+1} - (c-1)z^c \right. \\ &\left. +(z-1)z^3 - (5c-3)z^2 + 2(c-1)z\right] \end{aligned} \tag{17}$$

By definition $M(z)$ is a polynomial in $z$ of order $c-1$. Then $\mu_i,\ \ 1 \le i \le c-1$ can be found by inverting $M(z)$. (To be done)

**Case 2: When $\gamma = \frac{q}{p} \neq 1$.** In this case, from (17.31)

$$r_1 = \frac{\gamma^c - \gamma}{\gamma^c - 1}, \ \text{ and } \ r_{c-1} = \frac{\gamma^c - \gamma^{c-1}}{\gamma^c - 1},$$

and

$$1 - r_{c-1} = \frac{\gamma^{c-1} - 1}{\gamma^c - 1}, \ \text{ and } \ r_1 = \gamma(1 - r_{c-1}).$$

By subtracting (12) from (11), we find

$$0 = (q-p)r_1 + p(1 - r_{c-1})(1 - \gamma^{-c+1}) - \mu_{c-1}(q - p\gamma^{-c+1}),$$

from which we can solve for $\mu_{c-1}$:

$$\mu_{c-1} = \frac{q(1-\gamma^{-1})r_1 + p(1 - r_{c-1})(1 - \gamma^{-c+1})}{q(1 - \gamma^{-c})}.$$

Using the result (17.31), and after some manipulation we obtain

$$\mu_{c-1} = \frac{(\gamma^{c-1} - 1)(\gamma^c - q\gamma^{c-1} - p\gamma)}{q(\gamma^c - 1)^2}. \tag{18}$$

If we let $\gamma \to 1, p = q \to \frac{1}{2}$ in this expression, we find, applying l'Hopital's rule a few times,

$$\lim_{\gamma \to 1} \mu_{c-1} = \frac{c-1}{c},$$

which is off from the result in case 1 by factor of 2(?).
By substituting (18) into (9), we find

$$\begin{aligned}
p\mu_1 &= qr_1 + p(1 - r_{c-1}) - q\mu_{c-1} + c - 1 \\
&= q\frac{\gamma^c - \gamma}{\gamma^c - 1} + p\frac{\gamma^{c-1} - 1}{\gamma^c - 1} - \frac{(\gamma^{c-1} - 1)(\gamma^c - q\gamma^{c-1} - p\gamma)}{(\gamma^c - 1)^2} + c - 1 \\
&= \frac{(\gamma^{c-1} - 1)}{(\gamma^c - 1)^2}\left[(\gamma^c - 1)(q\gamma + p) - (\gamma^c - q\gamma^{c-1} - p\gamma)\right] + c - 1 \\
&= \frac{(\gamma^{c-1} - 1)(\gamma - 1)}{(\gamma^c - 1)^2}\left[q\gamma^c + q\gamma\frac{(\gamma^{c-2} - 1)}{\gamma - 1} + p\right] + c - 1
\end{aligned}$$

In the limit $\gamma \to 1$ and $p = q \to \frac{1}{2}$, we have

$$\lim \mu_1 = 2??$$

## 17.2 Brownian Motion or Wiener Process

**17.5 Properties of Wiener process.**

(a) The process $W(t)$, like the simple random walk $X_n$, has stationary independent increment with zero mean. From the temporal homogeneity, $W(t) - W(0)$ has the same distribution as $W(t+s) - W(s)$. Thus,

$$\mathrm{Var}[W(t)] = \mathrm{Var}[W(t+s) - W(s)]. \tag{19}$$

Since the increments are independent

$$\mathrm{Var}[W(t+s) - W(s)] + \mathrm{Var}[X(s)] = \mathrm{Var}[W(t+s)]. \tag{20}$$

Thus, if we define

$$g(t) = \mathrm{Var}[W(t)],$$

Then from (19) and (20), we have

$$g(t) + g(s) = g(t+s),\ t, s > 0. \tag{21}$$

The only real-valued nonnegative function $g(t)$ that satisfies this relationship is

$$g(t) = \alpha t,$$

where $\alpha$ is an arbitrary nonnegative constant.

(b) Let $t + s = u$ in (19). Then $u > s$, and

$$\mathrm{Var}[W(u) - W(s)] = \mathrm{Var}[W(u-s)] = g(u-s) = \alpha(u-s),\ \ u \geq s.$$

For $u < s$, we have, using the property $\mathrm{Var}[-X] = \mathrm{Var}[X]$,

$$\mathrm{Var}[W(u) - W(s)] = \mathrm{Var}[W(s) - W(u)] = g(s-u) = -\alpha(s-u),\ \ u < s.$$

Thus, combining the above two

$$\mathrm{Var}[W(t) - W(s)] = \alpha|t-s|.$$

**17.6 Transformation of a Wiener process.**

(a) . Let $t^2 = u$ and $s^2 = v$. Then

$$R(t,s) = E[W(t^2)W(s^2)] = E[W(u)W(v)] = \alpha \min(u,v) = \alpha(t^2, s^2).$$

(b) Recall the following formula of the moments of the multivariate normal RVs.

$$E[X_1X_2X_3X_4] = E[X_1X_2]E[X_3X_4] + E[X_1X_3]E[X_2X_4] + E[X_1X_4]E[X_2X_3] - 2m_1m_2m_3m_4.$$

Let $X_1 = X_2 = W(t_1)$ and $X_3 = X_4 = W(t_2)$. Then

$$E[Y(t_1)Y(t_2)] = E[X^2(t_1)]E[X^2(t_2)] + 2(E[W(t_1)W(t_2)])^2 = \alpha^2 t_1 t_2 + 2\alpha^2 t_1^2 = \alpha^2 t_1(t_2 + 2t_1),\ \ t_1 < t_2.$$

### 17.2.1 Diffusion Process and diffusion equations

**17.7 Diffusion equations of Brownian motion with drift.**

The forward equation (17.99) reduces to

$$\frac{\partial f(x,t|x_0,t_0)}{\partial t} = -\beta \frac{\partial f(x,t|x_0,t_0)}{\partial x} + \frac{\alpha}{2}\frac{\partial^2 f(x,t|x_0,t_0)}{\partial x^2}. \tag{22}$$

The backward equation becomes

$$\frac{\partial f(x,t|x_0,t_0)}{\partial t_0} = -\beta \frac{\partial f(x,t|x_0,t_0)}{\partial x_0} - \frac{\alpha}{2}\frac{\partial^2 f(x,t|x_0,t_0)}{\partial x_0^2}. \tag{23}$$

**17.8 Conditional PDFs of the standard Brownian motion.**

(a) We can use the formula for the conditional PDF of the binomial normal distribution, but we will show its steps below.
Since $R_{W_s}(t,s) = \min(t,s)$, the correlation matrix (i.e., covariance matrix, since mean zero) of $W_s(t) = x$ and $W_s(t_0) = x_0$ ($t < t_0$) is

$$\boldsymbol{C} = \begin{bmatrix} t & t \\ t & t_0 \end{bmatrix}.$$

Hence

$$\boldsymbol{C}^{-1} = \frac{1}{\det \boldsymbol{C}} \begin{bmatrix} t_0 & -t \\ -t & t \end{bmatrix}.$$

Thus, the joint PDF of $\boldsymbol{X} = (W_s(t), W_s(t_0))^\top$ is

$$f(x,x_0) = \frac{1}{2\pi\sqrt{t(t_0-t)}} \exp\left(-\frac{t_0x^2 - 2tx_0x + tx_0^2}{2t(t_0-t)}\right),$$

and

$$f(x_0) = \frac{1}{\sqrt{2\pi t_0}} \exp\left(-\frac{x_0^2}{2t_0}\right).$$

Thus, the conditional PDF is

$$f(x|x_0) = \frac{f(x,x_0)}{f(x_0)} = \frac{1}{\sqrt{\frac{t(t_0-t)}{t_0}}} \exp - \left\{ -\frac{\left(x - \frac{tx_0}{t_0}\right)^2}{\frac{2t(t_0-t)}{t_0}} \right\}.$$

Hence it is the normal distribution with mean $\frac{x_0}{t_0}t$ and variance $\frac{t_0-t}{t_0}t$.
(b) Since the process $W_s(t)$ satisfies both the spatial and temporal homogeneity, we consider the displacement from $t_1$ and $x_1$ and denote

$$t' = t - t_1,\;\; t'_0 = t_2 - t_1,\;\; x' = x - x_1,\;\; \text{and } x'_0 = x_2 - x_1.$$

Then from part (a), we know the conditional PDF of $x' = x - x_1$ is a normal distribution with mean

$$\frac{x'_0}{t'_0}t' = \frac{x_2 - x_1}{t_2 - t_1}(t - t_1),$$

and variance

$$\frac{(t'_0 - t')t'}{t'_0} = \frac{(t_2 - t_1)(t - t_1)}{t_2 - t_1}.$$

Hence the conditional distribution of $x$ has the same variance above and mean

$$x_1 + \frac{x_2 - x - 1}{t_2 - t_1}(t - t_1).$$

**17.9 Solution of the Fokker-Plank equation(17.77).**

(a) Suppose the steady-state distribution

$$\pi(x) = \lim_{t\to\infty} f(x,t)$$

exists. Then it has to satisfy the equation (17.77) with the LHS set to zero:

$$0 = -\beta \frac{d\pi(x)}{dx} + \frac{\alpha}{2}\frac{d^2\pi(x)}{dx^2}.$$

Then

$$\frac{d^2\pi(x)}{dx^2} = \frac{2\beta}{\alpha}\frac{d\pi(x)}{dx},$$

which gives

$$\frac{d\pi(x)}{dx} = Ae^{\frac{2\beta x}{\alpha}},$$

for some constant $A$. Then

$$\pi(x) = \frac{A\alpha}{2\beta}e^{\frac{2\beta x}{\alpha}} + B.$$

In order to make $\int_0^\infty \pi(x)\,dx < \infty$, we must set $A = B = 0$. Then $\pi(x) = 0$ for all $x \geq 0$. Thus, there is no bona fide PDF.

(b) In order to show (17.78) is the solution of (17.77), we evaluate the LHS and RHS:

$$\begin{aligned}
\text{LHS} &= \frac{\partial f(x,t)}{\partial t}\\
&= -\frac{2\pi\alpha}{2}(2\pi\alpha t)^{-\frac{3}{2}}e^{-\frac{(x-\beta t)^2}{2\alpha t}} - (2\pi\alpha t)^{-\frac{1}{2}}e^{-\frac{(x-\beta t)^2}{2\alpha t}}\frac{d}{dt}\left(\frac{(x-\beta t)^2}{2\alpha t}\right)\\
&= -\frac{1}{t}f(x,t) + \frac{x^2-\beta^2t^2}{2\alpha t^2} = \frac{x^2-\alpha t-\beta^2t^2}{2\alpha t^2}f(x,t).
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial f(x,t)}{\partial x} &= \frac{1}{\sqrt{2\pi\alpha t}}\left(-\frac{x-\beta t}{\alpha t}\right)e^{-\frac{(x-\beta t)^2}{2\alpha t}} = -\frac{x-\beta t}{\alpha t}f(x,t).\\
\frac{\partial^2 f(x,t)}{\partial x^2} &= -\frac{1}{\alpha t}f(x,t) + \left(\frac{x-\beta t}{\alpha t}\right)^2 f(x,t).
\end{aligned}$$

Hence

$$\begin{aligned}
\text{RHS} &= -\beta\frac{\partial f9x,t)}{\partial x} + \frac{\alpha}{2}\frac{\partial^2 f(x,t)}{\partial x^2}\\
&= \frac{\beta(x-\beta t)}{\alpha t}f(x,t) - \frac{f(x,t)}{2t} + \frac{(x-\beta t)^2}{2\alpha t^2}f(x,t)\\
&= \frac{x^2-\alpha t-\beta^2t^2}{2\alpha t^2}f(x,t)
\end{aligned}$$

Hence we have shown that (17.77) holds.

**17.10* Derivation of (17.104) and (17.106).**

(a) Let $X$ be a RV with mean $\mu$ and variance $\sigma^2$ and the PDF $f(x)$. Then for any function $g(x)$ that is continuous and at least twice differentiable at $x = \mu$, we can expand $g(x)$ using the Taylor series expansion:

$$g(x) = g(\mu) + g'(\mu)(x-\mu) + g''(\mu)\frac{(x-\mu)^2}{2} + o((x-\mu)^2).$$

Then

$$E[g(X)] = \int g(x)f(x)\,dx = g(\mu) + 0 + g''(\mu)\frac{\sigma^2}{2} + o(\sigma^2),$$

where the term $o(\sigma^2)$ approaches zero faster than $\sigma^2$ as $\sigma \to 0$. Thus, if $\sigma^2$ becomes very small, we can ignore the last term. Recall the following properties of Dirac's delta function

$$\int_{-\infty}^{\infty} \delta(x-a)g(x)\,dx = g(a), \tag{24}$$

$$\int_{-\infty}^{\infty} \delta^{(k)}(x-a)g(x)\,dx = (-1)^k g^{(k)}(a), \tag{25}$$

where (25) can be derived from (24) by applying integration by parts $k$ times. Thus, for very small $\sigma^2$, we can write

$$f(x) = \delta(x-\mu) + \frac{\sigma^2}{2}\delta^{(2)}(x-\mu) + o(\sigma^2).$$

**Note:** If the support of $f(x)$ is $[\mu-\epsilon, \mu+\epsilon]$ with very small $\epsilon$, the above condition $\sigma^2 \approx 0$ is satisfied. This condition of finite support is sufficient, but not necessary for the above formula to hold. If the distribution is Gaussian, the condition of finite support is, strictly speaking, not warranted.

(b) When a random process $X(t)$ is time-continuous as in a diffusion process, the value of $X(t+h) = x$ cannot be much different from $X(t) = x'$, because $x \to x'$ as $h \to 0$. Since we are given the drift rate and variance rate, we can write the conditional mean and conditional variance of $X(t+h)$ as follows:

$$E[X(t+h)|X(t)=x'] = x' + E[X(t+h) - X(t)|X(t)=x'] = x' = \beta(x',t)h + o(h),$$

and

$$\mathrm{Var}[X(t+h)|X(t)=x'] = E[(X(t+h)-X(t))^2|X(t)=x'] = \alpha(x',t) + o(h).$$

Clearly as $h \to 0$, the conditional PDF $f(x, t_h|x', t)$ satisfies the property of $f(x)$ having very small $\sigma^2$. By identifying $\mu$ as $x' + \beta(x',t)h$ and $\sigma^2$ as $\alpha(x',t)h$, we can write the conditional (or transitional) PDF as

$$f(x, t+h|x', t) = \delta(x - x' - \beta(x',t)h) + \delta^{(2)}(x - x' - \beta(x',t)h)\frac{\alpha(x',t)h}{2} + o(h).$$

**17.11* Derivation of the forward diffusion equation.** We start with the Chapman-Kolmogorov equation:

$$f(x, t+h|x_0, t_0) = \int f(x, t+h|x', t)f(x', t|x_0, t_0)\,dx'. \tag{26}$$

Then

$$\text{LHS} = f(x,t|x_0,t_0) + \frac{\partial f(x,t|x_0,t_0)}{\partial t} + o(h).$$

The conditional PDF $f(x, t+h||x', t)$ is a Gaussian PDF with mean $\mu(t+h|x',t) = x' + \beta(x',t)h$ and variance $\sigma^2(t+h|x',t)$, using the argument similar to the one in the derivation of the backward equation. Thus, for sufficiently small $h$, we can use the same approximation as (17.104):

$$f(y,t+h|x',t) = \delta(x-x'-\beta(x',t)h) + \delta^{(2)}(x-x'-\beta(x',t)h)\frac{\alpha(x',t)h}{2} + o(h). \quad (27)$$

Then the RHS of (26) is

$$\begin{aligned}\text{RHS} &= \int \left[\delta(x-x'-\beta(x',t)h) + \delta^{(2)}(x-x'-\beta(x',t)h)\frac{\alpha(x',t)h}{2}\right] f(x',t|x_0,t_0)\,dx' \\ &\triangleq I_1 + I_2 + o(h),\end{aligned}$$

where

$$\begin{aligned}I_1 &= \int \delta(x-x'-\beta(x',t)h) f(x',t|x_0,t_0)dx' \\ &= \int [\delta(x-x') - h\delta^{(1)}(x-x')\beta(x',t)] f(x',t|x_0,t_0)\,dx' + o(h) \\ &= f(x,t|x_0,t_0) - h\int \delta^{(1)}(x-x')[\beta(x',t)f(x',t|x_0,t_0)]\,dx' + o(h) \\ &= f(x,t|x_0,t_0) - h\frac{\partial(\beta(x,t)f(x,t|x_0,t_0))}{\partial x} + o(h), \end{aligned} \quad (28)$$

where we used the properties

$$\delta^{(1)}(x-x') = -\delta^{(1)}(x'-x), \ \text{ and } \ \int \delta^{(1)}(x'-x)(x')\,dx' = -f'(x).$$

Similarly

$$\begin{aligned}I_2 &= \frac{h}{2}\int \delta^{(2)}(x-x'-\beta(x',t)h)\alpha(x',t)f(x',t|x_0,t_0)\,dx' \\ &= \frac{h}{2}\int [\delta^{(2)}(x-x')\alpha(x',t) + o(h)] f(x',t|x_0,t_0)\,dx' \\ &= \frac{h}{2}\frac{\partial^2(\alpha(x,t)f(f,t|x_0,t_0))}{\partial x^2} + o(h).\end{aligned}$$

From these Kolmogorov's forward equation readily follows.

**Note:** The term $I_1$ can be alternatively calculated without explicit use of $\delta^{(1)}(x)$. Rewrite the argument of the delta function in the RHS of the first line of (28):

$$\begin{aligned}x - x' - \beta(x',t)h &= x - x' - h\left[\beta(x,t) - (x-x')\frac{\partial\beta(x,t)}{\partial x}\right] + o(h) \\ &= (x-x')B(x,t,h) - h\beta(x,t) + o(h),\end{aligned}$$

where

$$B(x,t,h) = 1 + h\frac{\partial \beta(x,t)}{\partial x},$$

Then

$$\begin{aligned} I_1 &= \int \delta(B(x,t,h)(x-x') - h\beta(x,t) + o(h) f(x',t|x_0,t_0)\,dx' \\ &= \int \delta\left(B(x,t,h)(x-x') - \frac{h\beta(x,t)}{B(x,t,h)}\right) f(x',t|x_0,t_0)\,dx' + o(h). \end{aligned}$$

Then, by identifying $C = B(x,t,h)$ and $c = \beta(x,t)h$, we have

$$\begin{aligned} I_1 &= \frac{f\left(x - \frac{\beta(x,t)h + o(h)}{B(x,t,h)}, t\,|x_0,t_0\right)}{B(x,t,h)} + o(h) \\ &= f(x,t|x_0,t_0)\left(1 - h\frac{\partial \beta(x,t)}{\partial x}\right) - \frac{\partial f(x,t|x_0,t_0)}{\partial x}\frac{\beta(x,t)h + o(h)}{B^2(x,t,h)} + o(h) \\ &= f(x,t|x_0,t_0) - hf(x,t|x_0,t_0)\frac{\partial \beta(x,t)}{\partial x} - h\beta(x,t)\frac{\partial f(x,t|x_0,t_0)}{\partial x} + o(h) \\ &= f(x,t|x_0,t_0) - h\frac{\partial(\beta(x,t)f(x,t|x_0,t_0))}{\partial x} + o(h), \end{aligned}$$

where we used

$$B^{-1}(x,t,h) = 1 - h\frac{\partial \beta(x,t)}{\partial x} + o(h), \text{ and } B^{-2}(x,t,h) = 1 + o(h).$$

The last expression agrees with the result of (28) obtained using the property of $\delta^{(1)}(x)$.

**17.12 Conditional expectation and pure prediction.** The discussion given in Example 22.3 (pp. 667-670) concerns about the discrete-time Gauss-Markov process (GMP), but essentially the same observation holds for the continuous-time GMP, i.e., the Ornstein-Uhlenbeck process. Theorem 22.4 essentially says that the minimum mean square prediction for a Gaussian process is given by an optimum linear predictor and is equal to the conditional estimate of $V(t)$ given the current value $V(t_0)$ $(t \geq t_0)$.

$S_{t+p}$ and $S_t$ in (22.126) and (22.127) correspond to $V(t)$ and $V(t_0)$ in (17.119) and (17.120). The correlation coefficient $\rho$ between the bivariate Gaussian variables is a geometric decay function $\rho = \alpha^p$ in the discrete-time case, where as it is an exponential decay function $\rho = e^{-\beta(t-t_0)}$. Thus, it should be apparent that the continuous-time analog of (22.126) and (22.127) takes the form of (17.119) and (17.120). See Problem 22.19 and its solution for related discussion on the continuous-time GMP.

**17.13 1st-order PDF of the O-U process.**

**Note:**Equation (17.126) has a typo error. In the RHS: $-\frac{\alpha_0}{2}$) should be $+\frac{\alpha_0}{2}$).

Since we are interested in the steady state distribution, we set the LHS to be zero. While in (17.126) $f_V = f_V(v,t)$, it is now a function of $v$ only, i.e., $f_V = f_V(v) = \lim_{t\to\infty} f_V(v,t)$:

$$\beta_1\frac{vf_V}{dv} + \frac{\alpha_0}{2}\frac{d^2 f_V}{dv^2} = 0.$$

We simplify the notation by writing $f_V$ simply as $f$, and

$$\frac{\alpha_0}{2\beta_1} \triangleq \sigma^2.$$

Then the above equation can be written as

$$\frac{d}{dv}\left[\sigma^2 \frac{df}{dv} + vf\right] = 0.$$

By integrating the above once, we obtain

$$\sigma^2 \frac{df}{dv} + vf = c_1,$$

where $c_1$ is some constant. Since any physically reasonable PDF $f$ would have both $f \to 0$ and $\frac{df}{dv} \to 0$ as $v \to \infty$. Thus $c_1$ must be zero. Thus, we reduce the problem to a first order differential equation:

$$\sigma^2 \frac{d}{dv} \ln f + v = 0,$$

which leads to

$$\ln f = \frac{v^2}{2\sigma^2} + c_2,$$

which implies

$$f = c_3 e^{-\frac{v^2}{2\sigma^2}}.$$

Since $f(v)$ is a PDF, we know the constant $c_3 = \frac{1}{\sqrt{2\pi}\sigma}$ (see p. 81).

**17.14 Time-dependent solution for the Ornstein-Uhlenbeck process.** [315]
**Note:**Equation (17.126) has a typo error. In the RHS: $-\frac{\alpha_0}{2}$) should be $+\frac{\alpha_0}{2}$).

**Note:**

1. As discussed in the solution of 17.12, the conditional mean and conditional variance (17.129) and (17.130) are given from the results (22.126) and (22.127) of the autoregressive process of order 1, AR(1): The Ornstein-Uhlenbeck process is the continuous-time analog of AR(1). Once the conditional mean and variance are given by (17.129) and (17.130), the time-dependent solution is given by (17.128)
2. The PDE is sometimes called Rayleigh equation and the solution (17.128) is known. The derivation is TBD.
3. Volume 2 of Feller derives (17.128) by making some transformation of the variables.

Here we follow Sweet & Hardin [315]. This solution technique, however, can handle more complicated cases where reflecting and or absorbing barriers exist. Our case can be considered as a special case where the barriers are at infinity. The final solution involves an infinite set of eigenvalues and eigenfuctions and do not lead to (17.128).
In order to use the notation and symbols commonly used in the theory of differential equation, we use the variables $x$ and $t$ instead of $v$ and $t$. So we write the PDF as $f(x,t)$. We also write

$\beta_1 = \beta$ and $\alpha_0 = \alpha$. The forward equation of (17.126) is

$$\frac{\partial f(x,t)}{\partial t} = \beta \frac{\partial [xf(x,t)]}{\partial x} + \frac{\alpha}{2}\frac{\partial^2 f(x,t)}{\partial x^2}, \ \ \alpha > 0, \beta > 0. \tag{29}$$

We follow Sweet & Hardin [315] and use the "separation of variables" method, a common technique in solving PDEs. Writing

$$f(x,t) = X(x)T(t),$$

we have

$$X(x)\dot{T}(t) = \beta(X(x) + xX'(t))T(t) + \frac{\alpha}{2}X''(x)T(t),$$

which we rewrite as

$$\frac{\dot{T}(t)}{T(t)} = \beta + \frac{xX(x)}{X(x)} + \frac{\alpha}{2}\frac{X''(x)}{X(x)} = \text{const},$$

where the constant must be a negative value, otherwise there is no stable solution for $T(t)$. By writing this constant as $-c^2$, we have

$$\dot{T}(t) + a^2 T(t) = 0 \tag{30}$$

$$\frac{\alpha}{2}X''(x) + \beta x X'(x) + (\beta + a^2)X(x) = 0. \tag{31}$$

Replace the function $X(x)$ by $Y(x)$ such that

$$X(x) = e^{-\frac{\beta x^2}{2\alpha}}Y(x), \tag{32}$$

Then change the variable $x$ to $s$ such that

$$x = s\sqrt{\frac{\alpha}{2\beta}}.$$

Then, we have

$$\frac{d^2Y(s)}{ds^2} + \left(\frac{1}{2} + \lambda - \frac{s^2}{4}\right)Y(s) = 0, \tag{33}$$

where

$$\lambda = \frac{a^2}{\beta}.$$

Equation (33) is known as Weber's differential equation.
As solutions to the above differential equation, we consider the even functions

$$y_e(s;\lambda) = \exp\left\{-\frac{s^2}{4}\right\} M\left(-\frac{\lambda}{2}, \frac{1}{2}, \frac{s^2}{2}\right),$$

and the odd function

$$y_o(s;\lambda) = \exp\left\{-\frac{s^2}{4}\right\} M\left(-\frac{\lambda+1}{2}, \frac{3}{2}, \frac{s^2}{2}\right),$$

where $M(a,b,c)$ is Kummer's function. The Wronskian of these solutions is

$$W(y_e, y_o; \lambda) = \det \begin{bmatrix} y_e(s;\lambda) & y_o(s;\lambda) \\ y_e'(s;\lambda) & y_o'(s;\lambda) \end{bmatrix} = 1.$$

Furthermore, there are the following relations:

$$y - e'(s;\lambda) + \frac{1}{2} s y_e(s;\lambda) = -\lambda y_0(s;\lambda - 1), \tag{34}$$

$$y_o'(s;\lambda) + \frac{1}{2} s y_o(s;\lambda) = y_e(s;\lambda). \tag{35}$$

Thus, we seek solutions of the form

$$Y(s) = A y_e(s;\lambda) + B y_o(s;\lambda).$$

**To continue:**

**17.15 Variance of the integration of the O-U process.**

**Note:** (17.132) is not correct as shown in (38).

We represent the O-U process V(t) by the stochastic differential equations (17.112) and (17.122), i.e.,

$$\begin{aligned} dV(t) &= -\beta_1 V(t)\, dt + \sqrt{\alpha_0} Z(t)\, dt \\ &= -\beta_1 V(t)\, dt + \sqrt{\alpha_0}\, dW(t), \end{aligned} \tag{36}$$

where $W(t)$ is the standard Brownian motion, and $Z(t)$ is white Gaussian noise with zero mean and unit variance. Observing that

$$d(e^{\beta_1 s} V(s)) = \sqrt{\alpha_0} e^{\beta_1 s} dW(s),$$

and integrating it from $s = 0$ to $s = u$, we obtain the stochastic integral representation

$$V(u) = v_0 e^{-\beta_1 u} + \sqrt{\alpha_0} \int_0^u e^{-\beta_1 (u-s)} dW(s), \tag{37}$$

where $v_0 = V(0)$. From this we find that the conditional mean and variance of $V(t)$ are given by (17.129) and (17.130) of p. 505.

The integrated process $X(t)$ of (17.131) is also a Gaussian process. Substituting (37) into (17.131), we have

$$X(t) = \int_0^t V(u)\, du = v_0 \int_0^t e^{-\beta_1 u}\, du + \sqrt{\alpha_0} \int_0^t \int_0^u e^{-\beta_1 (u-s)} dW(s)\, du.$$

Now we change the order of the integration in the second term, i.e.,

$$\begin{aligned} &\int_0^t \left( \int_0^u e^{-\beta_1 (u-s)} dW(s) \right) du = \int_0^t e^{\beta_1 s} \left( \int_s^t e^{-\beta_1 u}\, du \right) dW(s) \\ &= \frac{1}{\beta_1} \int_0^t \left( 1 - e^{-\beta_1 (t-s)} \right) dW(s). \end{aligned}$$

Hence, we obtain

$$X(t) = \frac{v_0}{\beta_1}\left(1 - e^{-\beta_1 t}\right) + \frac{\sqrt{\alpha_0}}{\beta_1}\int_0^t \left(1 - e^{-\beta_1(t-s)}\right) dW(s).$$

Clearly $X(0) = 0$, and its mean

$$E[X(t)] = \frac{E[V(0)]}{\beta_1}\left(1 - e^{-\beta_1 t}\right) = 0,$$

and variance

$$\begin{aligned}
\mathrm{Var}[X(t)] &= \frac{\alpha_0}{\beta_1^2}\int_0^t\int_0^t (1 - e^{-\beta_1(t-s)})(1 - e^{-\beta_1(t-s')})E[Z(s)Z(s')]\,ds\,ds' \\
&= \frac{\alpha_0}{\beta_1^2}\int_0^t\int_0^t (1 - e^{-\beta_1(t-s)})(1 - e^{-\beta_1(t-s')})\delta(s-s')\,ds\,ds' \\
&= \frac{\alpha_0}{\beta_1^2}\int_0^t (1 - e^{-\beta_1(t-s)})^2\,ds = \frac{\alpha_0}{\beta_1^2}\int_0^t \left[1 - 2e^{-\beta_1(t-s)} + e^{-2\beta_1(t-s)}\right] ds \\
&= \frac{\alpha_0}{\beta_1^2}\left[t - \frac{2}{\beta_1}(1 - e^{-\beta_1 t}) + \frac{1}{2\beta_1}(1 - e^{-2\beta_1 t})\right] \\
&= \frac{2\sigma_V^2}{\beta_1}\left[t - \frac{2}{\beta_1}(1 - e^{-\beta_1 t}) + \frac{1}{2\beta_1}(1 - e^{-2\beta_1 t})\right] \qquad (38)
\end{aligned}$$

(17.132) is incorrect since the third term in [ ] is missing.

## 17.3 Stochastic Differential Equations and Itô Process

**17.16* Conditional mean and variance of the geometric Brownian motion.**
We can write

$$\begin{aligned}
E[Y(t)|Y(u), 0 \le u \le s] &= E\left[e^{X(t)}|X(u), 0 \le u \le s\right] \\
&= E\left[e^{X(s)+X(t)-X(s)}|X(u), 0 \le u \le s\right] \\
&= e^{X(s)}E\left[e^{X(t)-X(s)}|X(u), 0 \le u \le s\right] \quad \text{(independent increments)} \\
&= Y(s)E\left[e^{X(t-s)-X(0)}\right] \quad \text{(temporal homogeneity)} \\
&= Y(s)E\left[e^{X(t-s)}\right] \quad (\text{because } X(0) = 0)
\end{aligned}$$

Recall the moment generating function (MGF) of a normal RV $X$:

$$M_X(\xi) = E[e^{\xi X}] = \exp\left\{E[X]\xi + \frac{\mathrm{Var}[X]\xi^2}{2}\right\}.$$

Thus,

$$E\left[e^{X(t-s)}\right] = M_{X(t-s)}(1) = \exp\left\{\beta(t-s) + \frac{\alpha(t-s)}{2}\right\} = e^{\left(\beta+\frac{\alpha}{2}\right)(t-s)}.$$

Therefore,

$$E[Y(t)|Y(u), 0 \le u \le s] = Y(s)e^{\left(\beta+\frac{\alpha}{2}\right)(t-s)}.$$

Similarly,

$$\begin{aligned}
E[Y(t)^2|Y(u), 0 \le u \le s] &= E\left[e^{2X(t)}|X(u), 0 \le u \le s\right] \\
&= E\left[e^{2X(s)+2(X(t)-X(s))}|X(u), 0 \le u \le s\right] \\
&= e^{2X(s)}E\left[e^{2(X(t)-X(s))}|X(u), 0 \le u \le s\right] \\
&= Y(s)^2E\left[e^{2(X(t-s)-X(0))}\right] = Y(s)^2E\left[e^{2X(t-s)}\right] \\
&= Y(s)^2e^{2(\beta+\alpha)(t-s)}
\end{aligned}$$

Thus, the conditional variance is

$$\begin{aligned}
\mathrm{Var}[Y(t)|Y(u), 0 \le u \le s] &= E[Y(t)^2|Y(u), 0 \le u \le s] - (E[Y(t)|Y(u), 0 \le u \le s])^2 \\
&= Y(s)^2e^{2(\beta+\alpha)(t-s)} - \left(Y(s)e^{\left(\beta+\frac{\alpha}{2}\right)(t-s)}\right)^2 \\
&= Y(s)^2e^{(2\beta+\alpha)(t-s)}\left(e^{\alpha(t-s)} - 1\right).
\end{aligned}$$

**17.17 Itô's lemma applied to geometric Brownian motion.** We apply the Taylor series expansion to the function $V = V(Y, t)$, similar to (17.148) :

$$dV = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial Y}\,dY + \frac{1}{2}\frac{\partial^2 V}{\partial Y^2}dY^2 + \dots.$$

For the GBM process $Y(t) = e^{X(t)}$,

$$dY = e^X\,dX = Y\,dX = Y(\beta dt + \sqrt{\alpha}\,dW_{\mathrm{s}}),$$

where $W_{\mathrm{s}}(t)$ is the standard Brownian motion, i.e., $W_{\mathrm{s}}(t) \sim N(0, t)$. Similarly

$$dY^2 = Y^2(\beta^2\,dt^2 + \alpha dW_{\mathrm{s}}^2 + 2\beta\sqrt{\alpha}\,dW_{\mathrm{s}}\,dt).$$

From Itô's lemma,

$$dW_{\mathrm{s}}^2 = dt, dW_{\mathrm{s}}\,dt = 0,$$

Hence,

$$\begin{aligned}
dV &= \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial Y}Y(\beta dt + \sqrt{\alpha}\,dW_{\mathrm{s}}) + \frac{1}{2}\frac{\partial^2 V}{\partial Y^2}Y^2\alpha dW_{\mathrm{s}}^2 + o(dt) \\
&= \left(\frac{\partial V}{\partial t} + \beta Y\frac{\partial V}{\partial Y} + \frac{\alpha Y^2}{2}\frac{\partial^2 V}{\partial Y^2}\right) + \sqrt{\alpha}\frac{\partial V}{\partial Y}\,dW_{\mathrm{s}},
\end{aligned}$$

where we used $dW_{\mathrm{s}}^2 = dt$ as given in (17.144). Thus, we have obtained (17.173).

**17.18 Conditional mean and variance of $X(T)$, given $Y(t)$.** Show that the conditional mean and variance of $X(T)$, given $Y(t)$ are given by (17.184).
Brownian motion $X(t)$ is normally distributed with mean and variance given by

$$E[X(t)|X(0) = x_0] = x_0 + \beta t, \text{ and } \mathrm{Var}[X(t)|X(0) = x_0] = \alpha t,$$

where $\beta = \beta_y - \alpha/2$ in the GBM $Y(t)$ of (17.172) should be replaced by $\beta = r - \alpha/2$ for the model (17.172). Using the temporal homogeneity, we have

$$E[X(T)|X(t) = x(t)] = x(t) + \beta(T-t), \text{ and } \mathrm{Var}[X(T)|X(t) = x(t)] = \alpha(T-t).$$

Since $x(t) = \ln y(t)$ we have

$$E[X(T)|Y(t) = y(t)] = \ln y(t) + \left(r - \frac{\alpha}{2}\right)(T-t), \text{ and } \mathrm{Var}[X(T)|Y(t) = y(t)] = \alpha(T-t).$$

**17.19* European call option.**

(a) The call option price is \$13.50.
(b) The call option price is \$17.03.
(c) The call option price is \$19.99. A MATLAB program is as follows:

```
 function option
%
%  Example in Chapter 16: European call option
%
Yt=100; C=90; Tt=0.5; sigma=0.2; r=0.1;
%
alpha=sigma^2; t1=log(Yt/C); t2=sqrt(alpha*Tt); t3=(r+alpha/2)*Tt;
t4=(r-alpha/2)*Tt; u1=(t1+t3)/t2; u2=(t1+t4)/t2; Phi1=normcdf(u1);
Phi2=normcdf(u2); v=Yt*Phi1-C*exp(-r*Tt)*Phi2;
fprintf('Current price= %5.2f \n', Yt);
fprintf('Exercise price= %5.2f \n', C);
fprintf('Expiration date (in month)= %5.2f \n', Tt*12);
fprintf('Volatility= %5.2f \n', sqrt(alpha));
fprintf('Risk-free interest rate= %5.2f \n', r);
fprintf('The value of the call option= %5.2f \n',v);
```

# 18 Solutions for Chapter 18: Statistical Estimation and Decision Theory

## 18.1 Parameter Estimation

**18.1 Sampling distribution**

(a)

$$F_T(t) = P[T(\boldsymbol{X}) \le t] = \int_{-\infty}^{\infty} I(T(\boldsymbol{x}) \le t) f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x}.$$

(b)

$$b_T = E[T(\boldsymbol{x})] - \theta = \int_{-\infty}^{\infty} [T(\boldsymbol{x}) - \theta] f_{\boldsymbol{X}}(\boldsymbol{x}; \theta)\, d\boldsymbol{\theta}.$$

(c)

$$\sigma_T^2 = \mathrm{Var}[T(\boldsymbol{x})] = E\left[(T(\boldsymbol{X}) - \theta - b_T)^2\right] = \int_{-\infty}^{\infty} (T(\boldsymbol{x}) - \theta - b_T)^2 f_{\boldsymbol{X}}(\boldsymbol{x}; \theta)\, d\boldsymbol{x}.$$

**18.2 Efficient estimator** By definition $\mathbf{Var}[\hat{\boldsymbol{\theta}}^*(\boldsymbol{X})] - \mathbf{Var}[\hat{\boldsymbol{\theta}}(\boldsymbol{X})]$ is a negative semi-definite, i.e.,

$$\boldsymbol{a}^\top \left(\mathbf{Var}[\hat{\boldsymbol{\theta}}^*(\boldsymbol{X})] - \mathbf{Var}[\hat{\boldsymbol{\theta}}(\boldsymbol{X})]\right) \boldsymbol{a} \le 0,$$

for any non-zero vector $\boldsymbol{a}$. Since the estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ is unbiased,

$$\mathbf{Var}[\hat{\boldsymbol{\theta}}(\boldsymbol{X})] = E[\hat{\boldsymbol{\theta}}(\boldsymbol{X})\hat{\boldsymbol{\theta}}(\boldsymbol{X})^\top].$$

Thus,

$$\boldsymbol{a}^\top \mathbf{Var}[\hat{\boldsymbol{\theta}}(\boldsymbol{X})]\boldsymbol{a} = E[(\boldsymbol{a}^\top \hat{\boldsymbol{\theta}}(\boldsymbol{X}))\hat{\boldsymbol{\theta}}(\boldsymbol{X})^\top \boldsymbol{a}] = E[(\boldsymbol{a}^\top \hat{\boldsymbol{\theta}}(\boldsymbol{X}))^2] = \mathrm{Var}\,[\boldsymbol{a}^\top \hat{\boldsymbol{\theta}}(\boldsymbol{X})].$$

The same expression holds for $\hat{\boldsymbol{\theta}}^*(\boldsymbol{X})$. Hence (**??**) follows.

**18.3 More on Exponential family distributions.**

(a) We have

$$f(x; \sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) = x \exp\left(x^2\eta + \log(-2\eta)\right)$$

where $\eta = -\frac{1}{2}\sigma^{-2}$. Hence $T(x) = x^2$ and $A(\eta) = -\log(-2\eta)$. We find $E[x^2] = A'(\eta) = -\eta^{-1} = 2\sigma^2$ and $Var[x^2] = A''(\eta) = \eta^{-2} = 4\sigma^4$.

(b) For the MLE of the exponential family, we have

$$\nabla_{\boldsymbol{\theta}} \log L_{\boldsymbol{x}}(\theta) = \nabla_{\boldsymbol{\theta}} \log h(x) + \boldsymbol{\eta}^\top(\boldsymbol{\theta})\boldsymbol{T}(\boldsymbol{x}) - \nabla_{\boldsymbol{\theta}} A(\boldsymbol{\theta}) = 0.$$

**18.4* Properties of the score function and the observed Fisher information matrix.**

(a) We assume that the regularity conditions for the validity of the following transformations are satisfied. Taking the gradient with respect to $\boldsymbol{\theta}$ of $E[\boldsymbol{T}^\top(\boldsymbol{X},\boldsymbol{\theta})] = \int_x f(\boldsymbol{x},\boldsymbol{\theta})\boldsymbol{T}^\top(\boldsymbol{x},\boldsymbol{\theta})d\boldsymbol{x}$ and using the formula for the gradient of a product (see Supplementary Materials), we obtain

$$\nabla_{\boldsymbol{\theta}} E[\boldsymbol{T}^\top(\boldsymbol{X},\boldsymbol{\theta})] = \int_x f(\boldsymbol{x},\boldsymbol{\theta})(\nabla_{\boldsymbol{\theta}}\boldsymbol{T}^\top(\boldsymbol{x},\boldsymbol{\theta}))d\boldsymbol{x} + \int_x (\nabla_{\boldsymbol{\theta}} f(\boldsymbol{x},\boldsymbol{\theta}))\boldsymbol{T}^\top(\boldsymbol{x},\boldsymbol{\theta})d\boldsymbol{x} \tag{1}$$

But, according to definition of the score,

$$\boldsymbol{s}(\boldsymbol{x},\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \log f(\boldsymbol{x},\boldsymbol{\theta}) = \frac{\nabla_{\boldsymbol{\theta}} f(\boldsymbol{x},\boldsymbol{\theta})}{f(\boldsymbol{x},\boldsymbol{\theta})}$$

so that the previous equation can be written as

$$\nabla_{\boldsymbol{\theta}} E[\boldsymbol{T}^\top(\boldsymbol{X},\boldsymbol{\theta})] = E[\nabla_{\boldsymbol{\theta}}\boldsymbol{T}^\top(\boldsymbol{X},\boldsymbol{\theta})] + E[s(\boldsymbol{X},\boldsymbol{\theta})\boldsymbol{T}^\top(\boldsymbol{X},\boldsymbol{\theta})] \tag{2}$$

which is equivalent to 18.111.

(b) Equation (18.111) with $\boldsymbol{T} = 1$ yields

$$E[\boldsymbol{s}(\boldsymbol{X};\boldsymbol{\theta})] = 0.$$

Alternatively, we can derive the formula directly, by expressing the expectation in terms of the PDF. Again we show the single parameter case:

$$\text{LHS} = \int \frac{\log f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{\partial\theta} f_{\boldsymbol{X}}(\boldsymbol{x};\theta)\,d\boldsymbol{x} = \int \frac{\partial f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{\partial\theta}\,d\boldsymbol{x} = \frac{\partial}{\partial\theta}\int f_{\boldsymbol{X}}(\boldsymbol{x};\theta)\,d\boldsymbol{x} = \frac{\partial 1}{\partial\theta} = 0.$$

(c) Substitute $\boldsymbol{T}(\boldsymbol{X};\boldsymbol{\theta}) = \boldsymbol{s}(\boldsymbol{X};\boldsymbol{\theta})$ in (18.111). Since $E[\boldsymbol{s}(\boldsymbol{X};\boldsymbol{\theta})] = \mathbf{0}$ and $\nabla_{\boldsymbol{\theta}}\boldsymbol{s}^\top(\boldsymbol{x};\boldsymbol{\theta}) = \boldsymbol{J}(\boldsymbol{x};\boldsymbol{\theta})$ according to (18.32), (18.111) becomes (18.113).

When $\boldsymbol{\theta}$ is a one-dimensional parameter, the LHS of (18.113) is $E[s(\boldsymbol{x};\theta)^2] = E\left[\left(\frac{\log f(\boldsymbol{x};\theta)}{\partial\theta}\right)^2\right]$, and the RHS is $\mathcal{I}(\theta) = -E\left[\frac{\partial^2 \log L_{\boldsymbol{X}}(\theta)}{\partial\theta^2}\right]$.

(d) Denote as $\boldsymbol{T}(\boldsymbol{x},\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$. Since $\hat{\boldsymbol{\theta}}$ is unbiased, $E[\boldsymbol{T}^\top(\boldsymbol{x},\boldsymbol{\theta})] = 0$. Also $\nabla_{\boldsymbol{\theta}}\boldsymbol{T}^\top(\boldsymbol{x},\boldsymbol{\theta}) = -\nabla_{\boldsymbol{\theta}}\boldsymbol{\theta}^\top = -\boldsymbol{I}$. Thus, equation (18.111) can be written as

$$E[\boldsymbol{s}(\boldsymbol{X},\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top] = \mathrm{Cov}[\boldsymbol{s}(\boldsymbol{X},\boldsymbol{\theta}),\hat{\boldsymbol{\theta}}] = \boldsymbol{I}.$$

Since

$$\mathrm{Cov}[\hat{\boldsymbol{\theta}},\boldsymbol{s}(\boldsymbol{X},\boldsymbol{\theta})] = (\mathrm{Cov}[\boldsymbol{s}(\boldsymbol{X},\boldsymbol{\theta}),\hat{\boldsymbol{\theta}}])^\top$$

we conclude that

$$\mathrm{Cov}[\boldsymbol{s}(\boldsymbol{X},\boldsymbol{\theta}),\hat{\boldsymbol{\theta}}] = \mathrm{Cov}[\hat{\boldsymbol{\theta}},\boldsymbol{s}(\boldsymbol{X},\boldsymbol{\theta})] = \boldsymbol{I}. \tag{3}$$

**An alternative proof:** Since $\nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})$ has zero mean, it suffices to show

$$\int f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) \left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}\right) dx = \mathbf{0}.$$

The unbiasedness of $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ gives,

$$\int f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta})\, d\boldsymbol{x} = 0.$$

By applying $\nabla_{\boldsymbol{\theta}}$ to the above, and using the formula $\nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) = \frac{\nabla_{\boldsymbol{\theta}} f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}$, the required result readily follows.

**18.5 Maximum-likelihood estimate of $g(\boldsymbol{\theta})$.**
Consider the single parameter case, since the multiple parameter case can be treated in the same way by applying the transformation to the individual parameter. The likelihood functions of the original parameter $\theta$ and the transformed parameter $\eta = g(\theta)$ is given by

$$\begin{aligned} L_{\boldsymbol{x}}(\theta) &= f_{\boldsymbol{X}}(\boldsymbol{x};\theta) \\ L_{\boldsymbol{x}}(\eta) &= f_{\boldsymbol{X}}(\boldsymbol{x};g^{-1}(\eta)) \triangleq f_{\boldsymbol{X}}(\boldsymbol{x};h(\eta)), \end{aligned}$$

where the $h$ is the inverse function of $g$. The MLE $\hat{\eta}$ must satisfy

$$\frac{dL_{\boldsymbol{x}}(\eta)}{d\eta} = 0 \text{ at } \eta = \hat{\eta}.$$

From the chain rule, we have

$$\begin{aligned} \frac{dL_{\boldsymbol{x}}(\eta)}{d\eta} = \frac{\partial f_{\boldsymbol{X}}(\boldsymbol{x};h(\eta))}{\partial \eta} &= \frac{\partial f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{\partial \theta}\frac{dh(\eta)}{d\eta} \\ &= \frac{dL_{\boldsymbol{x}}(\theta)}{d\theta}\frac{dh(\eta)}{d\eta}. \end{aligned}$$

Since $g$ is a monotone function, so is $h$. Thus $\frac{dh(\eta)}{d\eta} \neq 0$ for any $\eta$. Thus, $\frac{dL_{\boldsymbol{x}}(\eta)}{d\eta} = 0$ for $\eta = \eta^{\star}$, if and only if $\frac{dL_{\boldsymbol{x}}(\theta)}{d\theta)} = 0$ for $\theta = \theta^{\star}$, Since the MLE $\hat{\eta}$ is one of these solutions $\eta^{\star}$, the MLE estimate satisfies the property (18.117).

**An alternative proof:** Suppose that $\hat{\boldsymbol{\theta}}$ is an MLE of $L_{\boldsymbol{x}}(\boldsymbol{\theta}) = f(\boldsymbol{x};g(\boldsymbol{\theta}))$. Then

$$f(\boldsymbol{x};g(\hat{\boldsymbol{\theta}})) \geq f(\boldsymbol{x};g(\boldsymbol{\theta})), \text{ for all } \boldsymbol{\theta} \in V_{\hat{\boldsymbol{\theta}}},$$

where $V_{\hat{\boldsymbol{\theta}}}$ is a vicinity of $\hat{\boldsymbol{\theta}}$. Since $g(\cdot)$ is continuous and monotonic, $V_{\hat{\boldsymbol{\theta}}}$ is mapped into $U_{\hat{\boldsymbol{\eta}}}$, a vicinity of $\hat{\boldsymbol{\eta}} = g(\hat{\boldsymbol{\theta}})$, and

$$f(\boldsymbol{x};\hat{\boldsymbol{\eta}}) \geq f(\boldsymbol{x};\boldsymbol{\eta}), \text{ for all } \boldsymbol{\eta} \in U_{\hat{\boldsymbol{\eta}}}.$$

Thus, we have

$$\hat{\boldsymbol{\eta}} = g(\hat{\boldsymbol{\theta}})$$

as an MLE of the transformed parameter $\boldsymbol{\eta}$.

**18.6 The CRLB and the Cauchy-Schwarz inequality.**
The definition of the bias of the estimator

$$b(\theta) = E[\hat{\theta}(\boldsymbol{X})] - \theta$$

can be expressed, for a continuous RV $\boldsymbol{X}$, as

$$\theta + b(\theta) = \int_{\boldsymbol{x}} T(\boldsymbol{x}) f(\boldsymbol{x}, \theta)\, d\boldsymbol{x}. \tag{4}$$

By differentiating w.r.t. $\theta$, and assuming a regularity condition that allows exchange the order of differentiation and integration, we have

$$\begin{aligned} 1 + b'(\theta) &= \int_{\boldsymbol{x}} T(\boldsymbol{x}) \frac{\partial f(\boldsymbol{x}, \theta)}{\partial \theta}\, d\boldsymbol{x} = \int_{\boldsymbol{x}} T(\boldsymbol{x}) \left( \frac{\partial \log f(\boldsymbol{x}, \theta)}{\partial \theta} \right) f(\boldsymbol{x}, \theta)\, d\boldsymbol{x} \\ &= \mathrm{Cov}[T(\boldsymbol{x}), s(\boldsymbol{X}, \theta)], \end{aligned}$$

where in deriving the last equation, we used the property $E[s(\boldsymbol{X}, \theta)] = 0$, i.e., $E[(\theta + b(\theta))s(\boldsymbol{X}, \theta)] = 0$, independent of $\theta$.
Recall the inequality for a pair of RVs $X, Y$, given by (10.12):

$$|\mathrm{Cov}[X, Y]|^2 \leq \mathrm{Var}[X]\mathrm{Var}[Y]. \tag{5}$$

Let $X$ and $Y$ be replaced by the RVs $T(X)$ and $s(\boldsymbol{X}, \theta)$. Then, using the property of $s(\boldsymbol{X}, \theta)$ given by (18.113), we find

$$|1 + b'(\theta)|^2 \leq \mathrm{Var}[T(\boldsymbol{X})] I(\theta),$$

from which (18.40) readily follows.

**18.7* The CRLB and a sufficient statistic.**
Apply the inverse operation of the operator $\nabla_{\boldsymbol{\theta}}$ to both sides in (18.43), leading to

$$\log f_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\theta}) = \int (\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta})^\top \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})\, d\boldsymbol{\theta} + C(\boldsymbol{x}), \tag{6}$$

where $C(\boldsymbol{x})$ is an arbitrary function of $\boldsymbol{x}$ that must satisfy the normalization condition for $L_{\boldsymbol{x}}(\boldsymbol{x}; \boldsymbol{\theta})$). The integration notation $\int \boldsymbol{a}^\top(\boldsymbol{\theta}) d\boldsymbol{\theta}$ should not be confused as the regular multiple integrations of many variables. For a vector function $\boldsymbol{a}(\boldsymbol{\theta}) = (a_1(\boldsymbol{\theta}), a_2(\boldsymbol{\theta}), \ldots, a_M(\boldsymbol{\theta}))$, we define

$$\int \boldsymbol{a}^\top\, d\boldsymbol{\theta} \triangleq \sum_{i=1}^{M} \int a_i(\boldsymbol{\theta})\, d\theta_i. \tag{7}$$

Thus,

$$L_{\boldsymbol{x}}(\boldsymbol{x}; \boldsymbol{\theta}) = h(\boldsymbol{x}) \exp\left( \int (\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta})^\top \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})\, d\boldsymbol{\theta} \right) = \exp\left( \boldsymbol{\eta}(\boldsymbol{\theta})^\top \hat{\boldsymbol{\theta}}(\boldsymbol{x}) - A(\boldsymbol{\theta}) \right), \tag{8}$$

where $h(\boldsymbol{x}) = \exp C(\boldsymbol{x})$, and

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = \int \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \text{ and } A(\boldsymbol{\theta}) = \int \boldsymbol{\theta}^\top \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{9}$$

Hence, it is apparent from Theorem 18.1 that an *efficient* estimate $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ is a *sufficient statistic* for estimating $\boldsymbol{\theta}$.

**18.8 Minimum-variance unbiased linear estimator.**

(a) Let the linear estimator be

$$T(\boldsymbol{x}) = \sum_{i=1}^{n} a_i x_i = \boldsymbol{a}^\top \boldsymbol{x}.$$

Then the condition of unbiasedness requires that

$$E[T(\boldsymbol{x})] = \theta, \text{ or } E[\boldsymbol{a}^\top \boldsymbol{x}] = \theta \boldsymbol{a}^\top \boldsymbol{\mu} = \theta, \text{ i.e., } \boldsymbol{a}^\top \boldsymbol{\mu} = 1.$$

The variance of the estimator is

$$\begin{aligned}\mathcal{E} = \mathrm{Var}[T(\boldsymbol{x})] &= E\left[\left\{\boldsymbol{a}^\top(\boldsymbol{X} - E[\boldsymbol{X}])\right\}^2\right] \\ &= \boldsymbol{a}^\top E\left[(\boldsymbol{X} - E[\boldsymbol{X}])(\boldsymbol{X} - E[\boldsymbol{X}])^\top\right] \boldsymbol{a} = \boldsymbol{a}^\top \boldsymbol{\Sigma} \boldsymbol{a}.\end{aligned}$$

So we must find $\boldsymbol{a}$ that minimizes $\mathcal{E}$ subject to the constraint $\boldsymbol{a}^\top \boldsymbol{\mu} = 1$. We use the Lagrangian multiplier method, and define

$$J(\boldsymbol{a}, \lambda) = \boldsymbol{a}^\top \boldsymbol{\Sigma} \boldsymbol{a} - \lambda(\boldsymbol{a}^\top \boldsymbol{\mu} - 1).$$

By differentiating $J$ with respect to $\boldsymbol{a}$ and $\lambda$, we find the necessary conditions for extremum:

$$\begin{aligned}\nabla_{\boldsymbol{a}} J(\boldsymbol{a}, \lambda) &= 2\boldsymbol{\Sigma}\boldsymbol{a} - \lambda\boldsymbol{\mu} = \boldsymbol{0} \\ \frac{\partial J(\boldsymbol{a}, \lambda)}{\partial \lambda} &= \boldsymbol{a}^\top \boldsymbol{\mu} - 1 = 0,\end{aligned}$$

where the second equation is nothing but the original constraint equation. From the first equation, we have

$$\boldsymbol{a} = \frac{\lambda}{2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}, \text{ i.e., } \boldsymbol{a}^\top = \frac{\lambda}{2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}.$$

Thus, the linear estimator we seek is given by

$$T(\boldsymbol{x}) = \frac{\lambda}{2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}.$$

By inserting this into the unbiasedness condition, we have

$$E[T(\boldsymbol{x})] = \frac{\lambda}{2} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \theta = \theta.$$

Hence,

$$\frac{\lambda}{2} = \frac{1}{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}.$$

Thus, we find the efficient linear estimator is given by

$$\hat{\theta}(\boldsymbol{x}) = \frac{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{x}}{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}.$$

(b)

$$\hat{\theta}(\boldsymbol{X}) - E[\hat{\theta}(\boldsymbol{X})] = \frac{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - E[\boldsymbol{X}])}{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}.$$

Hence

$$\sigma_{\hat{\theta}}^2 = \mathrm{Var}[\hat{\theta}(\boldsymbol{X})] = \frac{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} E\left[(\boldsymbol{x} - E[\boldsymbol{X}])(\boldsymbol{x} - E[\boldsymbol{X}])^\top\right] \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^2}$$
$$= \frac{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}^2} = \frac{1}{\boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}}.$$

**18.9 Confidence interval.**
Because the functions $d_i(\cdot) \;\; (i = 1, 2)$ is monotone increasing, so is its inverse function $c_i(\cdot) = d_i^{-1}(\cdot) \;\; (i = 1, 2)$. Thus, we that the following three events are equivalent to each other:

$$\{d_1(\theta) < \hat{\theta} < d_2(\theta)\} \Longleftrightarrow \{\theta < d_1^{-1}(\hat{\theta}) \text{ and } d_2^{-1}(\hat{\theta}) < \theta\} \Longleftrightarrow \{c_1(\hat{\theta}) < \theta < c_2(\hat{\theta})\}.$$

Taking the probability of the LHS and the RHS, we prove that (18.54) implies (18.57) and vice versa.

**18.10 Confidence interval of an estimator of the binomial distribution parameter.**
According to the DeMoivre-Laplace theorem, the statistic $\hat{p} = k/n$ is asymptotically normally distributed $N(p, \sqrt{pq/n})$. Therefore, as in Example 18.6, we find

$$d_1(p) = p - u_\alpha \sqrt{\frac{pq}{n}}, \;\; d_2(p) = p + u_\alpha \sqrt{\frac{pq}{n}}.$$

These equations represent lower and upper parts of the ellipse

$$(\hat{p} - p)^2 = u_\alpha^2 \frac{p(1-p)}{n} \tag{10}$$

on the $(p, \hat{p})$ plane.
By solving this quadratic equation for $p$, we obtain the inverse functions representing the desired confidence interval bounding points

$$c_1(\hat{p}) = \frac{n}{n + u_\alpha} \left( \hat{p} + \frac{u_\alpha^2}{n} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{u_\alpha^2}{4n^2}} \right)$$
$$c_2(\hat{p}) = \frac{n}{n + u_\alpha} \left( \hat{p} + \frac{u_\alpha^2}{n} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{u_\alpha^2}{4n^2}} \right)$$

We need consider only the portion of the ellipse for $0 \leq \hat{p} \leq 1$.

## 18.2 Hypothesis Testing and Statistical Decision

**18.11 Matched filter.**

(a)

$$E\left[\tilde{Z}_n^2\right] = E\left[\sum_{k=0}^{n}(h_k Z_{n-k})^2\right] = \sum_{j=0}^{n}\sum_{k=0}^{n} h_j h_k \sigma^2 \delta_{j,k}$$
$$= \sigma^2 \sum_{k=0}^{n} h_k^2 = \sigma^2 \|\boldsymbol{h}\|^2.$$

(b) Define $h_k \triangleq g_{n-k}, \quad k = 0, 1, 2, \ldots, n$. Then

$$\text{SNR} = \frac{A^2 \tilde{s}_n^2}{E\left[\tilde{Z}_n^2\right]} = \frac{A^2\left(\sum_{k=0}^{n} g_k s_k\right)^2}{\sigma^2\|\boldsymbol{h}\|^2} = \frac{A^2\left(\boldsymbol{g}^\top \boldsymbol{s}\right)^2}{\sigma^2\|\boldsymbol{h}\|^2}$$
$$\le \frac{A^2\|\boldsymbol{g}\|^2\|\boldsymbol{s}\|^2}{\sigma^2\|\boldsymbol{h}\|^2},$$

where the equality holds when $\boldsymbol{g} = c\boldsymbol{s}$, where $c$ is an arbitrary constant. In other words, $h_k = g_{n-k} = s_{n-k},\ k = 0, 1, 2, \ldots, n$. SNR becomes

$$\text{SNR} \le \frac{A^2\|bs\|^2}{\sigma^2} = \frac{A^2}{\sigma^2},$$

where we used the assumption that $\|\boldsymbol{s}\|^2 = 1$.

(c) $y_n = \sum_{k=0}^{n} h_k x_{n-k} = c\sum_{k=0}^{n} s_{n-k}x_{n-k} = c\boldsymbol{s}^\top\boldsymbol{x} = cT(\boldsymbol{x})$.

**18.12 Correlation receiver.**

(a) Since

$$R = A\boldsymbol{g}^\top\boldsymbol{s} + \boldsymbol{g}^\top\boldsymbol{Z},$$

we can write SNR as

$$\text{SNR} = \frac{A^2\left(\boldsymbol{g}^\top\boldsymbol{s}\right)^2}{\sigma^2\|\boldsymbol{g}\|^2}.$$

(b) Using the Cauchy-Schwartz inequality again,

$$\text{SNR} \le \frac{A^2\|\boldsymbol{g}\|^2\|\boldsymbol{s}\|^2}{\sigma^2\|\boldsymbol{g}\|^2} = \frac{A^2\|\boldsymbol{s}\|^2}{\sigma^2} = \frac{A^2}{\sigma^2},$$

where the equality holds if and only if $\boldsymbol{g} = c\boldsymbol{s}$ for some constant $c$.

**18.13 The slope of the tangent of the ROC curve.** By differentiating (18.87) and (18.89) with respect to the parameter $t_\alpha$,

$$\frac{d\alpha}{dt_\alpha} = -\frac{1}{\sigma}\phi\left(\frac{t_\alpha}{\sigma}\right)$$
$$\frac{dP_d}{dt_\alpha} = -\frac{1}{\sigma}\phi\left(\frac{t_\alpha}{\sigma} - r\right),$$

where $= A/\sigma$ and $\phi(\cdot)$ is the PDF of $N(0,1)$. Taking the ratio of the above two expressions,

$$\frac{dP_d}{d\alpha} = \frac{\phi\left(\frac{t_\alpha}{\sigma} - r\right)}{\phi\left(\frac{t_\alpha}{\sigma}\right)}$$

From (18.83),

$$\frac{t_\alpha}{\sigma} = \frac{r}{2} + r^{-1}\log\lambda_\alpha$$
$$\frac{t_\alpha}{\sigma} - r = -\frac{2}{2} + r^{-1}A\log\lambda_\alpha.$$

By substituting these expressions into the argument of $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$, we find

$$\frac{dP_d}{d\alpha} = \frac{\exp\left\{-\frac{1}{2}\left(\frac{r}{2} + r^{-1}\log\lambda_\alpha\right)^2\right\}}{\exp\left\{-\frac{1}{2}\left(\frac{r}{2} + r^{-1}\log\lambda_\alpha\right)^2\right\}} = \exp\log\lambda_\alpha = \lambda_\alpha.$$

**18.14 The area under the ROC curve and SNR parameter $r$.** Using the relations $\alpha = 1 - \Phi(u_\alpha)$, $\frac{d\alpha}{du_\alpha} = -\phi(u_\alpha)$, and $P_d = 1 - \Phi(u_\alpha - r)$, we have

$$\int_0^1 P_d\,d\alpha = -\int_\infty^{-\infty}[1 - \Phi(u_\alpha - r)]\phi(u_\alpha)\,du_\alpha = \int_{-\infty}^{\infty}\Phi(x + r)\phi(x)\,dx \triangleq S(r)$$

where we used the properties $1 - \Phi(u) = \Phi(-u)$ and $\phi(-u) = \phi(u)$. It is clear $S(0) = \int_{-\infty}^{\infty}\Phi(x))\phi(x)\,dx = \int_0^1 y\,dy = \frac{1}{2}$, where we set $y = \Phi(x)$. In the limit $r \to \infty$, $\Phi(x + r) = 1$ for all $x$. Hence $S(\infty) = \int_{-\infty}^{\infty}\phi(x)\,dx = 1$. The function $S(r)$ is monotone increasing, since

$$S'(r) = \int_{-\infty}^{\infty}\phi(x + r)\phi(x)\,dx \geq 0,$$

because $\phi(x)$ is a PDF.
Thus, the area surrounded by the ROC curve and the straight line $P_d = \alpha$ is

$$\int_0^1 (P_d - \alpha)\,d\alpha = S(r) - \frac{1}{2}.$$

Hence it increases from 0 to $\frac{1}{2}$, as $r$ increases from 0 to $\frac{1}{2}$.

**18.15 Exponential distribution with two different parameters.**

(a) Since $f_X(x;\theta_i) = \theta_i e^{-\theta_i x}$, then the Neyman-Pearson test is

$$\text{Accept } H_1 \text{ if } \Lambda(x) = \frac{\theta_1}{\theta_0}e^{-(\theta_1 - \theta_0)x} \geq \lambda_\alpha,$$

which is equivalent to

$$\text{Accept } H_1 \text{ if } x \leq \frac{1}{\theta_1 - \theta_0}\log\left(\frac{\theta_1}{\theta_0\lambda_\alpha}\right) \triangleq x_\alpha$$

The critical value $\lambda_\alpha$ is found from

$$\alpha = \int_0^{x_\alpha}\theta_0 e^{-\theta_0 x}\,dx = 1 - \exp\left(-\theta_0 x_\alpha\right), \tag{11}$$

from which

$$\log(1-\alpha) = \frac{\theta_0}{\theta_1 - \theta_0} \log\left(\frac{\theta_1}{\theta_0 \lambda_\alpha}\right).$$

Then we find

$$\lambda_\alpha = \frac{\theta_1}{\theta_0} \exp\left[\frac{\theta_1 - \theta_0}{\theta_0} \log(1-\alpha)\right]$$

(b) The detection probability can be calculated as

$$P_d = \int_0^{x_\alpha} \theta_1 e^{-\theta_1 x}\, dx = 1 - \exp\left(-\theta_1 x_\alpha\right), \tag{12}$$

From (11) and (12), $\log(1-\alpha) = -\theta_0 x_\alpha$ and $\log(1-P_d) = -\theta_0 x_\alpha$. Then we find the closed expression for the ROC curve.

$$P_d = 1 - (1-\alpha)^{\frac{\theta_1}{\theta_0}}.$$

The slope of the tangent is

$$\frac{dP_d}{d\alpha} = \frac{\theta_1}{\theta_0}(1-\alpha)^{\frac{\theta_1}{\theta_0}-1}.$$

Thus, the slope at $(0,0)$ is $\frac{\theta_1}{\theta_0} = \lambda_0$, and the slope at $(1,1)$ is $0 = \lambda_1$.

# 19 Solutions for Chapter 19: Estimation Algorithms

## 19.1 Classical Numerical Methods of Estimation

**19.1* Nonnegativity of KLD.**
We can extend any of the methods used in proving Shannon's lemma (or Gibbs' inequality) discussed in Section 10.1.3.

(a) If we use the inequality $\ln x \leq x - 1$, then

$$\log \frac{g(\boldsymbol{x})}{f(\boldsymbol{x})} \leq (\log e)\left(\frac{g(\boldsymbol{x})}{f(\boldsymbol{x})} - 1\right).$$

Then

$$\begin{aligned} D(\|\boldsymbol{g}) &= \int f(\boldsymbol{x}) \log \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}\, dx = -\int f(\boldsymbol{x}) \log \frac{g(\boldsymbol{x})}{f(\boldsymbol{x})}\, d\boldsymbol{x} \\ &\geq -\log e \int f(\boldsymbol{x}) \left(\frac{g(\boldsymbol{x})}{f(\boldsymbol{x})} - 1\right) d\boldsymbol{x} \\ &= -\log e \left[\int g(\boldsymbol{x})\, d\boldsymbol{x} - \int f(\boldsymbol{x})\, d\boldsymbol{x}\right] = -\log e(1-1) = 0. \end{aligned}$$

(b) **Use of Jensen's inequality.**
Since $\log x$ is a concave function, we have from Jensen's inequality

$$E_f\left[\log \frac{g(\boldsymbol{X})}{f(\boldsymbol{X})}\right] \leq \log E_f\left[\frac{g(\boldsymbol{X})}{f(\boldsymbol{X})}\right] = \log \int f(\boldsymbol{x}) \frac{g(\boldsymbol{x})}{f(\boldsymbol{x})}\, d\boldsymbol{x} = 0, \tag{1}$$

where equality holds when $\frac{g(\boldsymbol{x})}{f(\boldsymbol{x})} =$ constant for all $i$. This constant must be unity, since $\int f(\boldsymbol{x})\, d\boldsymbol{x} = \int g(\boldsymbol{x})\, d\boldsymbol{x} = 1$. Thus $f(\boldsymbol{x}) = g(\boldsymbol{x})$ for all $\boldsymbol{x}$. Hence

$$\int f(\boldsymbol{x}) \log g(\boldsymbol{x})\, d\boldsymbol{x} \leq \int f(\boldsymbol{x}) \log f(\boldsymbol{x})\, d\boldsymbol{x}, \tag{2}$$

from which $D(\boldsymbol{f}\|\boldsymbol{g}) \geq 0$ follows.

(c) **Lagrangian multiplier method:** Consider

$$F(g) = \int f(\boldsymbol{x}) \log g(\boldsymbol{x})\, d\boldsymbol{x}.$$

Since $\log g$ is a concave function of $g$ and $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$, $F(g)$ is a concave function of $g(\boldsymbol{x})$. Thus, if we find a stationary point, it becomes the point of a global maximum.

Define

$$J(g,\lambda) = \int f(\boldsymbol{x}) \log g(\boldsymbol{x})\, d\boldsymbol{x} + \lambda \left( \int g(\boldsymbol{x})\, d\boldsymbol{x} - 1 \right). \tag{3}$$

Differentiate it with respect to $g$, and $\lambda$ and set them all to zero:

$$\begin{aligned} \frac{\partial J(\boldsymbol{g},\lambda)}{\partial g(\boldsymbol{x})} &= \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} + \lambda = 0, \ \text{ for } \ -\infty < \boldsymbol{x} < \infty, \\ \frac{\partial J(\boldsymbol{g},\lambda)}{\partial \lambda} &= \int g(\boldsymbol{x})\, d\boldsymbol{x} - 1 = 0. \end{aligned}$$

From the first equation we find

$$f(\boldsymbol{x}) = -\lambda g(\boldsymbol{x}), \ \text{ for } \ -\infty < \boldsymbol{x} < \infty, \tag{4}$$

and substituting to the last equation (i.e., the original constraint equation) and using $\int f(\boldsymbol{x})\, d\boldsymbol{x} = 1$, we find

$$\lambda = -1. \tag{5}$$

Thus, the condition for a stationary point is

$$g(\boldsymbol{x}) = f(\boldsymbol{x}), \ \text{ for } \ -\infty < \boldsymbol{x} < \infty. \tag{6}$$

Thus, by substituting the condition (6) into (3), we attain the maximum of $J$:

$$J_{\max} = F_{\max} = \int f(\boldsymbol{x}) \log f(\boldsymbol{x})\, d\boldsymbol{x},$$

from which the nonnegativity of KLD follows.

**19.2 KLD between $\theta$ and $\theta'$.**

$$\begin{aligned} D(\boldsymbol{\theta}\|\boldsymbol{\theta}') &= \sum_{\boldsymbol{z}} \frac{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta})}{L(\boldsymbol{\theta})} \log \left( \frac{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta}) L(\boldsymbol{\theta}')}{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta}') L(\boldsymbol{\theta})} \right) \\ &= \sum_{\boldsymbol{z}} \frac{p(\boldsymbol{z},\boldsymbol{y};\boldsymbol{\theta})}{L(\boldsymbol{\theta})} \left[ \log \frac{L(\boldsymbol{\theta}')}{L(\boldsymbol{\theta})} + \log \frac{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta})}{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta}')} \right] \\ &= \frac{p(\boldsymbol{y};\theta)}{L(\theta)} \log \frac{L(\boldsymbol{\theta}')}{L(\boldsymbol{\theta})} + \sum_{\boldsymbol{z}} \frac{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta})}{L(\boldsymbol{\theta})} \log \frac{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta})}{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta}')} \\ &= \log \frac{L(\boldsymbol{\theta}')}{L(\boldsymbol{\theta})} + \sum_{\boldsymbol{z}} \frac{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta})}{L(\boldsymbol{\theta})} \log \frac{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta})}{p(\boldsymbol{y}.\boldsymbol{z};\boldsymbol{\theta}')}. \end{aligned}$$

**19.3 Approximation of $\ln x$ and Kullback's information criterion $I(\boldsymbol{n};\boldsymbol{\theta})$ and $\chi^2(\boldsymbol{n};\boldsymbol{\theta})$ statistic.** The approximation formula (19.10) is most accurate around $x = 1$. At $x = 1$, both sides are zero, and the slope is unity on both sides. Since $\frac{n_i}{np_i(\boldsymbol{\theta})} \approx 1$, we have

$$\begin{aligned} I(\boldsymbol{n};\boldsymbol{\theta}) &= \sum_{i=1}^{r} n_i \log \left( \frac{n_i}{np_i(\boldsymbol{\theta})} \right) \approx \frac{1}{2} \left[ \sum_{i=1}^{r} n_i \left( \frac{n_i}{np_i(\boldsymbol{\theta})} - \frac{np_i(\boldsymbol{\theta})}{n_i} \right) \right] \\ &= \frac{1}{2} \left( \sum_{i=1}^{r} \frac{n_i^2}{np_i(\boldsymbol{\theta})} - n \right) = \frac{1}{2} \chi^2(\boldsymbol{n};\boldsymbol{\theta}). \end{aligned}$$

As $n \to \infty$, $\frac{n_i}{np_i(\boldsymbol{\theta})} \to 1$ for all $i = 1, 2, \ldots, r$, which means that the approximation becomes exact asymptotically.

## 19.2 Expectation-Maximization Algorithm

**19.4 Relation between $Q$-function and $\log L$-function.**
From (19.23)

$$\log L_{\boldsymbol{y}}(\boldsymbol{\theta}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) + H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}),$$
$$\log L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)}) = Q(\boldsymbol{\theta}^{(p)}|\boldsymbol{\theta}^{(p)}) + H(\boldsymbol{\theta}^{(p)}|\boldsymbol{\theta}^{(p)}).$$

Subtracting the second equation from the above,

$$\begin{aligned}\log L_{\boldsymbol{y}}(\boldsymbol{\theta}) - \log L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)}) &= Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) - Q(\boldsymbol{\theta}^{(p)}|\boldsymbol{\theta}^{(p)}) \\ &\quad + \left[H(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) - H(\boldsymbol{\theta}^{(p)}|\boldsymbol{\theta}^{(p)})\right] \\ &\geq Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) - Q(\boldsymbol{\theta}^{(p)}|\boldsymbol{\theta}^{(p)}),\end{aligned}$$

where we used the fact that the expression in $[\ldots]$ is nonnegative because of (**??**). Hence the relation (19.27) follows.
Alternatively, can derive the relation without using the equations in the hint. We can write

$$\begin{aligned}L_{\boldsymbol{y}}(\boldsymbol{\theta}) = p_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) &= \sum_{\boldsymbol{x}} p_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta}) = p_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}^{(p)}) \sum_{\boldsymbol{x}} \frac{p_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})}{p_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}^{(p)})} \\ &= L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)}) \sum_{\boldsymbol{x}} p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}^{(p)}) \frac{p_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})}{p_{\boldsymbol{XY}}(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta}^{(p)})} \\ &= L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)}) \sum_{\boldsymbol{x}} p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y};\boldsymbol{\theta}^{(p)}) \frac{p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}^{(p)})} \\ &= L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)}) E\left[\left.\frac{p_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{\theta})}{p_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{\theta}^{(p)})}\right| \boldsymbol{y};\boldsymbol{\theta}^{(p)}\right].\end{aligned} \tag{7}$$

Then, consider

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) - Q(\boldsymbol{\theta}^{(p)}|\boldsymbol{\theta}^{(p)}) &= E\left[\left.\left(\log p_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{\theta}) - \log p_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{\theta}^{(p)})\right)\right| \boldsymbol{y};\boldsymbol{\theta}^{(p)}\right] \\ &= E\left[\left.\log\left(\frac{p_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{\theta})}{p_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{\theta}^{(p)})}\right)\right| \boldsymbol{y};\boldsymbol{\theta}^{(p)}\right] \\ &\leq \log E\left[\left.\frac{p_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{\theta})}{p_{\boldsymbol{X}}(\boldsymbol{X};\boldsymbol{\theta}^{(p)})}\right| \boldsymbol{y};\boldsymbol{\theta}^{(p)}\right],\end{aligned} \tag{8}$$

where we used Jensen's inequality to obtain the last expression. From the last two equations, we find (19.27).

**19.5 KLD and $Q$-function.**

(a) From the last equation of Problem 19.2, and the definition of the $Q$-function, it readily follows.
(b) Since $D(\boldsymbol{\theta}\|\boldsymbol{\theta}') \geq 0$, the inequality follows.

**19.6 Lower bound $B$ to the log-likelihood function.**

(a) We can write the log-likelihood function as

$$\log p(\boldsymbol{y};\boldsymbol{\theta}) = \log \sum_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} p(\boldsymbol{y},\boldsymbol{z};\boldsymbol{\theta}) = \log \sum_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \alpha(\boldsymbol{z}) \frac{p(\boldsymbol{y},\boldsymbol{z};\boldsymbol{\theta})}{\alpha(\boldsymbol{z})}, \tag{9}$$

where $\alpha(\boldsymbol{z})$ is an arbitrary probability distribution. By applying Jensen's inequality, we have

$$\log L_{\boldsymbol{y}}(\boldsymbol{\theta}) \geq \sum_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \alpha(\boldsymbol{z}) \log \frac{p(\boldsymbol{y},\boldsymbol{z};\boldsymbol{\theta})}{\alpha(\boldsymbol{z})} = B(\boldsymbol{\theta};\alpha(\boldsymbol{z})). \tag{10}$$

The equality holds when

$$\frac{p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta})}{\alpha(\boldsymbol{z})} = c \;\; \text{(constant)} \;\; \text{or} \;\; p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta}) = c\alpha(\boldsymbol{z}), \;\; \text{for all} \;\; \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}. \tag{11}$$

Since $\{p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{\theta})\}$ and $\{\alpha(\boldsymbol{z})\}$ are both probability distributions, the constant $c$ must be unity, hence (19.50) must hold in order for the lower bound $B(\boldsymbol{\theta},\alpha(\boldsymbol{z}))$ to be equal to the log-likelihood $\log L_{\boldsymbol{y}}(\boldsymbol{\theta})$.
(b)

$$\begin{aligned}
B(\boldsymbol{\theta};\alpha(\boldsymbol{z})) &= \sum_{\boldsymbol{z}} \alpha(\boldsymbol{z}) \log \frac{p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta})p(\boldsymbol{y};\boldsymbol{\theta})}{\alpha(\boldsymbol{z})} \\
&= \sum_{\boldsymbol{z}} \alpha(\boldsymbol{z}) \log \frac{p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta})}{\alpha(\boldsymbol{z})} + \log p(\boldsymbol{y};\boldsymbol{\theta}) \\
&= -\sum_{\boldsymbol{z}} \alpha(\boldsymbol{z}) \log \frac{\alpha(\boldsymbol{z})}{p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta})} + \log p(\boldsymbol{y};\boldsymbol{\theta}) \\
&= -D(\alpha(\boldsymbol{z})||p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta})) + \log L_{\boldsymbol{y}}(\boldsymbol{\theta}) \leq L_{\boldsymbol{y}}(\boldsymbol{\theta}) \leq \log L_{\boldsymbol{y}}(\boldsymbol{\theta}),
\end{aligned}$$

From this, it is also apparent that $B(\boldsymbol{\theta};\alpha(\boldsymbol{z})) = \log L_{\boldsymbol{y}}(\boldsymbol{\theta})$, if and only if $D(\alpha(\boldsymbol{z})||p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{\theta})) = 0$, i.e., $\alpha(\boldsymbol{z}) = p(\boldsymbol{z}|\boldsymbol{y},\boldsymbol{\theta})$, as proved in part (a).
(c) Let

$$\begin{aligned}
J(\alpha(\boldsymbol{z}),\lambda) &= B(\boldsymbol{\theta};\alpha(\boldsymbol{z})) + \lambda \left( \sum_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \alpha(\boldsymbol{z}) - 1 \right) \\
&= \sum_{\boldsymbol{z}_i\in\boldsymbol{\mathcal{Z}}} \alpha(\boldsymbol{z}) \log p(\boldsymbol{y},\boldsymbol{z};\boldsymbol{\theta}) - \sum_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \alpha(\boldsymbol{z}) \log \alpha(\boldsymbol{z}) + \lambda \left( \sum_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} \alpha(\boldsymbol{z}) - 1 \right),
\end{aligned}$$

Note $\alpha(\boldsymbol{z})$ is the probability value assigned to $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$, the sample space of the RV $\boldsymbol{Z}$. Take the partial derivative of $J$ respect to $\alpha(\boldsymbol{z})$ for each $\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}$ and $\lambda$ and set all equal to zero:

$$\frac{\partial J}{\partial \alpha(\boldsymbol{z})} = \log p(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta}) - \log \alpha(\boldsymbol{z}) - 1 + \lambda = 0, \text{ for all } \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}; \tag{12}$$

$$\frac{\partial J}{\partial \lambda} = \sum_{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}} \alpha(\boldsymbol{z}) - 1 = 0, \tag{13}$$

yielding

$$\alpha(\boldsymbol{z}) = p(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta}) e^{\lambda - 1}. \text{ for all } \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}. \tag{14}$$

By substituting this into the constraint equation (13), we finally obtain

$$\alpha(\boldsymbol{z}) = \frac{p(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta})}{\sum_{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}} p(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta})} = p(\boldsymbol{z}|\boldsymbol{y}; \boldsymbol{\theta}), \;\; \boldsymbol{z} \in \boldsymbol{\mathcal{Z}}. \tag{15}$$

Since $-\sum_{\boldsymbol{z}} \alpha(\boldsymbol{z}) \log \alpha(\boldsymbol{z})$ (which is the entropy of the distribution $\alpha(\boldsymbol{z})$ is a concave function of $\boldsymbol{z}$, so is our objective function $B(\boldsymbol{\theta}; \alpha(\boldsymbol{z}))$. So the $\alpha(\boldsymbol{z})$ of (15) maximizes $B(\boldsymbol{\theta}; \alpha(\boldsymbol{z}))$. The concavity of the function $B$ also follows by showing that Hessian matrix of $J$ is negative semi-definite, namely.

$$\begin{aligned}
\frac{\partial^2 J}{\partial \alpha(z_i) \partial \alpha(z_j)} &= -\frac{\delta_{i,j}}{\alpha(\boldsymbol{z}_i)}, \text{ for all } \boldsymbol{z}_i, \boldsymbol{z}_j \in \boldsymbol{\mathcal{Z}}; \\
\frac{\partial^2 J}{\partial \alpha(z_i) \partial \lambda} &= 0, \text{ for all } \boldsymbol{z}_i \in \boldsymbol{\mathcal{Z}}; \\
\frac{\partial^2 J}{\partial \lambda^2} &= 0.
\end{aligned}$$

**19.7 The reachable lower bound $B^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ of the likelihood function.**

(a)

$$\begin{aligned}
B^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) &= \sum_{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}} p(\boldsymbol{z}|\boldsymbol{y}; \boldsymbol{\theta}^{(p)} \log \frac{p(\boldsymbol{z}, \boldsymbol{y}; \boldsymbol{\theta})}{p(\boldsymbol{z}|\boldsymbol{y}; \boldsymbol{\theta}^{(p)})} \\
&= \sum_{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}} p(\boldsymbol{z}|\boldsymbol{y}; \boldsymbol{\theta}^{(p)} \left[ \log p(\boldsymbol{z}, \boldsymbol{y}; \boldsymbol{\theta}) - \log p(\boldsymbol{z}|\boldsymbol{y}; \boldsymbol{\theta}^{(p)}) \right] \\
&= E\left[ \log p(\boldsymbol{Z}, \boldsymbol{y}; \boldsymbol{\theta}) | \boldsymbol{y}; \boldsymbol{\theta}^{(p)} \right] - \sum_{\boldsymbol{z} \in \boldsymbol{\mathcal{Z}}} p(\boldsymbol{z}|\boldsymbol{y}; \boldsymbol{\theta}^{(p)}) \log p(\boldsymbol{z}|\boldsymbol{y}; \boldsymbol{\theta}^{(p)}) \\
&= Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) + H(\boldsymbol{Z}|\boldsymbol{y}; \boldsymbol{\theta}^{(p)}).
\end{aligned} \tag{16}$$

(b) If we set $\boldsymbol{\theta} = \boldsymbol{\theta}^{(p)}$ in the above equation, the logarithmic term in the RHS of the first line becomes

$$\log \frac{p(\boldsymbol{z}, \boldsymbol{y}; \boldsymbol{\theta}^{(p)})}{p(\boldsymbol{z}|\boldsymbol{y}; \boldsymbol{\theta}^{(p)})} = \log p(\boldsymbol{y}; \boldsymbol{\theta}^{(p)}).$$

Thus,

$$B^*(\boldsymbol{\theta}^{(p)}|\boldsymbol{\theta}^{(p)}) = \sum_{\boldsymbol{z}\in\boldsymbol{\mathcal{Z}}} p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta}^{(p)} \log p(\boldsymbol{y};\boldsymbol{\theta}^{(p)})$$
$$= \log p(\boldsymbol{y};\boldsymbol{\theta}^{(p)}) = \log L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)}). \tag{17}$$

So the lower bound $B^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ touches $\log_{\boldsymbol{y}} L(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(p)}$.

**An alternative proof:** From the derivation of the inequality (19.51),

$$B(\boldsymbol{\theta};\alpha(\boldsymbol{z})) \le L_{\boldsymbol{y}}(\boldsymbol{\theta}),$$

we know the equality holds, if and only if

$$D(\alpha(\boldsymbol{z})||p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta})) = 0,$$

or when

$$\alpha(\boldsymbol{z}) = p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta}).$$

Since we set $\alpha(\boldsymbol{z}) = p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta}^{(p)})$ in defining $B^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$, it follows that $B^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = \log L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})$, if and only if $p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta}) = p(\boldsymbol{z}|\boldsymbol{y};\boldsymbol{\theta}^{(p)})$. This last equation is satisfied if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}^{(p)}$.

**19.8 An alternative derivation of the EM algorithm.**
Any $\boldsymbol{\theta}$ that increases $B^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ will in turn increase $L_{\boldsymbol{y}}(\boldsymbol{\theta})$, as discussed in the text. So choose $\boldsymbol{\theta} = \boldsymbol{\theta}^{(p+1)}$ such that $B^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ is maximized. But

$$\begin{aligned}\boldsymbol{\theta}^{(p+1)} &= \arg\max_{\boldsymbol{\theta}} B^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) \\ &= \arg\max_{\boldsymbol{\theta}} \left[Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) + H(\boldsymbol{Z}|\boldsymbol{y};\boldsymbol{\theta}^{(p)})\right] \\ &= \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}),\end{aligned}$$

where we dropped the $H$ term, which does not depend on $\boldsymbol{\theta}$. Hence, maximizing the tight lower bound function $B^*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ is equivalent to maximizing the $Q$-function, given the current estimate $\boldsymbol{\theta}^{(p)}$. Hence we have derived the EM algorithm.

**19.9 EM algorithm for a MAP estimate.**

(a) The EM algorithm for a MAP estimate can be stated as in Algorithm 19.1.
(b) The Q-function of the previous algorithm can be written as

$$Q_{MAP}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) + \log\pi(\boldsymbol{\theta}) = E[\log p(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{y},\boldsymbol{\theta}^{(p)}) + \log\pi(\boldsymbol{\theta})$$

According to the Bayes' theorem

$$p(\boldsymbol{X};\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\theta}^{(p)}) = \frac{p(\boldsymbol{X}|\boldsymbol{y},\boldsymbol{\theta}^{(p)})\pi(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{y},\boldsymbol{\theta}^{(p)})}{\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{\theta}^{(p)})}$$

Substituting this into the previous equation and taking into account that $p(\boldsymbol{X}|\boldsymbol{y},\boldsymbol{\theta}^{(p)})$ does not depend on $\boldsymbol{\theta}$, we obtain the desired result.

**Algorithm 19.1 EM Algorithm for MAP estimate**

1: Denote the initial estimate of the model parameters by $\boldsymbol{\theta}^{(0)}$. Set the iteration number $p = 0$.
2: Assume the $p$-th estimate $\boldsymbol{\theta}^{(p)}$.
3: Evaluate

$$E[\log p_{\boldsymbol{ZY}|\boldsymbol{\Theta}}(\boldsymbol{Z}, \boldsymbol{y}|\boldsymbol{\theta})|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}] \triangleq Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}). \text{ \textbf{(E-step)}}$$

4: Find

$$\boldsymbol{\theta}^{(p+1)} = \arg\max_{\boldsymbol{\theta}} \left[Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) + \log \pi(\boldsymbol{\theta})\right]. \text{ \textbf{(M-step for MAP)}}$$

5: If any of the stopping conditions is achieved, go to step 6. Otherwise, replace $\boldsymbol{\theta}^{(p)}$ by $\boldsymbol{\theta}^{(p+1)}$, set $p \leftarrow p + 1$, and repeat the Steps 2 through 5.
6: Output $\boldsymbol{\theta}^{(p+1)}$ as a MAP estimate.

**19.10* EM algorithm when the complete variables come from the exponential family of distributions.**

By substituting (19.56) into (19.24), we find

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}(p)) = E[\log h(\boldsymbol{X})|\boldsymbol{y}; \boldsymbol{\theta}^{(p)}] + \boldsymbol{\eta}^\top(\boldsymbol{\theta})\boldsymbol{T}^{(p)} - A(\boldsymbol{\theta}),$$

where

$$\boldsymbol{T}^{(p)} = E[\boldsymbol{T}(\boldsymbol{x})|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}].$$

Since the first term $E[\log h(\boldsymbol{X})|\boldsymbol{y}; \boldsymbol{\theta}^{(p)}]$ in the above expansion is independent of $\boldsymbol{\theta}$, the M-step is reduced to

$$\boldsymbol{\theta}^{(p+1)} = \arg\max_{\boldsymbol{\theta}} \left[\boldsymbol{\eta}^\top(\boldsymbol{\theta})\boldsymbol{T}^{(p)} - A(\boldsymbol{\theta})\right]$$

# 20 Solutions for Chapter 20: Hidden Markov Models and Applications

## 20.1 Introduction

## 20.2 Formulation of a Hidden Markov Model

**20.1* Observable process $\boldsymbol{Y}(t)$.** Since $Y_t$ is a probabilistic function of $S_{t-1}$ and $S_t$, we write

$$y_t = f(s_t, s_{t-1}), \ \ y_{t-1} = f(s_{t-1}, s_{t-2}), \ldots.$$

In order for $Y_t$ to be a Markov chain, we must have

$$p(y_t|y_{t-1}, y_{t-2}, \ldots) = p(y_t|y_{t-1}).$$

The LHS can be written as

$$\text{LHS} = p(y_t|s_t, s_{t-1}, s_{t-2}, s_{t-3}, \ldots)$$

and the RHS is representable as

$$\text{RHS} = p(y_t|y_{t-1}) = p(y_t|s_t, s_{t-1}, s_{t-2}),$$

which contradicts to the definition that $Y_t$ depends only on $S_t$ and $S_{t-1}$, not on $S_{t-2}$. Thus, $Y_t$ is not a simple Markov process.

**21.2 An HMM representation of the convolutional encoder with a state-based output model.** Suppose we define the state of the encoder in terms of the three bits in the shift register, i.e., $S_t = (I_t, I_{t-1}, I_{t-2})$. Then the state space has eight elements $\{(000), (001), (010) \ldots, (111)\} = \{0, 1, 2, \ldots, 7\}$. Thus both the state transition diagram and trellis diagram become rather large. It is clear $S_t$ is also a simple Markov chain.

The encoder output $O_t = O_t(1)O_t^{(2)}$ is given by the same set of equations (20.34) through (20.36). If the channel is a discrete memoryless channel as in Example 20.1, $Y_t = Y_t^{(1)}Y_t^{(2)}$ depends only on the channel input of the present time $t$: $O_t = O_t^{(1)}O_t^{(2)}$. Thus, $(S_t, Y_t)$ is a hidden Markov process with

$$p(y_t|s_{t-1}, s_t) = p(y_t|s_t).$$

Thus, it is a state-based HMM.

The model parameters $c(i; j, k)$ takes a simple product form

$$c(i; j, k) = a(i; j)b(j; k), \ \ i, j \in \mathcal{S}, \ \ k \in \mathcal{Y},$$

where

$$b(j;k) = P(Y_t = k|S_t = j), \;\; j \in \mathcal{S}, \;\; k \in \mathcal{Y}.$$

If the channel output $Y_t$ is a continuous random variable as in the AWGN channel, we define the conditional PDF $f_{Y|S}(y|j)$ by

$$f_{Y|S}(y|j)\,dy = P[y < Y_t \leq y + dy|S_t = j].$$

If we assume that the signal is sent as bipolar pulses over an AWGN, as in Problem 20.5, then the coded two bits $X_t(j)$ are sent as a pair $(X_t^{(1)}(j), X_t^{(2)}(j))$ both of which are pulses of amplitude $+A$ or $-A$.

For instance, if the state (the register contents) is $S_t = (1,1,0)$, then $O_t = 01 = 1$, hence

$$f_{Y|S}(y_t^{(1)}, y_t^{(2)}|(1,1,0)) = \frac{1}{2\pi\sigma^2}\exp - \frac{(y_t^{(1)} + A)^2 + (y_t^{(2)} - A)^2}{2\sigma^2}.$$

All other cases of $f_{Y|S}(y_t^{(1)}, y_t^{(2)}|s_t)$ can be found in a similar fashion.

**20.3 $S_t$ of the convolutional encoder is a Markov chain.**

Since $S_t = (I_t, I_{t-1})$, and $I_{t-1}$ partially defines $S_{t-1}(= (I_{t-1}, I_{t-2}))$, $S_t$ depends partially on $S_{t-1}$ Since $I_t$ is statistically independent of $S_{t-k};\;\; k \geq 1$, it follows $S_t$ is a probabilistic function of $S_{t-1}$ only, hence it is a simple Markov chain.

**20.4 Equilibrium state distribution and mean state sojourn time.**

(a) The equilibrium state distribution $\boldsymbol{\pi}^\top = [\pi_G, \pi_B]$ of a Markov chain should satisfy (15.99):

$$\begin{aligned}\boldsymbol{\pi}^\top \tilde{\boldsymbol{A}} &= \boldsymbol{\pi}^\top \\ \boldsymbol{\pi}^\top \mathbf{1} &= 1,\end{aligned}$$

which gives

$$\begin{aligned}&\pi_0 p_{00} + \pi_B(1 - p_{11}) = \pi_0 \\ &\pi_0 + \pi_1 = 1.\end{aligned}$$

from which we obtain (20.44).

(b) Suppose that the Markov chain $\tilde{S}_t$ enters into state 0 at time $t$. The probability that it stays there for $k$ consecutive time units is geometrically distributed as follows:

$$P[S_{t+1} = 0, \ldots, S_{t+k-1} = 0, S_{t+k} = 1|S_t = 0] = p_{00}^{k-1}(1 - p_{00}), \;\; k \geq 1.$$

Then, the average of the random variable $k$ is $T_0$, and can be easily derived as we studied in Chapter 3.

$$T_0 = (1-p_{00})\sum_{k=0}^{\infty} k p_{00}^{k-1}$$
$$(1-p_{00})\frac{\partial}{\partial p_{00}}\sum_{k=0}^{\infty} p_{00}^{k}$$
$$= (1-p_{00})\frac{\partial}{\partial p_{00}}\frac{1}{1-p_{00}} = \frac{1}{1-p_{00}}.$$

**20.5 A convolutional encoder and an additive white Gaussian noise (AWGN) channel.**
In the rate 1/2 convolutional encoder, for every message bit $I_t$, there are two encoder output symbols, which are sent in the form of bipolar signal $\pm A$, i.e., the binary 1 and 0 in the encoder output are sent at $+A$ and $-A$, respectively. We denote the transmitted symbols as $X_t^{(1)}$ and $X_t^{(2)}$. We either send them in series or in parallel, and we are not concerned about its specific implementation here. We write the noisy output from the AWGN channel by $Y_t^{(1)}$ and $Y_t^{(2)}$, respectively:

$$Y_t^{(1)} = X_t^{(1)} + N_t^{(1)}$$
$$Y_t^{(2)} = X_t^{(2)} + N_t^{(2)}$$

Because of the whiteness of noise $\{N_t^{(1)}, N_t^{(2)}\}$ are independent to each other, and have a common distribution $N(0,\sigma^2)$. Then we have

$$c(S_{t-1}=i; S_t=j, Y_t) = a(i;j)\frac{1}{2\pi\sigma^2}\exp -\frac{[Y_t^{(1)} - X_t^{(1)}(i;j)]^2 + [Y_t^{(2)} - X_t^{(2)}(i;j)]^2}{2\sigma^2},$$

For example, if $i=(1,0)=2$ and $j=(1,1)=3$, then $a(i;j)=1/2$ and $O_t=01$. Then the transmitted signals are $X_t^{(1)} = -A$ and $X_t^{(2)} = +A$. The model parameters $\boldsymbol{\theta}$ in this case are: $\boldsymbol{\theta} = (A, \sigma, \{a(i;j)\})$.

**20.6* Partial-response channel.**

(a) Define the *state of the transmitter*, which is *hidden* as the transmitted information itself, i.e.,

$$S_t = I_t.$$

Then, the output $X_t$ from the partial-response channel, in the absence of noise, can be written as

$$X_t = A(S_t - S_{t-1}).$$

Hence, the conditional PDF of $Y_t = y$ given a state transition $S_{t-1} = 1 \to S_t = j, \quad (t \geq 1)$ is, in referring to (**??**),

$$f_{Y_t|S_{t-1}S_t}(y\,|\,i,j) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(y-x(i,j))^2}{2\sigma^2}\right\},$$

where the noise-free output $x(i,j)$ associated with a state transition $i \to j$ is

$$x(i,j) = A(j-i), \ \ i,j \in \{0,1\}.$$

The Markov chain $\{S_t\}$ in this case is simply a zero-th order Markov chain. The state transition probability matrix is

$$\boldsymbol{A} = [a(i;j)] = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}. \tag{1}$$

**Figure 20.1** Trellis diagrams of the HMM representations of a partial-response channel output with additive white Gaussian noise. We assume the initial bit $I_0 = 0$: (a) the transition-based output model; (b) the state-based output model.

(b) Define the state as

$$S_t \triangleq (I_{t-1}, I_t), \ \ t = 1, 2, \ldots.$$

Thus, the state space now consists of four states:

$$\mathcal{S} = \{00, 01, 10, 11\} \triangleq \{0, 1, 2, 3\}. \tag{2}$$

with the state transition matrix

$$\boldsymbol{A} = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}, \tag{3}$$

The conditional PDF (20.32) of the output, given the current states is:

$$f_{Y_t|S_t}(y|s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y - x(s))^2}{2\sigma^2}\right\},$$

where the output $x(s)$ for state $s \in \{0, 1, 2, 3\}$ is defined by

$$x(s) = \begin{cases} +A & \text{for } s = 1; \\ 0 & \text{for } s = 0, \text{ or } 3; \\ -A & \text{for } s = 2. \end{cases}$$

Figure 20.1 (b) shows the HMM with state-based output; state $S_t = s$ corresponds to $s = (i,j) = (I_{t-1}, I_t)$, and the number attached to each state $s$ is $x(s)$. In both (a) and (b) we assume that $I_0 = 0$, and the receiver should exploit this information in attempting to recover the transmitted information sequence $\{I_0\}$, as we will discuss further later in this chapter.

## 20.3 Evaluation of a Hidden Markov Model

**20.7* Likelihood function as a sum of products.**
Using the Markovian property of the sequence $\boldsymbol{S}$, we can write

$$p(\boldsymbol{s};\boldsymbol{\theta}) = \pi_0(s_0)\prod_{t=0}^{T} a(s_{t-1};s_t), \tag{4}$$

Under the state-based output model, where $\boldsymbol{\theta} = (\boldsymbol{\pi}_0, \boldsymbol{A}, \boldsymbol{B})$, $p(\boldsymbol{y}|\boldsymbol{s};\boldsymbol{\theta})$ can be written as the product of the conditional probabilities $b(s_t;y_t)$:

$$p(\boldsymbol{y}|\boldsymbol{s};\boldsymbol{\theta}) = \prod_{t=0}^{T} b(s_t;y_t). \tag{5}$$

Thus, by taking the product of the last two expressions,

$$p(\boldsymbol{s},\boldsymbol{y};\boldsymbol{\theta}) = p(\boldsymbol{s};\boldsymbol{\theta})p(\boldsymbol{y}|\boldsymbol{s};\boldsymbol{\theta}),$$

and substituting it into (20.48), we have

$$\begin{aligned} L_{\boldsymbol{y}}(\boldsymbol{\theta}) &= \sum_{\boldsymbol{s}\in\mathcal{S}^{T+1}} \pi_0(s_0)b(s_0,y_0)\prod_{t=1}^{T} a(s_{t-1};s_t)b(s_t;y_t) \\ &= \sum_{s_0\in\mathcal{S}}\sum_{s_1\in\mathcal{S}}\cdots\sum_{s_T\in\mathcal{S}} \pi_0(s_0)b(s_0;y_0)a(s_0;s_1)b(s_1;y_1)\cdots a(s_{T-1};s_T)b(s_T;y_T). \end{aligned} \tag{6}$$

Therefore, the likelihood function is again expressed as a *sum of products*.

**20.8 Derivation of the product forms.**

(a) Since $X_t = (S_t, Y_t)$ is a hidden Markov process, we can write

$$\begin{aligned} p(\boldsymbol{x}_0^t) &= p(x_0)p(x_1^t|x_0) \\ &= p(x_0)p(x_1|x_0)p(x_2^t|x_0^1) \\ &= p(x_0)p(x_1|s_0)p(x-2|x_0^1)p(x_3|x_0^2) \\ &= \vdots \\ &= p(x_0)\prod_{k=1}^{t} p(x_k|s_{k-1}) \end{aligned}$$

Thus

$$p(\boldsymbol{s},\boldsymbol{y}) = p(s_0,y_0)\prod_{t=1}^{T} p(s_t,y_t|s_{t-1}).$$

which gives (20.49).

(b) Similarly, for an HMM $\boldsymbol{x}_0^t = (\boldsymbol{s}_0^t, \boldsymbol{y}_0^t)$, we have

$$\begin{aligned}p(\boldsymbol{x}_0^t) &= p(\boldsymbol{x}_0^{t-1}, x_t) = p(\boldsymbol{x}_0^{t-1})p(x_t|\boldsymbol{x}_0^{t-1}) \\ &= p(\boldsymbol{x}_0^{t-1})p(x_t|x_{t-1}) = p(\boldsymbol{x}_0^{t-1})p(x_t|s_{t-1}) \\ &= p(\boldsymbol{s}_0^t, \boldsymbol{y}_0^t)p(s_t, y_t|s_{t-1}),\end{aligned}$$

which yields (20.127)

**20.9* Forward recursion formula when $Y_t$ is a continuous random variable.**
Define the functions $c(i; j, y_t)$ as

$$c(i; j, y_t) = a_t(i; j) f_{Y_t|S_{t-1}S_t}(y_t|(i, j)), \tag{7}$$

where $f_{Y_t|S_{t-1}S_t}(y_t|(i, j))$ is the conditional PDF defined by (**??**). Then from (20.55) we have the same *forward recursion* algorithm:

$$\alpha_t(j, \boldsymbol{y}_0^t) = \sum_{i \in \mathcal{S}} \alpha_{t-1}(i, \boldsymbol{y}_0^{t-1}) c(i; j, y_t), \;\; j \in \mathcal{S}, \; 1 \le t \le T, \tag{8}$$

**20.10 Backward recursion formula.**

$$\begin{aligned}\beta_t(i) = P(\boldsymbol{y}_{t+1}^T|S_t = i) &= \sum_{j \in \mathcal{S}} P(\boldsymbol{y}_{t+1}^T, S_{t+1} = j|S_t = i) \\ &= \sum_{j \in \mathcal{S}} P(y_{t+1}\boldsymbol{y}_{y+2}^T, S_{t+1} = j|S_t = i) \\ &= \sum_{j \in \mathcal{S}} P(S_{t+1} = j, y_{t+1}|S_t) P(\boldsymbol{y}_{t+2}^T|S_t = i, S_{t+1} = j, y_{t+1}) \\ &= \sum_{j \in \mathcal{S}} c(i; j, y_{t+1}) P(\boldsymbol{y}_{t+2}^T|S_{t+1} = j) \\ &= \sum_{j \in \mathcal{S}} c(i; j, y_{t+1}) \beta_{t+1}(j).\end{aligned}$$

where in deriving the next to the last line, we used the Markov property that if $S_{t+1}$ is known, knowledge of $S_t$ does not help in predicting $S_{t+2}$ and beyond. The state-based output model suggests that $y_{t+2}$ is a probabilistic function of $S_{t+2}$ alone. Thus, knowledge of $y_{t+1}$ does not provide any more information than knowledge of $S_{t+1}$.

**20.11 Backward algorithm.** As we dropped the arguments $\boldsymbol{y}_0^t$ from the forward variables, we delete the argument $\boldsymbol{y}_{t+1}^T$ in $\beta_t(i)$:

**20.12 Forward-backward formula** From the definitions of $\alpha_t(j)$ and $\beta_t(j)$,

$$\begin{aligned}\alpha_t(j)\beta_j(t) &= P(S_t = j, \boldsymbol{y}_0^t) P(\boldsymbol{y}_{t+1}^T|S_t = j) \\ &= P(S_t = j, \boldsymbol{y}_0^t) P(\boldsymbol{y}_{t+1}^T|S_t = j, \boldsymbol{y}_0^t) \\ &= P(S_t = j, \boldsymbol{y}_0^t, \boldsymbol{y}_{t+1}^T) = P(S_t = j, \boldsymbol{y}_0^T),\end{aligned}$$

where in obtaining the second line, we applied the argument used in the solution of the preceding exercise in the reverse direction, i.e., adding the information $\boldsymbol{y}_0^t$ to $S_t = j$ does not change the probability of the future evolution $\boldsymbol{y}_{t+1}^T$. By summing the above over $j \in \mathcal{S}$ we obtain (20.64).

**Algorithm 20.1 Backward Algorithm for $L(\boldsymbol{\theta})$**

1: Compute the backward variables recursively:

$$\beta_T(j) = 1, \quad j \in \mathcal{S},$$
$$\beta_t(i) = \sum_{j \in \mathcal{S}} c(i; j, y_{t+1}) \beta_{t+1}(j), \quad i \in \mathcal{S}, \quad t = T-1, T-2, \ldots, 1, 0.$$

2: Compute

$$L(\boldsymbol{\theta}) = \sum_{i \in \mathcal{S}} c(\emptyset; i, y_0) \beta_0(i).$$

**20.13 Forward-backward algorithm** See Algorithm 20.2. Note that evaluation of the likelihood function $L(\boldsymbol{\theta})$ can be done for any choice of $t$. For $t = T$, $L(\boldsymbol{\theta}) = \sum_{i \in \mathcal{S}} \alpha_T(i)$, and this reduces to Algorithm 20.1 of the text, and no backward variables are required. For $t = 0$, we have $L(\boldsymbol{\theta}) = \sum_{i \in \mathcal{S}} \alpha_0(i) \beta_0(i)$, hence the algorithm reduces to the backward algorithm obtained in Solution 20.5.

**Algorithm 20.2 Forward-Backward Algorithm for $L(\boldsymbol{\theta})$**

1: As we collect observations $y_t$, compute the forward variables recursively:

$$\alpha_0(i) = \pi_0(i) b(i; y_0), \quad i \in \mathcal{S},$$
$$\alpha_t(j) = \sum_{i \in \mathcal{S}} \alpha_{t-1}(i) c(i; j, y_t), \quad j \in \mathcal{S}, \quad t = 1, 2, \ldots, T.$$

2: After all observations $\boldsymbol{y}_0^T$ have been collected, compute the backward variables recursively:

$$\beta_T(j) = 1, \quad j \in \mathcal{S},$$
$$\beta_t(i) = \sum_{j \in \mathcal{S}} c(i; j, y_{t+1}) \beta_{t+1}(j), \quad i \in \mathcal{S}, \quad t = T-1, T-2, \ldots, 1, 0.$$

3: Compute

$$L(\boldsymbol{\theta}) = \sum_{i \in \mathcal{S}} \alpha_t(i) \beta_t(i), \quad \text{for any } t = 0, 1, \ldots, T.$$

## 20.4 Estimation Algorithms for State Sequence

**20.14* Viterbi algorithm**

$$
\begin{aligned}
\tilde{\alpha}_t(j) &= \max_{\boldsymbol{S}_0^{t-1}} P[\boldsymbol{S}_0^{t-1}, S_t = j, \boldsymbol{y}_0^t] = \max_{i\in\mathcal{S}} \max_{\boldsymbol{S}_0^{t-2}} P[\boldsymbol{S}_0^{t-2}, S_{t-1} = i, S_t = j, \boldsymbol{y}_0^{t-1}, y_t] \\
&= \max_{i\in\mathcal{S}} \max_{\boldsymbol{S}_0^{t-2}} P[\boldsymbol{S}_0^{t-2}, S_{t-1} = i, \boldsymbol{y}_0^{t-1}] P[S_t = j, y_t | \boldsymbol{S}_0^{t-2}, S_{t-1} = i, \boldsymbol{y}_0^{t-1}] \\
&= \max_{i\in\mathcal{S}} \max_{\boldsymbol{S}_0^{t-2}} P[\boldsymbol{S}_0^{t-2}, S_{t-1} = i, \boldsymbol{y}_0^{t-1}] P[S_t = j, y_t | S_{t-1} = i] \\
&= \max_{i\in\mathcal{S}} \left( \max_{\boldsymbol{S}_0^{t-2}} P[\boldsymbol{S}_0^{t-2}, S_{t-1} = i, \boldsymbol{y}_0^{t-1}] \right) P[S_t = j, y_t | S_{t-1} = i] \\
&= \max_{i\in\mathcal{S}} \{\tilde{\alpha}_{t-1}(i) c(i; j, y_t)\}.
\end{aligned}
$$

Note that in deriving the 3rd line, we used the defining property that the Markov proces $X_t = (S_t, Y_t)$ depends on only $S_{t-1}$, if it is an HMM.

**20.15 The Viterbi algorithm for a convolutional encoded sequence sent over a BSC.**

For a legitimate state transition $s_{t-1} = i \to s_t = j$ on the trellis diagram, $o_t = o_t^{(1)} o_t^{(2)} = k$ be the encoder output. Here $o_t^{(1)} o_t^{(2)}$ is the binary representation of the integer $k$. For example, if $k = 2$, then $o_t^{(1)} = 1$ and $o_t^{(2)} = 0$.

When we observe $y_t$ at the BSC output, $y_t^{(1)} \oplus o_t^{(1)} = 1$, if the first bit disagrees. Similarly if the second bits disagree, $y_t^{(2)} \oplus o_t^{(2)} = 1$. Then the Hamming distance

$$
h(y_t, o_t) = y_t^{(1)} \oplus o_t^{(1)} + y_t^{(2)} \oplus o_t^{(2)}
$$

represents the number of bits in disagreement and $2 - h(y_t, o_t)$ is the number of bit positions in agreement, where $h(y_t, o_t) = 0, 1$, or 2. Then, we can calculate, by generalizing (20.41), the joint conditional probabilities as

$$
c(i; j, y_t) = a_{ij} \epsilon^{d(y_t, o_t)} (1 - \epsilon)^{2 - d(y_t, o_t)}, \;\; i, j \in \mathcal{S}, \;\; y_t \in \mathcal{Y}.
$$

Then, by taking the logarithm

$$
d(i; j, y_t) = \log a_{i,j} + d(y_t, o_t) \log \epsilon + (2 - d(y_t, o_t)) \log(1 - \epsilon).
$$

If the error rate $\epsilon = 2^{-m} \ll 1$, it will be convenient to use the base 2 logarithm. We can have the following approximate expression, by assuming $a_{i,j} = 1/2$, and $\ln(1 - x) \approx x$,

$$
d(i, j; y_t) \approx -1 - d(y_t, o_t) m + (2 - d(y_t, o_t)) m \log_2 e.
$$

Then we apply Algorithm 20.2 with the initial condition to find the MLSE of $\hat{\boldsymbol{s}}_0^T$

$$
\vec{\alpha}_0(i) = \begin{cases} 0 & \text{for } \; i = 0 \\ -\infty & \text{for } \; i \neq 0, \end{cases},
$$

Once we determine the MLSE $\hat{\boldsymbol{s}}^*$, it is straightforward to find the MLSE of $I_t$, using the relation

$$
\hat{s}_t^* = \hat{I}_t^* - \hat{I}_{t-1}^*.
$$

**20.16 MAP state estimation versus maximum-likelihood state estimation.**

$$\boldsymbol{s}_0^{t\,*} = \arg\max_{\boldsymbol{s}_0^t} p(\boldsymbol{y}_0^t|\boldsymbol{s}_0^t) = \arg\max_{s_t}\max_{\boldsymbol{s}_0^{t-1}} p(\boldsymbol{y}_0^t|\boldsymbol{s}_0^{t-1}, s_t).$$

Define

$$\begin{aligned}
\tilde{\alpha}_t(i, \boldsymbol{y}_0^t) &= \max \boldsymbol{s}_0 Tt - 1 P[\boldsymbol{Y}_0^t|\boldsymbol{S}_0^{t-1}, S_t = j] = \max_{i\in\mathcal{S}}\max_{\boldsymbol{S}_0^{t-2}} P[\boldsymbol{Y}_0^{t-1}, Y_t|\boldsymbol{S}_0^{t-2}, S_{t-1} = i, S_t = j] \\
&= \max_{i\in\mathcal{S}}\max_{\boldsymbol{S}_0^{t-2}} P[\boldsymbol{Y}_0^{t-1}|\boldsymbol{S}_0^{t-2}, S_{t-1} = i] P[Y_t|\boldsymbol{S}_0^{t-2}, S_{t-1} = i, S_t = j, \boldsymbol{Y}_0^{t-1}] \\
&= \max_{i\in\mathcal{S}}\max_{\boldsymbol{S}_0^{t-2}} P[\boldsymbol{Y}_0^{t-1}|\boldsymbol{S}_0^{t-2}, S_{t-1} = i,] P[Y_t|S_{t-1} = i, S_t = j] \\
&= \max_{i\in\mathcal{S}}\left(\max_{\boldsymbol{S}_0^{t-2}} P[\boldsymbol{Y}_0^{t-1}|\boldsymbol{S}_0^{t-2}, S_{t-1} = i]\right) P[Y_t|S_{t-1} = i, S_t = j] \\
&= \max_{i\in\mathcal{S}}\{\tilde{\alpha}_{t-1}(i)\tilde{b}(i; j, Y_t)\}.
\end{aligned}$$

**Algorithm 20.3 Backward Algorithm for Maximum Likelihood Sequence Estimation**

1: Compute the backward variables recursively, starting with the known final value $S_T^*$:

$$\overrightarrow{\beta}_T(j) = \begin{cases} 0 & \text{for } j = S_T^* \\ -\infty & \text{for } j \neq S_T^*, \end{cases},$$

$$\overrightarrow{\beta}_t(i) = \max_{j\in\mathcal{S}}\left\{d(i; j, y_{t+1}) + \overrightarrow{\beta}_{t+1}(j)\right\},\quad i \in \mathcal{S},\quad t = T-1, T-2, \ldots, 1, 0.$$

Keep a pointer to the state $j^*$, which the surviving path leads to, i.e.,

$$j^* = \arg\max_{j\in\mathcal{S}}\left\{d(i; j, y_{t+1}) + \overrightarrow{\beta}_{t+1}(j)\right\}.$$

2: Find the surviving state at $t = 0$:

$$\arg\max_{i\in\mathcal{S}} \overrightarrow{\beta}_0(i) \triangleq S_0^*.$$

3: Starting from $S_0^*$, back-track the state sequence $S_1^*, \ldots, S_t^*, \ldots, S_{T-1}^*, S_T^*$, as the pointer to each surviving state indicates.

**20.17 Viterbi algorithm for an additive white gaussian channel.**

(a)

$$g(y_t|(i,j)) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{[y_t - x_t(i,j)]^2}{2\sigma^2}\right\}$$

$$\mu_t(i;j) = c_t(i; j, y_t) = a(i;j)g(y_t|(i,j)).$$

(b) Take the logarithm of $\mu_t(i;j)$:

$$\ln\mu_t(i;j) = \ln a(i;j) - \frac{[y_t - x_t(i;j)]^2}{2\sigma^2} - \ln(\sqrt{2\pi}\sigma) = -\frac{1}{2\sigma^2}\left\{[y_t - x_t(i;j)]^2 - 2\sigma^2\right\} + \text{const.}$$

Thus, (20.71) can be replaced by (20.131). Since $\pi_0(j) = a(\emptyset; j) = 0$ for $j \neq S_0^*$, the initial value of $\breve{a}_0(j) = \ln\pi_0(j) + \ln g(y_0|(\emptyset; j))$ is infinity for $j \neq S_0^*$. For $S_0 = S_0^*$, we set any finite value. So we can set, for instance

$$\breve{a}_0(S_0^*) = 0.$$

**20.18* Viterbi algorithm for a partial-response channel [199,200].**

(a) Since $a(i;j) = 1/2$ for all $(i,j)$, we can drop the term $\sigma \ln a(i;j)$ in the recursion. Furthermore, noting

$$(y_t - x_t)^2 = -2\left(y_t x_t - \frac{x_t^2}{2}\right) + y_t,$$

we can replace (20.131) by (20.133).

(b)

$$\breve{\alpha}_t(0) = \max\left\{\breve{\alpha}_{t-1}(0),\ \ \breve{\alpha}_{t-1}(1) - Ay_t - \frac{A^2}{2}\right\},$$
$$\breve{\alpha}_t(1) = \max\left\{\breve{\alpha}_{t-1}(0) + Ay_t - \frac{A^2}{2},\ \ \breve{\alpha}_{t-1}(1)\right\} \tag{9}$$

In the above procedure, if the left term in the parenthesis gives the maximum, then the survivor emanates from state $S_{t-1} = 0$, otherwise from $S_{t-1} = 1$.

**20.19 Backward version of the Viterbi algorithm.**

(a) By noting in analogous to the forward variable (20.70) we can rewrite the backward variable as

$$\beta_t(j) = P[\boldsymbol{y}_{t+1}^T | S_t = i] = \sum_{\boldsymbol{s}_{t+1}^T} p(\boldsymbol{s}_{t+1}^T, \boldsymbol{y}_{t+1}^T | S_t = i). \tag{10}$$

Then we define a backward auxiliary variable for the Viterbi algorithm by

$$\tilde{\beta}_t(i) = \max_{\boldsymbol{s}_{t+1}^T} p(\boldsymbol{s}_{t+1}^T, \boldsymbol{y}_{t+1}^T | S_t = i). \tag{11}$$

(b) Then we can find the following recursion:

$$\begin{aligned}
\tilde{\beta}_t(i) &= \max_{\boldsymbol{S}_{t+2}^T,\, j} P[\boldsymbol{S}_{t+2}^T, S_{t+1} = j, \boldsymbol{y}_{t+2}^T, y_{t+1} | S_t = i] \\
&= \max_{\boldsymbol{s}_{t+2}^T,\, j} p(s_{t+1} = j, y_{t+1} | S_t = i) p(\boldsymbol{s}_{t+2}^T, \boldsymbol{y}_{t+2}^T | S_t = i, S_{t+1} = j, y_{t+1}) \\
&= \max_{\boldsymbol{s}_{t+2}^T,\, j} \{c(i;j,y_{t+1}) p(\boldsymbol{s}_{t+2}^T, \boldsymbol{y}_{t+2}^T | S_{t+1} = j)\} \\
&= \max_{\boldsymbol{s}_{t+2}^T,\, j} \{c(i;j,y_{t+1}) \tilde{\beta}_{t+1}(j)\}.
\end{aligned} \tag{12}$$

(c) **Backward Viterbi Algorithm and its program:**

a. Starting with

$$\tilde{\beta}_T(j) = \delta(j, S_T^*),$$

calculate $\tilde{\beta}_t(i), i \in \mathcal{S}$ for $t = T-1, \ldots, 1, 0$, using the recursion (12). While computing the survivor's score $\tilde{\beta}_t(i)$, keep a pointer to the state $j^*$ to which the surviving path enters, i.e.,

$$j^* = \arg\max_{j\in\mathcal{S}} c(i;j,y_{t+1})\tilde{\beta}_{t+1}(j) \triangleq S^*_{t+1}. \tag{13}$$

b. Find the surviving state at $t = 0$, i.e.,

$$j^* = \arg\max_{i\in\mathcal{S}} \tilde{\beta}_0(i) \triangleq S^*_0. \tag{14}$$

c. Starting from this state $S^*_0$, the state sequence $S^*_0, S^*_1, \dots, S^*_t, \dots, S^*_{T-1}, S^*_T$ is back-tracked (to the forward direction in $t$) as the pointer to each surviving state indicates, where $S^*_T$ is the known terminal state defined by (**??**).

d. The state sequence

$$\boldsymbol{S}^* = S^*_0 S^*_1 S^*_2 \dots S^*_t \dots S^*_T$$

thus found maximizes the likelihood function (20.78), i.e.,

$$\boldsymbol{S}^* = \arg\max\{L(\boldsymbol{S})\} = \arg\max_{\boldsymbol{S}} p(\boldsymbol{y}|\boldsymbol{S})$$

**Algorithm 20.4 Backward Algorithm for Maximum Likelihood Sequence Estimation**

(d) 1: Compute the backward variables recursively, starting with the known final value $S^*_T$:

$$\overrightarrow{\beta}_T(j) = \begin{cases} 0 & \text{for } \ j = S^*_T \\ -\infty & \text{for } \ j \neq S^*_T, \end{cases},$$

$$\overrightarrow{\beta}_t(i) = \max_{j\in\mathcal{S}} \left\{ d(i;j,y_{t+1}) + \overrightarrow{\beta}_{t+1}(j) \right\}, \ \ i \in \mathcal{S}, \ \ t = T-1, T-2, \dots, 1, 0.$$

Keep a pointer to the state $j^*$, which the surviving path leads to, i.e.,

$$j^* = \arg\max_{j\in\mathcal{S}} \left\{ d(i;j,y_{t+1}) + \overrightarrow{\beta}_{t+1}(j) \right\}.$$

2: Find the surviving state at $t = 0$:

$$\arg\max_{i\in\mathcal{S}} \overrightarrow{\beta}_0(i) \triangleq S^*_0.$$

3: Starting from $S^*_0$, back-track the state sequence $S^*_1, \dots, S^*_t, \dots, S^*_{T-1}, S^*_T$, as the pointer to each surviving state indicates.

**20.20 Backward Viterbi algorithms for an AWGN channel and the partial-response channel.**

(a) Referring to the solution to Problem 20.4, we take the logarithm of $\mu_{t+1}(i;j)$:

$$\begin{aligned} \ln \mu_{t+1}(i;j) &= \ln a(i;j) - \frac{[y_{t+1} - x_{t+1}(i;j)]^2}{2\sigma^2} - \ln(\sqrt{2\pi}\sigma) \\ &= -\frac{1}{2\sigma^2}\left\{[y_{t+1} - x_{t+1}(i;j)]^2 - 2\sigma^2\right\} + \text{const.} \end{aligned}$$

Thus, (20.135) can be replaced by (20.136). Since $\pi_T(j) = 0$ for $j \neq S^*_0$, the initial value of $\breve{a}_T(j)$ is infinity for $j \neq S^*_T$. For $S_T = S^*_T$, we set any finite value. So we can set, for instance

$$\breve{a}_T(S^*_T) = 0.$$

(b) Since $a(i;j) = 1/2$ for all $(i,j)$, we can drop the term $\sigma \ln a(i;j)$ in the recursion. Furthermore, noting

$$(y_{t+1} - x_{t+1})^2 = -2\left(y_{t+1}x_{t+1} - \frac{x_{t+1}^2}{2}\right) + y_{t+1},$$

we can replace (20.131) by (20.137).

$$\begin{aligned}
\breve{\beta}_t(0) &= \max\left\{\breve{\beta}_{t+1}(0),\ \ \breve{\beta}_{t+1}(1) + Ay_{t+1} - \frac{A^2}{2}\right\}, \\
\breve{\alpha}_t(1) &= \max\left\{\breve{\beta}_{t+1}(0) - Ay_{t+1} - \frac{A^2}{2},\ \ \breve{\beta}_{t+1}(1)\right\}
\end{aligned} \tag{15}$$

In the above procedure, if the left term in the parenthesis gives the maximum, then the survivor enters state $S_{t+1} = 0$, otherwise from $S_{t-1} = 1$.

**20.21 Forward and backward variables of the Viterbi algorithm.** Similar to Problem 20.2, we have from the definitions of $\tilde{\alpha}_t(j)$ and $\tilde{\beta}_t(j)$,

$$\begin{aligned}
\tilde{\alpha}_t(j)\tilde{\beta}_j(t) &= \max_{\boldsymbol{S}_0^{t-1}} P[\boldsymbol{S}_0^{t-1}, S_t = j, \boldsymbol{y}_0^t] \max_{\boldsymbol{S}_{t+1}^T} P[\boldsymbol{S}_{t+1}^T, \boldsymbol{y}_{t+1}^T | S_t = j] \\
&= \max_{\boldsymbol{S}_0^{t-1}, \boldsymbol{S}_{t+1}^T} P[\boldsymbol{S}_0^{t-1}, S_t = j, \boldsymbol{y}_0^t] \max_{\boldsymbol{S}_{t+1}^T} P[\boldsymbol{S}_{t+1}^T, \boldsymbol{y}_{t+1}^T | \boldsymbol{S}_0^{t-1}, S_t = j, \boldsymbol{y}_0^t] \\
&= \max_{\boldsymbol{S}_0^{t-1}, \boldsymbol{S}_{t+1}^T} P[\boldsymbol{S}_0^{t-1}, S_t = j, \boldsymbol{S}_{t+1}^T, \boldsymbol{y}_0^t, \boldsymbol{y}_{t+1}^T] \\
&= \max_{\boldsymbol{S}_0^{t-1}, \boldsymbol{S}_{t+1}^T} P[\boldsymbol{S}_0^{t-1}, S_t = j, \boldsymbol{S}_{t+1}^T, \boldsymbol{y}_0^T],
\end{aligned}$$

where in obtaining the second line, we applied the argument that adding the information $\boldsymbol{S}_0^{t-1}$ and $\boldsymbol{y}_0^t$ to $S_t = j$ does not change the probability of the future evolution of $\boldsymbol{y}_{t+1}^T$ and $\boldsymbol{S}_{t+1}^T$.

## 20.5 The BCJR Algorithm

**20.22 APP $\xi_t(i,j|\boldsymbol{y})$ and joint probability $\sigma_t(i;j,\boldsymbol{y})$.**

(a)

$$\sum_{t=o}^{T} \xi_t(i,j\,|\boldsymbol{y})$$

is the expected number of transitions from state $i$ to state $j$.

(b)

$$\begin{aligned}\sigma_t(i,j,\boldsymbol{y}) &= P[S_t=i,S_{t+1},\boldsymbol{y}] = P[S_t,S_{t+1},\boldsymbol{y}_0^t,y_{t+1},\boldsymbol{y}_{t+2}^T]\\ &= P[S_t=i,\boldsymbol{y}_0^t]P[S_{t+1},y_{t+1},\boldsymbol{y}_{t+2}^T|S_t,\boldsymbol{y}_0^t]\\ &= P[S_t=i,\boldsymbol{y}_0^t]P[S_{t+1},y_{t+1}|S_t=i,\boldsymbol{y}_0^t]P[\boldsymbol{y}_{t+2}^T|S_t=i,S_{t+1}=j,\boldsymbol{y}_0^t,y_{t+1}]\\ &= P[S_t=i,\boldsymbol{y}_0^t]P[S_{t+1},y_{t+1}|S_t=i]P[\boldsymbol{y}_{t+2}^T|S_{t+1}=j]\\ &= \alpha_t(i)c_{t+1}(i;j,y_{t+1})\beta_{t+1}(j)\end{aligned}$$

c Since $\xi_t(i,j\,|\boldsymbol{y}) = \frac{\sigma_t(i;j,\boldsymbol{y})}{p(\boldsymbol{y})}$, it is apparent that (20.142) holds.

**20.23 Alternative derivation of the MAP estimate**

(a) From the hint (20.145), we have the following identity:

$$\begin{aligned}P[S_t=i,S_{t+1}=i|\boldsymbol{y}\,] &= P[S_t=i|\boldsymbol{y}\,] - \sum_{j\in\mathcal{S}\setminus i} P[S_t=i,S_{t+1}=j|\boldsymbol{y}\,]\\ &= P[S_{t+1}=i|\boldsymbol{y}\,] - \sum_{j\in\mathcal{S}\setminus i} P[S_t=j,S_{t+1}=i|\boldsymbol{y}\,].\end{aligned} \tag{16}$$

Then, from (21), (20.81) and (20.145), we obtain the backward recursion formula (20.143) for the variable $\gamma_t(i)$:

$$\gamma_t(i) = \gamma_{t+1}(i) + \sum_{j\in\mathcal{S}\setminus i}[\xi_t(i,j)-\xi_t(j,i)], \tag{17}$$

with the initial condition (20.144):

$$\gamma_T(i) = \frac{\alpha_T(i)}{p(\boldsymbol{y})} = \frac{\alpha_T(i)}{\sum_{i\in\mathcal{S}}\alpha_T(i)}. \tag{18}$$

(b) Once we obtain $\gamma_t(i)$ for all $i\in\mathcal{S}$ for $t=T-1,T-2,\dots,1,0$ using the backward recursion (20.143), then the MAP estimate can be expressed as

$$\hat{S}_t = \arg\max_{i\in\mathcal{S}}\gamma_t(i), \quad \text{for } t=T,T-1,\dots,1,0. \tag{19}$$

## 20.6 Maximum-Likelihood Estimation of Model Parameters

**20.24 Derivation of (20.85).**

$\gamma_t(i)$ is the conditional distribution of $S_t$ given the observations $\boldsymbol{y}$, as defined in (20.81):

$$\gamma_t(i) \triangleq P[S_t=i|\,\boldsymbol{y}],\ i\in\mathcal{S}. \tag{20}$$

$\xi_t(i,j)$ is the joint conditional distribution of $(S_t,S_{t+1})$ given the observations $\boldsymbol{y}$, as defined in (20.139):

$$\xi_t(i,j) \triangleq P[S_t=i,S_{t+1}=j|\boldsymbol{y}]. \tag{21}$$

Then by obtaining the marginal distribution with respect to the first or the second argument, it is apparent that

$$\sum_{i\in\mathcal{S}} \xi_t(i,j) = P[S_{t+1}=j|\boldsymbol{y}] = \gamma_{t+1}(j)$$

and

$$\sum_{j\in\mathcal{S}} \xi_t(i,j) = P[S_t=i|\boldsymbol{y}] = \gamma_t(i).$$

From the second formula, and the forward-backward variable formula for $\xi(i,j)$, we find

$$\gamma_t(i) = \frac{\alpha_t(i,\boldsymbol{y}_0^t)\beta_t(i;\boldsymbol{y}_{t+1}^T)}{p(\boldsymbol{y})}.$$

**20.25* Alternative derivation of the FBA for the transition-based HMM.**

(a) We begin with the general auxiliary function derived in (19.38) of Section 19.2.2

$$\boxed{Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) \triangleq E\left[\log p(\boldsymbol{S},\boldsymbol{y};\boldsymbol{\theta})|\,\boldsymbol{y};\boldsymbol{\theta}^{(p)}\right] = \sum_{\boldsymbol{s}} p(\boldsymbol{s}|\boldsymbol{y};\boldsymbol{\theta}^{(p)}) \log p(\boldsymbol{s},\boldsymbol{y};\boldsymbol{\theta}),} \tag{22}$$

where $\boldsymbol{\theta}^{(p)}$ is the $p$th estimate of the model parameters, $p=0,1,2,\ldots$.
By referring to $p(\boldsymbol{S},\boldsymbol{y};\boldsymbol{\theta})$ in the above expression for $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$, we find from (20.49)

$$p(\boldsymbol{S},\boldsymbol{y};\boldsymbol{\theta}) = \alpha(S_0,y_0)\prod_{t=1}^{T} c(S_{t-1};S_t,y_t). \tag{23}$$

Since each $c(S_{t-1};S_t,y_t)$ is equal to $c(i;j,k)$ for some $i,j\in\mathcal{S}$ and $k\in\mathcal{Y}$, we can write the above as

$$p(\boldsymbol{S},\boldsymbol{y};\boldsymbol{\theta}) = \alpha_0(S_0,y_0)\prod_{i,j\in\mathcal{S},k\in\mathcal{Y}} c(i;j,k)^{M(i;j,k)}, \tag{24}$$

where $M(i;j,k)$ is the number of times that $(s_t,s_{t+1},y_{t+1})=(i,j,k)$ is found in the sequence $(\boldsymbol{s},\boldsymbol{y})$. For each $t=1,2,\ldots,T$, $(s_t,s_{t+1},y_{t+1})$ belongs to one and only one of the possible triplets $(i,j,k)\in\mathcal{S}\times\mathcal{S}\times\mathcal{Y}$. Thus, the following identity must hold for any sequence $(\boldsymbol{s},\boldsymbol{y})$:

$$\sum_{i,j\in\mathcal{S},k\in\mathcal{Y}} M(i;j,k) = T.$$

Although we observe an instance $\boldsymbol{y}$, we cannot observe the associated instance $\boldsymbol{s}$, so we must treat this missing data as a RV. Hence, $M(i;j,k)$, being a function of $\boldsymbol{S}$, is also a RV and so is $p(\boldsymbol{S},\boldsymbol{y};\boldsymbol{\theta})$. Taking the logarithm of both sides yields

$$\log p(\boldsymbol{S},\boldsymbol{y};\boldsymbol{\theta}) = \log\alpha_0(S_0,y_0) + \sum_{i,j\in\mathcal{S},k\in\mathcal{Y}} M(i;j,k)\log c(i;j,k). \tag{25}$$

Thus, we can write $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(p)})$ as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = Q_0(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) + Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}), \tag{26}$$

where

$$Q_0(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = E\left[\log \alpha_0(S_0, y_0)|\boldsymbol{y};\boldsymbol{\theta}^{(p)}\right], \tag{27}$$

$$Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = \sum_{i,j\in\mathcal{S},k\in\mathcal{Y}} E\left[M(i;j,k)|\boldsymbol{y},\boldsymbol{\theta}^{(p)}\right]\log c(i;j,k). \tag{28}$$

(b) We denote the above conditional expectation of the random variable $M(i;j,k)$ as

$$E\left[M(i;j,k)|\boldsymbol{y},\boldsymbol{\theta}^{(p)}\right] \triangleq \overline{M}^{(p)}(i;j,k|\boldsymbol{y}). \tag{29}$$

By counting only those sequences in which $y_t = k$ for some $t$, we can rewrite the last expression by using the *forward and backward variables* that are obtained together with the updated model parameters:

$$\overline{M}^{(p)}(i;j,k|\boldsymbol{y}) = \frac{\sum_{t=1}^{T}\alpha_{t-1}^{(p)}(i,\boldsymbol{y}_0^{t-1})c^{(p)}(i;j,k)\beta_t^{(p)}(j;\boldsymbol{y}_{t+1}^T)\delta_{y_t,k}}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})}, \tag{30}$$

where $c^{(p)}(i;j,y_t)$ is the $p$th update of the conditional probability (20.15), and $\alpha_t^{(p)}(i,\boldsymbol{y}_0^t)$ and $\beta_t^{(p)}(i;\boldsymbol{y}_{t+1}^T)$ are the variables (20.54) and (20.60) computed under the assumption $\boldsymbol{\theta} = \boldsymbol{\theta}^{(p)}$; and $\delta_{y_t,k}$ is one for $y_t = k$, and is zero otherwise.
(30) can be derived as follows: We can write

$$\overline{M}^{(p)}(i;j,k|\boldsymbol{y}) = \frac{\sum_{t=1}^{T}\delta_{y_t,k}P[S_{t-1}=i,S_t=j,\boldsymbol{Y}_0^T=\boldsymbol{y}_0^T;\boldsymbol{\theta}^{(p)}]}{p(\boldsymbol{y})}.$$

By noting

$$\begin{aligned}
&P[S_{t-1}=i,S_t=j,\boldsymbol{Y}=\boldsymbol{y};\boldsymbol{\theta}^{(p)}] = P[S_{t-1}=i,\boldsymbol{Y}_0^{t-1}=\boldsymbol{y}_0^{t-1};\boldsymbol{\theta}^{(p)}] \\
&\qquad\cdot P[S_t=j,\boldsymbol{Y}_t^T|S_{t-1}=i,\boldsymbol{Y}_0^{t-1}=\boldsymbol{y}_0^{t-1};\boldsymbol{\theta}^{(p)}] \\
&= P[S_{t-1}=i,\boldsymbol{Y}_0^{t-1}=\boldsymbol{y}_0^{t-1};\boldsymbol{\theta}^{(p)}]P[S_t=j,Y_t=y_t|S_{t-1}=i,\boldsymbol{Y}_0^{t-1};\boldsymbol{\theta}^{(p)}] \\
&\qquad\cdot P[\boldsymbol{Y}_{t+1}^T|S_{t-1}=i,S_t=j,\boldsymbol{Y}_0^t=\boldsymbol{y}_0^t;\boldsymbol{\theta}^{(p)}] \\
&= \alpha_{t-1}^{(p)}(i,\boldsymbol{y}_0^{t-1})P[S_t=j,Y_t=y_t|S_{t-1}=i;\boldsymbol{\theta}^{(p)}]P[\boldsymbol{Y}_{t+1}^T=\boldsymbol{y}_{t+1}^T|S_t=j;\boldsymbol{\theta}^{(p)}] \\
&= \alpha_{t-1}^{(p)}(i,\boldsymbol{y}_0^{t-1})c^{(p)}(i;j,y_t)\beta_t^{(p)}(j;\boldsymbol{y}_{t+1}^T)
\end{aligned}$$

Thus, we obtain (30).
Similarly We can write

$$\overline{M}_0^{(p)}(j,k|\boldsymbol{y}) = \frac{\delta_{y_0,k}P[S_0=j,Y_0=k;\boldsymbol{\theta}^{(p)}]}{p(\boldsymbol{y})}.$$

By noting

$$P[S_0=j,Y_0=k;\boldsymbol{\theta}^{(p)}] = \alpha_0^{(p)}(j,y_0)\beta_0^{(p)}(j;\boldsymbol{y}_1^T),$$

we obtain (33) to be given below.

(c) **Maximization step:**
Since the model parameters are the joint probability and conditional joint probability distributions, they must satisfy constraints

$$\sum_{j\in\mathcal{S},k\in\mathcal{Y}} \alpha_0(j,k) = 1, \tag{31}$$

$$\sum_{j\in\mathcal{S},k\in\mathcal{Y}} c(i;j,k) = 1, \text{ for all } i \in \mathcal{S}. \tag{32}$$

We wish to find the value of $\boldsymbol{\theta}$ that maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ under the set of constraints (32). Maximization of $Q_0(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ can be found as follows. For a given $j \in \mathcal{S}, k \in \mathcal{Y}$, we denote by $M_0(j,k)$ the number of times that the initial state $(S_0, Y_0) = (j,k)$ occurs. Clearly $M_0(j,k)$ is a 1-0 random variable such that $\sum_{j\in\mathcal{S},k\in\mathcal{Y}} M_0(j,k) = 1$. We can write its conditional expectation, given $\boldsymbol{Y} = \boldsymbol{y}$ (see the second result in part (b).) as

$$\overline{M}_0^{(p)}(j,k|\boldsymbol{y}) = \begin{cases} \frac{\alpha_0^{(p)}(j,y_0)\beta_0^{(p)}(j;\boldsymbol{y}_1^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})}, & k = y_0 \\ 0, & k \neq y_0 \end{cases} \tag{33}$$

Thus, $Q_0(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ of (27) can be written as

$$Q_0(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = \sum_{j\in\mathcal{S}} \overline{M}_0^{(p)}(j,k|\boldsymbol{y}) \log \alpha_0(j,y_0). \tag{34}$$

Using the log sum inequality of (10.21), we find the above expression can be maximized when $\alpha_0(j,y_0) = \alpha_0^{(p+1)}(j,y_0)$, where

$$\alpha_0^{(p+1)}(j,y_0) = \frac{\overline{M}_0^{(p)}(j,k|\boldsymbol{y})}{\sum_{j\in\mathcal{S},k\in\mathcal{Y}} \overline{M}_0^{(p)}(j,k|\boldsymbol{y})} = \frac{\alpha_0^{(p)}(j,y_0)\beta_0^{(p)}(j;\boldsymbol{y}_1^T)}{\sum_{j\in\mathcal{S}} \alpha_0^{(p)}(j,y_0)\beta_0^{(p)}(j;\boldsymbol{y}_1^T)} \tag{35}$$

$$= \frac{\alpha_0^{(p)}(j,y_0)\beta_0^{(p)}(j;\boldsymbol{y}_1^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})}. \tag{36}$$

Maximization of $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ is equivalent to maximizing the following expression for each $i \in \mathcal{S}$.

$$\sum_{j\in\mathcal{S},k\in\mathcal{Y}} \overline{M}^{(p)}(i;j,k|\boldsymbol{y}) \log c(i;j,k), \quad i \in \mathcal{S}. \tag{37}$$

By using the log sum inequality (10.21) again, we find that the maximum of (37) can be achieved when

$$c(i;j,k) = \frac{\overline{M}^{(p)}(i;j,k|\boldsymbol{y})}{\sum_{j\in\mathcal{S},k\in\mathcal{Y}} \overline{M}^{(p)}(i;j,k|\boldsymbol{y})}, \text{ for all } j \in \mathcal{S}, k \in \mathcal{Y}. \tag{38}$$

By substituting (30) into (38), we obtain the following expression for the $(p+1)$st update of

**Algorithm 20.5 EM Algorithm for a transition-based HMM**

1: Set $p \longleftarrow 0$, and denote the initial estimate of the model parameters as $\boldsymbol{\alpha}_0^{(0)} = [\alpha_0^{(0)}(i, y_0),\ i \in \mathcal{S}]$ and $\boldsymbol{C}^{(0)}(y_0) = [c^{(0)}(i; j, k)\delta_{k,y_0};\ \ i, j \in \mathcal{S}, k \in \mathcal{Y}]$.

2: **Forward part of E-step:** Compute and save the forward vector variables $\boldsymbol{\alpha}_t^{(p)}$ recursively:

$$\boldsymbol{\alpha}^{(p)}{}_t^{\top} = \boldsymbol{\alpha}^{(p)}{}_{t-1}^{\top}\boldsymbol{C}^{(p)}(y_t),\ \ t = 1, 2, \ldots, T,$$

3: Compute the likelihood function: $L^{(p)} = \mathbf{1}^{\top}\boldsymbol{\alpha}_T^{(p)}$.

4: **Backward Part of E-step:** Compute the backward vector variables $\boldsymbol{\beta}_t^{(p)}$ recursively. Compute and accumulate the APPs $\xi_t^{(p)}(i, j|\boldsymbol{y})$ and $\gamma_t^{(p)}(i|\boldsymbol{y})$.

a. Set $\boldsymbol{\beta}_T^{(p)} = \mathbf{1}$, $\Xi^{(p)}(i, j, k) = 0$, and $\Gamma^{(p)}(i, k) = 0$, $i, j \in \mathcal{S}, k \in \mathcal{Y}$.

b. For $t = T-1, T-2, \ldots, 0$:

i.Compute $\boldsymbol{\beta}_t^{(p)} = \boldsymbol{C}^{(p)}(y_{t+1})\boldsymbol{\beta}_{t+1}^{(p)}$.

ii.Compute $\xi^{(p)}(i, j|\boldsymbol{y}) = \alpha_{t-1}^{(p)}(i)c^{(p)}(i; j, k)\beta_{t+1}^{(p)}(j)$ and add to $\Xi^{(p)}(i, j, k)$:

$$\Xi^{(p)}(i, j, k) \leftarrow \Xi^{(p)}(i, j, k) + \xi_t^{(p)}(i, j\boldsymbol{y})\delta_{k,y_t}.$$

iii.Compute $\gamma^{(p)}(i|\boldsymbol{y}) = \sum_{j\in\mathcal{S}} \xi^{(p)}(i, j|\boldsymbol{y})$ and add to $\Gamma^{(p)}(i, k)$:

$$\Gamma^{(p)}(i, k) \leftarrow \Gamma^{(p)}(i, k) + \gamma_t^{(p)}(i|\boldsymbol{y})\delta_{k,y_t},\ \ \text{for all}\ \ i \in \mathcal{S}, k \in \mathcal{Y}.$$

5: **M-step:** Update the model parameters:

$$\alpha_0^{(p+1)}(j) \leftarrow \frac{\alpha_0^{(p)}(j)\beta_0^{(p)}(j)}{L^{(p)}},\ \ \text{for all}\ \ j \in \mathcal{S}$$

$$c^{(p+1)}(i; j, k) \leftarrow \frac{\Xi^{(p)}(i, j, k)}{\Gamma^{(p)}(i, k)}\ \ \text{for all}\ \ i, j \in \mathcal{S},\ \ k \in \mathcal{Y}.$$

6: If any of the stopping conditions is met, stop the iteration and output the estimated $\boldsymbol{\alpha}_0^{(p+1)}$ and $\boldsymbol{C}^{(p+1)}$; else set $p \leftarrow p + 1$ and repeat the Steps 2 through 5.

---

the model parameter $c(i; j, k)$:

$$\begin{aligned} c^{(p+1)}(i; j, k) &= \frac{\sum_{t=1}^{T} \alpha_{t-1}^{(p)}(i, \boldsymbol{y}_0^{t-1})c^{(p)}(i; j, k)\beta_t^{(p)}(j; \boldsymbol{y}_{t+1}^T)\delta_{y_t,k}}{\sum_{j\in\mathcal{S}}\sum_{t=1}^{T} \alpha_{t-1}^{(p)}(i, \boldsymbol{y}_0^{t-1})c^{(p)}(i; j, y_t)\beta_t^{(p)}(j; \boldsymbol{y}_{t+1}^T)} \\ &= \frac{\sum_{t=1}^{T} \xi_{t-1}^{(p)}(i, j|\boldsymbol{y})\delta_{y_t,k}}{\sum_{j\in\mathcal{S}}\sum_{t=1}^{T} \xi_{t-1}^{(p)}(i, j|\boldsymbol{y})}, \end{aligned} \tag{39}$$

where we used the property (20.142) of the APP $\xi_t(i, j|\boldsymbol{y})$:

$$\boxed{\xi_t^{(p)}(i, j|\boldsymbol{y}) = \frac{\alpha_t^{(p)}(i, \boldsymbol{y}_0^t)c^{(p)}(i; j, y_{t+1})\beta_{t+1}^{(p)}(j; \boldsymbol{y}_{t+2}^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})},\ \ i, j \in \mathcal{S},\ \ t \in \mathcal{T},} \tag{40}$$

which is the APP of observing a transition $S_t = i \longrightarrow S_{t+1} = j$, given the observations $\boldsymbol{y}$ and the model parameter $\boldsymbol{\theta}^{(p)}$. We can relate $\xi_t^{(p)}(i, j|\boldsymbol{y})$ to the APP $\gamma_t(i|\boldsymbol{y})$ defined in (20.81):

$$\gamma_t^{(p)}(i|\boldsymbol{y}) = \sum_{j \in \mathcal{S}} \xi_t^{(p)}(i, j|\boldsymbol{y}) = \frac{\alpha_t^{(p)}(i, \boldsymbol{y}_0^t)\beta_t^{(p)}(i; \boldsymbol{y}_{t+1}^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})} \tag{41}$$

with $\gamma_T^{(p)}(i|\boldsymbol{y})$ given, from (20.85) and (20.63), by

$$\gamma_T^{(p)}(i|\boldsymbol{y}) = \frac{\alpha_T^{(p)}(i, \boldsymbol{y}_0^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})} = \frac{\alpha_T^{(p)}(i, \boldsymbol{y}_0^T)}{\sum_{i \in \mathcal{S}} \alpha_T^{(p)}(i, \boldsymbol{y}_0^T)}, \ i \in \mathcal{S}. \tag{42}$$

Thus, we can express $c^{(p+1)}(i; j, k)$ of (39), using (40) and (20.148), as

$$\boxed{c^{(p+1)}(i; j, k) = \frac{\sum_{t=1:\ y_t=k}^T \xi_{t-1}^{(p)}(i, j|\boldsymbol{y})}{\sum_{t=1}^T \gamma_{t-1}^{(p)}(i|\boldsymbol{y})},} \tag{43}$$

where $\sum_{t=1:\ y_t=k}^T$ in the numerator means summation with respect to $t$ for which $y_t = k$. Algorithm 20.5 implements the EM algorithm discussed above. The forward part of the E-step is the same as Algorithms 20.1 and 20.3, and we use the vector-matrix notation as before. The backward part is basically the same as Algorithm 20.3, as far as the computation of the backward vector variables $\boldsymbol{\beta}_t^{(p)}$ is concerned. However, we need to compute the APP variables $\xi_t(i, j|\boldsymbol{y})$ and $\gamma_t(i|\boldsymbol{y})$ as well, and sum them with respect to $t$. For this purpose we create arrays $\Xi(i, j, k)$ and $\Gamma(i, k)$, where

$$\Xi(i, j, k) = \sum_{t=1}^T \xi_t(i, j|\boldsymbol{y})\delta_{y_t=k} \tag{44}$$

$$\Gamma(i, k) = \sum_{t=1}^T \gamma_t(i|\boldsymbol{y})\delta_{y_t=k}. \tag{45}$$

For the parameter variables used in the algorithm, we explicitly show the superscript $^{(p)}$, although we suppress the observed data $\boldsymbol{y}$. If we do not need to keep all the computation results in the iterative procedure, we can overwrite the parameter values of the previous iteration and can suppress $(p)$.

**20.26* Alternative derivation of the Baum-Welch Algorithm.**
Then

$$\log p(\boldsymbol{S}, \boldsymbol{y}|\boldsymbol{\theta}) = \log \pi_0(S_0) + \sum_{i,j \in \mathcal{S}} M(i, j) \log a(i; j) + \sum_{j \in \mathcal{S}, k \in \mathcal{Y}} N(j, k) \log b(j; k). \tag{46}$$

Thus we can write $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ as

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = E[\log \pi_0(S_0)|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}] &+ \sum_{i,j \in \mathcal{S}} E\left[M(i, j)|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}\right] \log a(i; j) \\ &+ \sum_{j \in \mathcal{S}, k \in \mathcal{Y}} E\left[N(j, k)|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}\right] \log b(j; k). \end{aligned} \tag{47}$$

By denoting

$$E\left[M(i,j)|\boldsymbol{y},\boldsymbol{\theta}^{(p)}\right] \triangleq \bar{M}^{(p)}(i,j|\boldsymbol{y}),$$
$$E\left[N(j,k)|\boldsymbol{y},\boldsymbol{\theta}^{(p)}\right] \triangleq \bar{N}^{(p)}(j,k|\boldsymbol{y})$$

We can write the above expectations by using the forward and backward variables:

$$\bar{M}^{(p)}(i;j|\boldsymbol{y}) = \frac{\sum_{t=1}^{T} \alpha_{t-1}^{(p)}(i,\boldsymbol{y}_0^{t-1})a^{(p)}(i;j)\beta_t^{(p)}(j;\boldsymbol{y}_{t+1}^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})} \tag{48}$$

$$\bar{N}^{(p)}(j,k) = \frac{\sum_{t=0}^{T} \alpha_t^{(p)}(j,\boldsymbol{y}_0^t)\beta_t^{(p)}(j;\boldsymbol{y}_{t+1}^T)\delta(y_t,k)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})} \tag{49}$$

where $L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)}) = p^{(p)}(\boldsymbol{y})$. Note that the range of summation for $\bar{M}$ and $\bar{N}$ differ.
The first term of (47) can be written as

$$E[\log \pi_0(S_0)|\boldsymbol{y},\boldsymbol{\theta}^{(p)}] = \sum_{j\in\mathcal{S}} \log \pi_0(j)P[S_0 = j|\boldsymbol{y},\boldsymbol{\theta}^{(p)}].$$

By applying the equality condition for the log-sum inequality, we find that (47) can be maximized when we set the model parameters to the following values in the $(p+1)$st iteration:

$$\pi_0^{(p+1)}(j) = P[S_0 = j|\boldsymbol{y},\boldsymbol{\theta}^{(p)}] = \gamma_0^{(p)}(j|\boldsymbol{y}) = \frac{\alpha_0^{(p)}(j,y_0)\beta_0^{(p)}(j;\boldsymbol{y}_1^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})}$$

$$a^{(p+1)}(i;j) = \frac{\bar{M}^{(p)}(i,j|\boldsymbol{y})}{\sum_{j\in\mathcal{S}} \bar{M}^{(p)}(i,j|\boldsymbol{y})} \quad i,j\in\mathcal{S}$$

$$b^{(p+1)}(j;k) = \frac{\bar{N}^{(p)}(j,k|\boldsymbol{y})}{\sum_{k\in\mathcal{Y}} \bar{N}^{(p)}(j,k|\boldsymbol{y})}, \quad j\in\mathcal{S}.$$

The mean values $\bar{M}^{(p)}(i;j|\boldsymbol{y}$ of (48) and $\bar{N}^{(p)}(j,k)$ of (49) can be written in terms of the APP $\xi_t(i,j|\boldsymbol{y})$ of (20.142) and the APP $\gamma_t(i|\boldsymbol{y})$ of (20.85) as follows:

$$\bar{M}^{(p)}(i;j|\boldsymbol{y}) = \sum_{t=0}^{T} \xi_{t-1}^{(p)}(i,j|\boldsymbol{y}), \text{ and } \bar{N}^{(p)}(j,k) = \sum_{t=0}^{T} \gamma_t^{(p)}(j|\boldsymbol{y})b^{(p)}(j;k)\delta(y_t,k), \tag{50}$$

Thus, we have the expressions (20.105) as the M-step solution:

## 20.7 Application Example: Parameter Estimation in Mixture Distributions

**20.27 Derivation of (20.118).**
To maximize $Q_1$ of (20.117) with respect to $\pi_l$, we use the standard Lagrangian method and solve the following equation:

$$\frac{\partial}{\partial \pi_i}\left[\sum_{i\in\mathcal{S}}\sum_{t=0}^{T}(\log \pi_i)\gamma_t^{(p)}(i|y_t) + \lambda\left(\sum_{i\in\mathcal{S}}\pi_i - 1\right)\right] = 0, \tag{51}$$

where $\gamma_t^{(p)}(i|y_t) \triangleq P[S_t = i|y_t, \boldsymbol{\theta}^{(p)}]$. Thus, we have

$$\pi_i = -\lambda^{-1} \sum_{t=0}^{T} \gamma_t^{(p)}(i|y_t), \quad i \in \mathcal{S}. \tag{52}$$

By substituting these into the constraint $\sum_{i\in\mathcal{S}} \pi_i = 1$, we find $\lambda = -(T+1)$. Thus, the updated estimate of $\pi_i$ is

$$\pi_i^{(p+1)} = \frac{1}{T+1} \sum_{t=0}^{T} \gamma_t^{(p)}(i|y_t). \tag{53}$$

**20.28 Mixture of Gaussian distribution.**

By writing $\phi_i = (\mu_i, \sigma_i)$, we have

$$f_i(y; \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(y-\mu_i)^2}{2\sigma_i^2}\right). \tag{54}$$

The update formula for the probability distribution $\{\pi_i\}$ is given by (20.122), where the conditional probability in that expression can be computed for given $y_t$ and $\boldsymbol{\theta}^{(p)} \triangleq \{(\pi_i^{(p)}, \mu_i^{(p)}, \sigma_i^{(p)});\ i \in \mathcal{S}\}$, as

$$\gamma_t^{(p)}(i|y_t) = \pi_i^{(p)} \frac{\frac{1}{\sigma_i^{(p)}} \exp\left(-\frac{(y_t-\mu_i^{(p)})^2}{2{\sigma_i^{(p)}}^2}\right)}{\sum_{j\in\mathcal{S}} \frac{\pi_j^{(p)}}{\sigma_j^{(p)}} \exp\left(-\frac{(y_t-\mu_j^{(p)})^2}{2{\sigma_j^{(p)}}^2}\right)}, \quad i \in \mathcal{S} \tag{55}$$

By substituting

$$\log f_i(y; \mu_i, \sigma_i) = -\frac{(y-\mu_i)^2}{2\sigma_i^2} - \log\sigma_i - \frac{1}{2}\log(2\pi). \tag{56}$$

into (20.118), we have

$$Q_2(\boldsymbol{\phi}|\boldsymbol{\theta}^{(p)}) = \sum_{i\in\mathcal{S}} \sum_{t=0}^{T} \left[-\frac{(y_t-\mu_i)^2}{2\sigma_i^2} - \log\sigma_i - \frac{1}{2}\log(2\pi)\right] \gamma_t^{(p)}(i|y_t). \tag{57}$$

If we differentiate this with respect to $\mu_i$ and set it to zero, we find

$$\sigma_i^{-2} \sum_{t=0}^{T} (y_t - \mu_i)\gamma_t^{(p)}(i|y_t) = 0, \tag{58}$$

which we can solve for $\mu_i$. By denoting this solution as $\mu_i^{(p+1)}$, we have

$$\mu_i^{(p+1)} = \frac{\sum_{t=0}^{T} y_t \gamma_t^{(p)}(i|y_t)}{\sum_{t=0}^{T} \gamma_t^{(p)}(i|y_t)}. \tag{59}$$

Updating the estimate of $\sigma_i$ can be obtained in a similar fashion. Differentiate $Q_2$ with respect to $\sigma_i$

$$\frac{\partial}{\partial\sigma_i} Q_2(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = \sum_{i\in\mathcal{S}} \sum_{t=0}^{T} \left[\frac{(y_t-\mu_i)^2}{\sigma_i^3} - \frac{1}{\sigma_i}\right] \gamma_t^{(p)}(i|y_t) = 0,$$

**Algorithm 20.6 EM Algorithm for Gaussian distribution mixture**

1: Set the iteration number: $p \leftarrow 0$, and denote the initial estimate of the parameters: $\boldsymbol{\theta}^{(0)} = (\pi^{(0)}, \mu_i^{(0)}, \sigma_i^{(0)};\ i \in \mathcal{S})$. .

2: Estimate new parameters:

$$\pi_i^{(p+1)} = \frac{1}{n}\sum_{t=0}^{T}\gamma_t^{(p)}(i|y_t);$$
$$\mu_i^{(p+1)} = \frac{\sum_{t=0}^{T} y_t\gamma_t^{(p)}(i|y_t)}{\sum_{t=0}^{T}\gamma_t^{(p)}(i|y_t)};$$
$$\sigma_i^{2\,(p+1)} = \frac{\sum_{t=0}^{T} y_t^2\gamma_t^{(p)}(i|y_t)}{\sum_{t=0}^{T}\gamma_t^{(p)}(i|y_t)} - (\mu_i^{(p+1)})^2.$$

3: Replace $\boldsymbol{\theta}^{(p)}$ by $\boldsymbol{\theta}^{(p+1)}$, set $p \leftarrow p+1$, and repeat the Steps 2 through 4.

which gives

$$\sigma_i^2 = \frac{\sum_{t=0}^{T}\gamma_t^{(p)}(i|y_t)(y_t-\mu_i)^2}{\sum_{t=0}^{T}\gamma_t^{(p)}(i|y_t)}.$$

Since the $\mu$ in the RHS is unknown, we should substitute the best estimate, i.e., $\mu_i(p+1)$. So we obtain

$$\left(\sigma_i^{(p+1)}\right)^2 = \frac{\sum_{t=0}^{T} y_t^2\gamma_t^{(p)}(i|y_t)}{\sum_{t=0}^{T}\gamma_t^{(p)}(i|y_t)} - (\mu_i^{(p+1)})^2. \tag{60}$$

Equations (20.122), (59) and (60) perform both the E-step and the M-step simultaneously. We summarize the above results in Algorithm 20.6.

**20.29 Mixture of exponential family distributions.**

Consider canonical exponential family distributions:

$$f_i(y_t;\eta_i) = h_i(y_t)\exp\{\eta_i^\top \boldsymbol{T}_i(y_t) - A_i(\boldsymbol{\eta}_i)\},\ \ i \in \mathcal{S},\ \ t\imath\mathcal{T}.$$

where $\eta_i = \eta_i(\phi_i)$ is a transformed parameter of $\phi_i$ , $i \in \mathcal{S}$, where both $\phi_i$ and $\eta_i$ may be vector-valued. And let $\boldsymbol{\eta} = \{\eta_i;\ i \in \mathcal{S}\}$.

$$\log f_i(y_t;\eta_i) = \log h_i(y_t) + \eta_i^\top \boldsymbol{T}_i(y_t) - A_i(\eta_i).$$

Thus, $Q_2$-function takes the form

$$Q_2(\boldsymbol{\eta};\boldsymbol{\theta}^{(p)}) = \sum_{t=0}^{T}\left(\log h_i(y_t) + \eta_i^\top \boldsymbol{T}_i(y_t) - A_i(\eta_i)\right)\gamma_t^{(p)}(i|y_t).$$

Differentiate the above with respect to $\eta_i$ and set it to zero:

$$\nabla_{\eta_i}Q_2 = \sum_{t=0}^{T}\left(\boldsymbol{T}_i(y_t) - \nabla_{\eta_i}A_i(\eta_i)\right)\gamma_t^{(p)}(i|y_t) = \mathbf{0},\ \ i \in \mathcal{S},$$

from which we derive the update formula for $\eta_i$:

$$\nabla_{\eta_i} A_i(\eta_i^{(p+1)}) = \sum_{t=0}^{T} \boldsymbol{T}_i(y_t)\gamma_t^{(p)}(i|y_t). \tag{61}$$

For the normal distribution, we have

$$\eta_i^\top = [\eta_{i,1}, \eta_{i,2}] = \left[\frac{1}{\sigma_i^2} . \frac{\mu_I}{\sigma_i^2}\right], \;\; \boldsymbol{T}_i(y) = \begin{bmatrix} -\frac{y^2}{2} \\ y \end{bmatrix}$$

$$h_i(y) = \frac{1}{\sqrt{2\pi}}, \;\; A_i(\eta_i) = \frac{\mu_i^2}{2\sigma_i^2} + \log \sigma_i.$$

We can write

$$\mu_i = \frac{\eta_{i,2}}{\eta_{i,1}}, \;\text{ and }\; \sigma_i^2 = \frac{1}{\eta_1},$$

and

$$A_i(\eta_i) = \frac{\eta_{i,2}^2}{2\eta_{i,1}} - \frac{\log \eta_{i,1}}{2}.$$

Then from (61) we have

$$\left[-\frac{\eta_{i,2}^{(p+1)^2}}{2\eta_{i_1}^{(p+1)^2}}, \frac{\eta_{i,2}^{(p+1)}}{\eta_{i,1}^{(p+1)}}\right] = \sum_{t=0}^{T}\left[-\frac{y_t^2}{2}, y_t\right]\gamma_t^{(p)}(i|y_t).$$

Hence

$$\mu_i^{(p+1)^2} + \sigma_i^{(p+1)^2} = \sum_{t=0}^{T} y_t^2 \gamma_t^{(p)}(i|y_t)$$

$$\mu_i^{(p+1)} = \sum_{t=0}^{T} y_t \gamma_t^{(p)}(i|y_t).$$

Thus

$$\sigma_i^{(p+1)^2} = \sum_{t=0}^{T} y_t^2 \gamma_t^{(p)}(i|y_t) - \mu_i^{(p+1)^2}.$$

**20.30 Markov modulated Poisson process and Markov modulate Poisson sequence.**

(a) Transition rate matrix of the CTMC over the time interval $[0, \tau]$ is given by

$$\boldsymbol{P}(t) = [P_{ij}(t)] = [\exp(\boldsymbol{Q}t)]. \tag{62}$$

Therefore, the TPM $\boldsymbol{A}$ of the corresponding MMPS ought to be

$$\boldsymbol{A} = [a(i;j)] = \boldsymbol{P}(\Delta) = [\exp(\boldsymbol{Q}\Delta)]. \tag{63}$$

(b) The number of Poisson arrivals in the interval $\Delta$ is Poisson distributed with mean $\Delta\mu_i$ when $S(\tau) = i$. Hence it is apparent

$$\lambda_i = \mu_i \Delta. \tag{64}$$

**20.31 Traffic modeling based on MMPS.**

By writing instances of the observable packet traffic and the hidden state sequence by $\boldsymbol{y} = (y_0, y_1, \ldots, y_T)$ and $\boldsymbol{s} = (s_0, s_1, \ldots, s_T)$, we list below the probability functions that are needed to compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$:

$$p(\boldsymbol{s};\boldsymbol{\theta}) = \pi_0(s_0)\prod_{t=1}^{T} a(s_{t-1};s_t), \tag{65}$$

$$p(\boldsymbol{y}|\boldsymbol{s};\boldsymbol{\theta}) = \prod_{t\in\mathcal{T}} p(y_t|s_t;\boldsymbol{\theta}) = \prod_{t\in\mathcal{T}} \frac{\lambda_{s_t}^{y_t}}{y_t!}e^{-\lambda_{s_t}}, \tag{66}$$

$$\log p(\boldsymbol{s},\boldsymbol{y};\boldsymbol{\theta}) = \log\pi_0(s_0) + \sum_{t=1}^{T}\log a(s_{t-1};s_t) + \sum_{t\in\mathcal{T}}[y_t\log\lambda_{s_t} - \lambda_{s_t} - \log(y_t!)] \tag{67}$$

Then the $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ of (19.24) can be evaluated as

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) &= E\left[\log p(\boldsymbol{S},\boldsymbol{y};\boldsymbol{\theta})|\boldsymbol{y};\boldsymbol{\theta}^{(p)}\right] \\
&= E\left[\log\pi_0(S_0)|\boldsymbol{y};\boldsymbol{\theta}^{(p)}\right] + \sum_{t=1}^{T} E\left[\log a(S_{t-1};S_t)|\boldsymbol{y};\boldsymbol{\theta}^{(p)}\right] \\
&\quad + \sum_{t\in\mathcal{T}} E\left[(y_t\log\lambda_{S_t} - \lambda_{S_t})|\boldsymbol{y};\boldsymbol{\theta}^{(p)}\right] - \sum_{t\in\mathcal{T}} E\left[\log(y_t!)|\boldsymbol{y};\boldsymbol{\theta}^{(p)}\right] \\
&= Q_1(\boldsymbol{\pi}_0) + Q_2(\boldsymbol{A}) + Q_3(\boldsymbol{\lambda}) - Q_4
\end{aligned} \tag{68}$$

The 4th term $Q_4 = \sum_{t\in\mathcal{T}}\log(y_t!)$ does not contain any of the parameters of $\boldsymbol{\theta}$, so it has nothing to do with maximization of the $Q$-function with respect to $\boldsymbol{\theta}$. The first three terms contain different parameter groups, and they can be maximized separately.

1. **Maximization of** $Q_1(\boldsymbol{\pi}_0)$**:** We can rewrite $Q_1(\boldsymbol{\pi}_0)$ as

$$Q_1(\boldsymbol{\pi}_0) = \sum_{\boldsymbol{s}\in\mathcal{S}^{|\mathcal{T}|}} p(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{\theta}^{(p)})\log\pi_0(s_0) = \sum_{s_0\in\mathcal{S}} P[S_0 = s_0|\boldsymbol{y},\boldsymbol{\theta}^{(p)}]. \tag{69}$$

   We see the above expression is maximized by setting $\pi(i) = \pi^{(p+1)}(i),\ i\in\mathcal{S}$, where

$$\pi^{(p+1)}(i) = P[S_0 = i|\boldsymbol{y},\boldsymbol{\theta}^{(p)}] = \gamma_0(i|\boldsymbol{y},\boldsymbol{\theta}^{(p)}) \text{ for } i\in\mathcal{S}. \tag{70}$$

   $\gamma_0(i|\boldsymbol{y},\boldsymbol{\theta})$ is the APP (*a posteriori* probability) defined in (20.81) of Section 20.5.
   This part is essentially the same as the initial state probability update in the Baum-Welch algorithm. The computation of $\gamma_0(i|\boldsymbol{y},\boldsymbol{\theta}^{(p)})$ was already discussed in Algorithm 20.5 of Section 20.6.1 and Algorithm 20.5 in Section 20.6.2.

2. **Maximization of** $Q_2(\boldsymbol{A})$**:** Maximization of the $Q_2(\boldsymbol{A})$ can be done in the same way as we did in deriving the re-estimation formula of $a(i;j)$ in the Baum-Welch Algorithm. Thus, $Q_2(\boldsymbol{A})$ can be maximized when we set $\boldsymbol{A} = \boldsymbol{A}^{(p+1)} = [a^{(p+1)}(i;j)]$, where

$$a^{(p+1)}(i;j) = \frac{\sum_{t=1}^{T}\xi_{t-1}(i,j\,|\boldsymbol{y},\boldsymbol{\theta}^{(p)})}{\sum_{t=1}^{T}\gamma_{t-1}(i\,|\boldsymbol{y},\boldsymbol{\theta}^{(p)})},\ \ i,j\in\mathcal{S}; \tag{71}$$

where $\xi_t(i,j\,|\boldsymbol{y},\theta^{(p)})$ is defined in (20.83) and its computation algorithm is also a part of Algorithm 20.5.

**Algorithm 20.7 Baum-Welch Algorithm for a Markov modulated Poisson Sequence (MMPS) packet traffic model**

1: Denote the initial estimate of the model parameters as $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\pi}_0^{(0)}, \boldsymbol{A}^{(0)}, \boldsymbol{\lambda}^{(0)})$. Set the iteration number $p=0$.
2: Assume the $p$-th estimate of model parameters $\boldsymbol{\theta}^{(p)} = (\boldsymbol{\pi}_0^{(p)}, \boldsymbol{A}^{(p)}, \boldsymbol{\lambda}^{(p)})$.
3: **E-step:** Evaluate the forward and backward variables, $L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})$, and APPs:

$$\begin{aligned}
\alpha_0^{(p)}(i) &= \pi_0^{(p)} b^{(p)}(i;y_0),\ \ i\in\mathcal{S},\\
\alpha_t^{(p)}(j) &= \left(\sum_{i\in\mathcal{S}} \alpha_{t-1}^{(p)}(i)a^{(p)}(i;j)\right)\frac{\lambda_j^{(p)\,y_t}}{y_t!}e^{-\lambda_j^{(p)}},\ \ j\in\mathcal{S},\ \ t=1,2,\ldots,T,\\
\beta_T^{(p)}(j) &= 1,\ \ j\in\mathcal{S};\\
\beta_t^{(p)}(i) &= \sum_{j\in\mathcal{S}} a^{(p)}(i;j)\frac{\lambda_j^{(p)\,y_{t+1}}}{y_{t+1}!}e^{-\lambda_j^{(p)}}\beta_{t+1}^{(p)}(j),\ \ i\in\mathcal{S},\ \ t=T-1,T-2,\ldots,1,0,\\
L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)}) &= \sum_{i\in\mathcal{S}} \alpha_t^{(p)}(i)\beta_t^{(p)}(i),\ \ \text{for some}\ \ t\in\mathcal{T},\\
\xi_t^{(p)}(i,j) &= \frac{\alpha_t^{(p)}(i)a^{(p)}(i;j)\frac{\lambda_j^{(p)\,y_{t+1}}}{y_{t+1}!}e^{-\lambda_j^{(p)}}\beta_{t+1}^{(p)}(j)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})},\ \ i,j\in\mathcal{S},\ \ t\in\mathcal{T},\\
\gamma_t(i)^{(p)} &= \sum_{j\in\mathcal{S}} \xi_t^{(p)}(i,j),\ \ i\in\mathcal{S},\ \ t\in\mathcal{T}.
\end{aligned}$$

4: **M-step:** Update the model parameters:

$$\begin{aligned}
\pi_0^{(p+1)}(i) &= \gamma_0^{(p)}(i),\ \ i\in\mathcal{S},\\
a^{(p+1)}(i;j) &= \frac{\sum_{t=1}^T \xi_{t-1}^{(p)}(i,j)}{\sum_{t=1}^T \gamma_{t-1}^{(p)}(i)},\ \ i,j\in\mathcal{S},\\
\lambda^{(p+1)}(i) &= \frac{\sum_{t\in\mathcal{T}} y_t\gamma_t^{(p)}(i)}{\sum_{t\in\mathcal{T}}\gamma_t^{(p)}(i)},\ \ i\in\mathcal{S}.
\end{aligned}$$

5: Replace $\boldsymbol{\theta}^{(p)}$ by $\boldsymbol{\theta}^{(p+1)}$, set $p\leftarrow p+1$, and repeat the Steps 2 through 4.

3. **Maximization of $Q_3(\boldsymbol{\lambda})$:** We proceed in the same way as in the above. Since there is no constraints on $\lambda_i$'s except that $\lambda\geq 0$, we set

$$\begin{aligned}
Q_3(\boldsymbol{\lambda}) &= \sum_{t\in\mathcal{T}}\sum_{\boldsymbol{s}\in\mathcal{S}^{|\mathcal{T}|}} p(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{\theta}^{(p)})(y_t\log\lambda_{s_t}^{(p)}-\lambda_{s_t}^{(p)})\\
&= \sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{S}}\Big[P[S_t=i|\boldsymbol{y},\boldsymbol{\theta}^{(p)}](y_t\log\lambda_i-\lambda_i)\Big] \qquad (72)
\end{aligned}$$

By differentiating the above with respect to $\lambda_i$, and set it zero, we have

$$\frac{\partial Q_3(\boldsymbol{\lambda})}{\partial \lambda_i} = \sum_{t\in\mathcal{T}} \left[ P[S_t = i|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}] \left( \frac{y_t}{\lambda_i} - 1 \right) \right] = 0 \tag{73}$$

Thus, $Q_3(\boldsymbol{\lambda})$ is maximized when the updated Poisson rate parameters are $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{(p+1)}$, where

$$\begin{aligned} \lambda_i^{(p+1)} &= \frac{\sum_{t\in\mathcal{T}} y_t P[S_t = i|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}]}{\sum_{t\in\mathcal{T}} P[S_t = i|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}]} \\ &= \frac{\sum_{t\in\mathcal{T}} y_t \gamma_t(i|\boldsymbol{y}, \boldsymbol{\theta}^{(p)})}{\sum_{t\in\mathcal{T}} \gamma_t(i|\boldsymbol{y}, \boldsymbol{\theta}^{(p)})}, \;\; i \in \mathcal{S}. \end{aligned} \tag{74}$$

Algorithm 20.7 summarizes the above result.

**Remarks: An alternative derivation of (70) and (71).**
We can apply the Lagrangian multiplier method to obtain (70) and (71) as follow.
Define

$$J_1 = Q_1(\boldsymbol{\pi}_0) + \mu \left( \sum_{i\in\mathcal{S}} \pi_0(i) - 1 \right), \tag{75}$$

where $\mu$ is the Lagrangian multiplier. We differentiate the above with respect to $\pi_0(i)$ and set it to zero.

$$\frac{\partial J_1}{\partial \pi_0(i)} = \frac{\partial}{\partial \pi_0(i)} \left[ \sum_{i\in\mathcal{S}} p(\boldsymbol{y}, S_0 = i|\boldsymbol{\theta}^{(p)}) \log \pi_0(i) \right] + \mu = 0, \;\; i \in \mathcal{S}, \tag{76}$$

from which we obtain

$$\mu = -\sum_{i\in\mathcal{S}} p(\boldsymbol{y}, S_0 = i|\boldsymbol{\theta}^{(p)}) = -p(\boldsymbol{y}|\boldsymbol{\theta}^{(p)}). \tag{77}$$

Hence $Q_1(\boldsymbol{\pi}_0)$ is maximized when we choose $\boldsymbol{\pi}_0 = \boldsymbol{\pi}_0^{(p+1)} = [\boldsymbol{\pi}_0^{(p+1)}(i);\ i \in \mathcal{S}]$, where

$$\pi_0^{(p+1)}(i) = \frac{p(\boldsymbol{y}, S_0 = i|\boldsymbol{\theta}^{(p)})}{p(\boldsymbol{y}|\boldsymbol{\theta}^{(p)})} = P[S_0 = i|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}] = \gamma_0(i|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}), \tag{78}$$

where $\gamma_0(i|\boldsymbol{y}, \boldsymbol{\theta})$ was defined in (20.81) of Section 20.5.
Similarly, we set

$$J_2 = Q_2(\boldsymbol{A}) + \sum_{i\in\mathcal{S}} \mu_i \left( \sum_{j\in\mathcal{S}} a(i;j) - 1 \right). \tag{79}$$

Then differentiate the above with respect to $a(i;j)$ and set it to zero. By solving the Lagrangian multipliers $\mu_i$'s, we find $Q_2(\boldsymbol{A})$ is maximized by setting $\boldsymbol{A} = \boldsymbol{A}^{(p+1)} = [a^{(p+1)}(i;j)]$, where

$$\begin{aligned} a^{(p+1)}(i;j) &= \frac{\sum_{t=1}^T P[S_{t-1} = i, S_t = j, \boldsymbol{y}|\boldsymbol{\theta}^{(p)}]}{\sum_{t=1}^T P[S_{t-1} = i, \boldsymbol{y}|\boldsymbol{\theta}^{(p)})} \\ &= \frac{\sum_{t=1}^T \xi_{t-1}(i, j\,|\boldsymbol{y}, \boldsymbol{\theta}^{(p)})}{\sum_{t=1}^T \gamma_{t-1}(i\,|\boldsymbol{y}, \boldsymbol{\theta}^{(p)})}, \;\; i, j \in \mathcal{S}; \end{aligned} \tag{80}$$

where $\xi_t(i, j\,|\boldsymbol{y}, \theta^{(p)})$ is defined in (20.83).

**20.32 Markov modulated Bernoulli sequence (MMBS).**

$$p(\boldsymbol{y}|\boldsymbol{s},\boldsymbol{\theta}) = \prod_{t=0}^{T} p(y_t|s_t,\boldsymbol{\theta}) = \prod_{t=0}^{T} p(y_t|s_t,\boldsymbol{\theta}).$$

From (20.163) we can write

$$p(y_t|s_t,\boldsymbol{\theta}) = (1-b_{s_t})\,\delta(y_t,0) + b_{s_t}\delta(y_t,1),$$

where $\delta(x, y)$ is used as an indicator function, i.e.,

$$\delta(x,y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otheerwise.} \end{cases} \tag{81}$$

Then (68) remains the same, except that $Q_3(\boldsymbol{\lambda})$ should be replaced by $Q_3(\boldsymbol{b})$, where

$$Q_3(\boldsymbol{b}) = \sum_{t=0}^{T} \sum_{\boldsymbol{s}\in\mathcal{S}^{T+1}} p(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{\theta}^{(p)}) \log\left[(1-b_{s_t})\delta(y_t,0) + b_{s_t}\delta(y_t,1)\right]$$

$$\sum_{t=0}^{T} \sum_{i\in\mathcal{S}} p(i|\boldsymbol{y},\boldsymbol{\theta}^{(p)})\delta(s_t,i)\left(\log\left[(1-b_i)\delta(y_t,0) + b_i\delta(y_t,1)\right]\right\} \tag{82}$$

Since there is no constraint among $\boldsymbol{b}$'s except that $0 \leq b_i \leq 1$ for the individuals $b_i$'s, we differentiate the above with respect to $b_i$ and set it to zero:

$$\frac{\partial Q(\boldsymbol{b})}{\partial b_i} = \sum_{t=0}^{T} p(i|\boldsymbol{y},\boldsymbol{\theta}^{(p)})\delta(s_t,i)\left[\frac{\delta(y_t,1)-\delta(y_t,0)}{(1-b_i)\delta(y_t,0)+b_i\delta(y_t,1)}\right] = 0. \tag{83}$$

The term in the square bracket becomes $\frac{1}{b_i}$ when $y_t = 1$ and $-\frac{1}{1-b_i}$, when $y_t = 0$. Thus, by multiplying $b_i(1-b_i)$ to both sides, we have,

$$\sum_{t=0}^{T} P[S_t = i|\boldsymbol{y},\boldsymbol{\theta}^{(p)}][(1-b_i)\delta(y_t,1) - b_i\delta(y_t,0)] = 0. \tag{84}$$

or

$$P[S_t = i|\boldsymbol{y},\boldsymbol{\theta}^{(p)}]\delta(y_t,1) = b_i \sum_{t=0}^{T} P[S_t = i|\boldsymbol{y},\boldsymbol{\theta}^{(p)}].$$

Hence, $Q_3(\boldsymbol{b})$ is maximized when the probability $b_i$ of packet generation is set to $b_i^{(p+1)}$, where

$$b_i^{(p+1)} = \frac{\sum_{t=0}^{T} P[S_t = i|\boldsymbol{y},\boldsymbol{\theta}^{(p)}]\delta(y_t,1)}{\sum_{t=0}^{T} P[S_t = i|\boldsymbol{y},\boldsymbol{\theta}^{(p)}]}. \tag{85}$$

This result makes an intuitive sense, since the denominator is the expected number of times $t$ that such that $S_t = i$ under the model parameter $\boldsymbol{\theta}^{(p)}$, and the numerator is the expected number of times that $S_t = i$ and $Y_t = 1$. The updating formulas for $\pi^{(p+1)}(i)$ and $[a^{(p+1)}(i;j)]$ are the same as (70) and (71), respectively .

# 21 Solutions for Chapter 21: Elements of Machine Learning

**21.1 Markov chain as a DBN.** Since the probability of a sequence of states has the form

$$p(s_0, s_1, ..., s_T) = \prod_{i=1}^{T} p(s_t|s_{t-1}).$$

This probability has the form of equation (21.15), therefore, it is a BN in which the node corresponding to $s_{t-1}$ is a parent of node $s_t$.
The Markov blanket of the node corresponding to $s_t$ consists of nodes $s_{t-1}$ and $s_{t+1}$ because

$$p(s_t|\boldsymbol{s}_{-t}) = \frac{p(s_t|s_{t-1})p(s_{t+1}|s_t)}{p(s_{t+1}|s_{t-1})}$$

depends only on the values of the previous and next nodes.

**21.2 An HMM as a DBN** Consider a transition-based HMM depicted in Figure 21.6.

(a) To obtain the Markov blanket, we compute

$$\begin{aligned}
p(s_t|\boldsymbol{s}_{-t}, \boldsymbol{y}_0^T) &= \frac{p(\boldsymbol{s}_0^T, \boldsymbol{y}_0^T)}{p(\boldsymbol{s}_{-t}, \boldsymbol{y}_0^T)} \\
&= \frac{p(\boldsymbol{s}_0^{t-1}, \boldsymbol{y}_0^{t-1})p(s_t, \boldsymbol{y}_t|s_{t-1})p(s_{t+1}, \boldsymbol{y}_{t+1}|s_t)p(\boldsymbol{s}_{t+1}^T, \boldsymbol{s}_{t+1}^T)}{\sum_{s_t} p(\boldsymbol{s}_0^{t-1}, \boldsymbol{y}_0^{t-1})p(s_t, \boldsymbol{y}_t|s_{t-1})p(s_{t+1}, \boldsymbol{y}_{t+1}|s_t)p(\boldsymbol{s}_{t+1}^T, \boldsymbol{s}_{t+1}^T)} \\
&= \frac{p(s_t, \boldsymbol{y}_t|s_{t-1})p(s_{t+1}, \boldsymbol{y}_{t+1}|s_t)}{\sum_{s_t} p(s_t, \boldsymbol{y}_t|s_{t-1})p(s_{t+1}, \boldsymbol{y}_{t+1}|s_t)}.
\end{aligned}$$

This shows that $s_{t-1}, s_{t+1}, \boldsymbol{y}_t, \boldsymbol{y}_{t+1}$ is the Markov blanket of $s_t$

(b) Let us compute the likelihood function described in Section 20.3. To comply with notation of this section, we denote messages sent from left to right in Figure 21.7 as

$$\alpha_t(s_t, \boldsymbol{y}_0^t) = \mu_{f_t \to s_t}.$$

Then Eq. (21.30) takes the form

$$\mu_{s_t \to f_{t+1}} = \alpha_t(s_t, \boldsymbol{y}_0^t)$$

while (21.29) becomes

$$\alpha_{t+1}(s_{t+1}, \boldsymbol{y}_0^{t+1}) = \sum_{i \in \mathcal{S}} p(s_{t+1}, y_{t+1}|s_i)\alpha_t(s_i, \boldsymbol{y}_0^t).$$

This equation is equivalent to (20.55) describing the forward algorithm.

The backward algorithm is derived similarly by considering messages sent from right to left in Figure 21.7. The BCJR algorithm is obtained if we compute $p(s_t|\boldsymbol{y}_0^T)$ using the sum-product algorithm:

$$p(s_t|\boldsymbol{y}_0^T \propto \alpha_t(s_t, \boldsymbol{y}_0^t)\beta(s_t, \boldsymbol{y}_{t+1}^T).$$

(c) Since for nonnegative real-valued quantities $a$, $b$, and $c$, the maximization operator follows the distributive law

$$a \max(b, c) = \max(ab, ac),$$

we can replace the multiplication operator in the previous subproblem to obtain the Viterbi algorithm (20.71)

**21.3 Conditional probability of a Markov blanket.** Our derivation generalizes that of part (a) of Solution 21.2. We have

$$p(x_j|\boldsymbol{x}_{-j}) = \frac{p(\boldsymbol{x})}{\sum_{x_j} p(\boldsymbol{x})} = \frac{\prod_{v\in\mathcal{V}} p(x_v|\boldsymbol{x}_{\mathrm{pa}(v)})}{\sum_{x_j}\prod_{v\in\mathcal{V}} p(x_v|\boldsymbol{x}_{\mathrm{pa}(v)})}.$$

In the above expression, all the components that do not depend on $j$ cancel out in the nominator and denominatore, and we obtain

$$p(x_j|\boldsymbol{x}_{-j}) = \frac{p(x_j|\boldsymbol{x}_{\mathrm{pa}(j)}\prod_{v\in\mathrm{ch}(j)} p(x_v|\boldsymbol{x}_{\mathrm{pa}(v)})}{\sum_{x_j} p(x_j|\boldsymbol{x}_{\mathrm{pa}(j)}\prod_{v\in\mathrm{ch(j)}} p(x_v|\boldsymbol{x}_{\mathrm{pa}(v)})}.$$

It then follows that the Markov blanket of the node $j$ consists of its parents $\mathrm{pa}(j)$, its children $\mathrm{ch}(j)$, and other parents $\mathrm{pa}(v)$ of its children $v \in \mathrm{ch}(j)$. It also follows that this Markov blanket is minimal.

**21.4* Sum-product algorithm for a phylogenetic tree.**
The character $\chi$ of a phylogenetic tree can be represented in the form of a vector $\boldsymbol{w} = (w_i, i \in \tilde{V})$ such that

$$w_{\phi(\ell)} = \chi(\ell), \;\; \ell \in \mathcal{L}.$$

Define $\boldsymbol{X} = (X_j, j \in \mathcal{V})$ to be the vector of node variables of the tree, and let $\tilde{\boldsymbol{X}}$ denote the restriction of $\boldsymbol{X}$ to the node variables associated with the leaves of the tree, i.e., $\tilde{\boldsymbol{X}} = (X_u, u \in \tilde{\mathcal{V}})$. Then the probability that the character $\chi$ is realized by the phylogenetic tree can be expressed as

$$P_\chi = p_{\tilde{\boldsymbol{X}}}(\boldsymbol{w}) = P[\tilde{\boldsymbol{X}} = \boldsymbol{w}] = \sum_{\boldsymbol{x}:\tilde{\boldsymbol{x}}=\boldsymbol{w}} P[\boldsymbol{X} = \boldsymbol{x}], \tag{1}$$

where $P[\boldsymbol{X} = \boldsymbol{x}]$ is the joint probability distribution of all of the node variables, $X_i$, associated with the tree $\mathcal{T}$. Let us assume that the nodes in $\mathcal{V}$ are labeled as $0, 1, \ldots, |\mathcal{V}| - 1$, reflecting a total ordering of the nodes, with 0 denoting the root node. Then the probability $P[\boldsymbol{X} = \boldsymbol{x}]$ can be written as

$$
\begin{aligned}
P[\boldsymbol{X}=\boldsymbol{x}] &= P[X_0=x_0]\prod_{v\in\tilde{\mathcal{V}}\setminus\{0\}} P[X_v=x_v \big| X_u=x_u,\ u\le v] \\
&\overset{(a)}{=} \pi_{x_0}(0)\prod_{e=(u,v)\in\mathcal{E}} P_{x_u x_v}(e), \qquad (2)
\end{aligned}
$$

where the Markov property (16.68) is applied in step (a). Combining (1) and (2) we obtain

$$
P_\chi = \sum_{\boldsymbol{x}:\tilde{\boldsymbol{x}}=\boldsymbol{w}} \pi_{x_0}(0)\prod_{e=(u,v)\in\mathcal{E}} P_{x_u x_v}(e). \qquad (3)
$$

We now develop a sum-product algorithm to compute $P_\chi$ by recursively decomposing the tree $\mathcal{T}$ into its constituent subtrees. Corresponding to the subtree, $\mathcal{T}^u$, rooted at node $u$, we define the random vector $\boldsymbol{X}^u = (X_v, v\in\mathcal{V}^u)$. By restricting the components of $\boldsymbol{X}^u$ to the node variables corresponding to the leaves of the subtree $\mathcal{T}^u$, we define the random vector $\tilde{\boldsymbol{X}}^u = (X_v, v\in\tilde{\mathcal{V}}^u)$. By conditioning on $X_0$, (1) can be written as

$$
P_\chi = \sum_{i\in\mathcal{S}} P[\tilde{\boldsymbol{X}}=\boldsymbol{w}|X_0=i]P[X_0=i] = \sum_{i\in\mathcal{S}} P[\tilde{\boldsymbol{X}}=\boldsymbol{w}|X_0=i]\pi_i(0). \qquad (4)
$$

For an arbitrary node $u\in\mathcal{V}$, the conditional probability of $\{\tilde{\boldsymbol{X}}^u=\boldsymbol{w}^u\}$ given $\{X_u=i\}$ can be expressed as

$$
\begin{aligned}
P[\tilde{\boldsymbol{X}}^u=\boldsymbol{w}^u|X_u=i] &= P[\tilde{\boldsymbol{X}}^v=\boldsymbol{w}^v,\ v\in\mathrm{ch}(u)|X_v=i] \\
&= \prod_{v\in\mathrm{ch}(u)}\sum_{j\in\mathcal{S}} P[\tilde{\boldsymbol{X}}^v=\boldsymbol{w}^v, X_v=j|X_v=i] \\
&\overset{(a)}{=} \prod_{v\in\mathrm{ch}(u)}\sum_{j\in\mathcal{S}} P[\tilde{\boldsymbol{X}}^v=\boldsymbol{w}^v|X_v=i]P_{ij}((u,v)), \qquad (5)
\end{aligned}
$$

where the Markov property (16.68) was used in step (a). Conditioning on $X_0$ and applying (5) in (1), we have

$$
\begin{aligned}
P_\chi = P[\tilde{\boldsymbol{X}}=\boldsymbol{w}] &= \sum_{i\in\mathcal{S}} P[\tilde{\boldsymbol{X}}=\boldsymbol{w}|X_0=i]\pi_d(0) \\
&= \sum_{i\in\mathcal{S}}\pi_d(0)\prod_{v\in\mathrm{ch}(0)}\sum_{j\in\mathcal{S}} P[\tilde{\boldsymbol{X}}^v=\boldsymbol{w}^v|X_v=c]P_{ij}((0,v)). \qquad (6)
\end{aligned}
$$

By applying (5) recursively to (6), we obtain an efficient sum-product algorithm to compute $P_\chi$: We start at the leaves of $\mathcal{T}$ by applying (5) and work up to the root node 0, finally applying (6). Such an algorithm has a computational complexity that is linear in the number of nodes, $|\mathcal{V}|$, in the tree.

**Section 21.7: Markov Chain Monte Carlo (MCMC) Methods**

**21.5* Second-order Markov chains.** The second-order MC is defined by the TPM

$$
\boldsymbol{P} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}.
$$

The stationary distribution satisfies the following equations:

$$\begin{aligned} \pi_1 P_{11} + \pi_2 P_{21} &= \pi_1, \\ \pi_1 P_{12} + \pi_2 P_{22} &= \pi_2 \end{aligned} \tag{7}$$

To construct an ergodic MC whose stationary distribution is $\boldsymbol{\pi} = (\pi_1, \pi_2)$, we must solve for $P_{ij}$ these equations together with

$$\begin{aligned} P_{11} + P_{21} &= 1, \\ P_{12} + P_{22} &= 1 \end{aligned}$$

which is a system of three independent equations (since equations in (7) are linearly dependent) with four unknowns. If we denote $x = P_{12}$, we can write a general solution as

$$\boldsymbol{P} = \begin{bmatrix} 1-x & x \\ \frac{\pi_1}{\pi_2}x & 1-\frac{\pi_1}{\pi_2}x \end{bmatrix}$$

where $x$ is a free variable. It is clear that the matrix is stochastic and ergodic if

$$0 < x < 1 \text{ and } x < \frac{\pi_2}{\pi_1}$$

.

Thus, we conclude that infinitely many MCs with the TPMs of the form

$$\boldsymbol{P} = \begin{bmatrix} 1-x & x \\ \frac{\pi_1}{\pi_2}x & 1-\frac{\pi_1}{\pi_2}x \end{bmatrix} \tag{8}$$

have the same steady-state distribution $\boldsymbol{\pi} = (\pi_1, \pi_2)$, if

$$0 < x < \min\{1, \frac{\pi_2}{\pi_1}\}.$$

This inequality is equivalent to

$$0 < x < \begin{cases} \frac{\pi_2}{\pi_1} & \text{if } \pi_2 \le 0.5 \\ 1 & \text{otherwise} \end{cases}$$

Note that if $x = 0$, then

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Any distribution is a stationary distribution of this chain. The chain is reversible but not ergodic (because it is reducible). Similarly, if $x = 1$, then

$$\boldsymbol{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

The chain has a unique stationary distribution $\boldsymbol{\pi} = (0.5, 0.5)$. The chain is reversible, but not ergodic (because it is periodic). Thus, we see that not every reversible MC can be used for MCMC. The chain must be ergodic.

**21.6 Non-reversible MC.**

(a) Both matrices are doubly-stochastic (i.e. their transpose are also stochastic). Thus, their stationary distributions are uniform because $\mathbf{1}^\top \boldsymbol{P} = \mathbf{1}^\top$. Since both TPMs are symmetrical, they represent reversible MCs. The product

$$\boldsymbol{P}_1 \boldsymbol{P}_2 = \begin{bmatrix} 0.43 & 0.27 & 0.3 \\ 0.32 & 0.4 & 0.28 \\ 0.25 & 0.33 & 0.42 \end{bmatrix} \tag{9}$$

represents but still has the uniform stationary distribution but is not a symmetric matrix. Thus, this the corresponding MC is not reversible.

(b) To compare the convergence rates, we find the eigenvalues of the matrices:

$$\begin{aligned} \mathrm{Eig}(\boldsymbol{P}_1) &= (-0.6, & 0.4, && 1), \\ \mathrm{Eig}(\boldsymbol{P}_2) &= (-0.2646, & 0.2646, && 1), \\ \mathrm{Eig}(\boldsymbol{P}_1\boldsymbol{P}_2) &= (0.1250 + 0.0343i, & 0.1250 - 0.0343i, && 1). \end{aligned} \tag{10}$$

Since the absolute values of the eigenvalues of the matrix $\boldsymbol{P}_1\boldsymbol{P}_2$, that are different from 1, are smaller than that of the other matrices, its powers tend to zero faster than that of the other matrices and, therefore, the state probability distribution of $\boldsymbol{P}_1\boldsymbol{P}_2$ also converges faster to the equilibrium uniform distribution (see Sec. 15.2.2).

**21.7* Stationary distribution in the block MH algorithm.**

$$\int_{\boldsymbol{x}_1}\int_{\boldsymbol{x}_2} f_1(\boldsymbol{x}_1;\boldsymbol{y}_1|\boldsymbol{x}_2) f_2(\boldsymbol{x}_2;\boldsymbol{y}_2|\boldsymbol{y}_1)\pi(\boldsymbol{x}_1,\boldsymbol{x}_2)\,d\boldsymbol{x}_1\,d\boldsymbol{x}_2$$

$$= \int_{\boldsymbol{x}_2} f_2(\boldsymbol{x}_2;\boldsymbol{y}_2|\boldsymbol{y}_1)\pi_2(\boldsymbol{x}_2)\,d\boldsymbol{x}_2 \left(\int_{\boldsymbol{x}_1} f_1(\boldsymbol{x}_1;\boldsymbol{y}_1|\boldsymbol{x}_2)\pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\,d\boldsymbol{x}_1\right) \tag{11}$$

$$= \int_{\boldsymbol{x}_2} f_2(\boldsymbol{x}_2;\boldsymbol{y}_2|\boldsymbol{y}_1)\pi_2(\boldsymbol{x}_2)\pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\,d\boldsymbol{x}_2 \tag{12}$$

$$= \int_{\boldsymbol{x}_2} f_2(\boldsymbol{x}_2;\boldsymbol{y}_2|\boldsymbol{y}_1)\pi(\boldsymbol{y}_1,\boldsymbol{x}_2)\,d\boldsymbol{x}_2 \tag{13}$$

$$= \pi_1(\boldsymbol{y}_1)\int_{\boldsymbol{x}_2} f_2(\boldsymbol{x}_2;\boldsymbol{y}_2|\boldsymbol{y}_1)\pi_2(\boldsymbol{x}_2|\boldsymbol{y}_1)\,d\boldsymbol{x}_2 \tag{14}$$

$$= \pi_1(\boldsymbol{y}_1)\pi_2(\boldsymbol{x}_2|\boldsymbol{y}_1) \tag{15}$$

$$= \pi(\boldsymbol{y}_1,\boldsymbol{y}_2). \tag{16}$$

where equations (11), (13), (14) and (16) follow from Bayes' rule, while (12) and (15) follow from (21.50) and (21.51), respectively.

**21.8* Stationary distribution in the Gibbs sampler.**

$$f(\boldsymbol{x};\boldsymbol{y}) = \pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\pi_2(\boldsymbol{y}_2|\boldsymbol{y}_1), \text{ where } \boldsymbol{x} = (\boldsymbol{x}_1,\boldsymbol{x}_2),\ \ \boldsymbol{y} = (\boldsymbol{y}_1, by_2)$$

$$
\begin{aligned}
\int_{\boldsymbol{x}} \pi(\boldsymbol{x}) f(\boldsymbol{x};\boldsymbol{y}) d\boldsymbol{x} &= \int_{\boldsymbol{x}_1}\int_{\boldsymbol{x}_2} \pi(\boldsymbol{x}_1, bx_2)\pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\pi_2(\boldsymbol{y}_2|\boldsymbol{y}_1) d\boldsymbol{x}_1 dx_2 \\
&= \int_{\boldsymbol{x}_1}\int_{\boldsymbol{x}_2} \pi_1(\boldsymbol{x}_1|bx_2)\pi_2(\boldsymbol{x}_2)\pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\pi_2(\boldsymbol{y}_2|\boldsymbol{y}_1) d\boldsymbol{x}_1 dx_2 \\
&= \int_{\boldsymbol{x}_2} \pi_2(\boldsymbol{x}_2)\pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\pi(\boldsymbol{y}_2|\boldsymbol{y}_1) d\boldsymbol{x}_2 \\
&= \int_{\boldsymbol{x}_2} \pi(\boldsymbol{y}_1, \boldsymbol{x}_2)\pi(\boldsymbol{y}_2|\boldsymbol{y}_1) d\boldsymbol{x}_2 \\
&= \pi_1(\boldsymbol{y}_1)\pi_2(\boldsymbol{y}_2|\boldsymbol{y}_1) = \pi(\boldsymbol{y}_1, \boldsymbol{y}_2) \\
&= \pi(\boldsymbol{y}).
\end{aligned}
$$

**21.9 Gibbs sampler for multidimensional normal distribution.** Suppose that $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})$ has a multivariate normal distribution

$$
f_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{1}{2\pi|\det \boldsymbol{C}_z|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu}_z)^\top \boldsymbol{C}_z^{-1}(\boldsymbol{z}-\boldsymbol{\mu}_z)\right\}, \tag{17}
$$

where

$$
\boldsymbol{\mu}_z = E[\boldsymbol{Z}] = \begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}
$$

and

$$
\boldsymbol{C}_z = E[(\boldsymbol{Z}-\boldsymbol{\mu}_z)(\boldsymbol{Z}-\boldsymbol{\mu}_z)^\top] = \begin{bmatrix} \boldsymbol{C}_{xx} & \boldsymbol{C}_{xy} \\ \boldsymbol{C}_{yz} & \boldsymbol{C}_{yy} \end{bmatrix}. \tag{18}
$$

Thenthe conditional distributions $f_{\boldsymbol{Y}|\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{x})$ and $f_{\boldsymbol{X}|\boldsymbol{y}}(\boldsymbol{x}|\boldsymbol{y})$ are also Gaussian with means

$$
\boldsymbol{\mu}_{y|x} = \boldsymbol{\mu}_y + \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_x),\ \boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \boldsymbol{C}_{xy}\boldsymbol{C}_{yy}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_y) \tag{19}
$$

and variance matrices

$$
\boldsymbol{C}_{y|x} = \boldsymbol{C}_{yy} - \boldsymbol{C}_{yx}\boldsymbol{C}_{xx}^{-1}\boldsymbol{C}_{xy},\ \boldsymbol{C}_{x|y} = \boldsymbol{C}_{xx} - \boldsymbol{C}_{xy}\boldsymbol{C}_{yy}^{-1}\boldsymbol{C}_{yx}. \tag{20}
$$

Thus, we can generate a normal RV $\boldsymbol{Z}$ using lower dimension normal RVs $\boldsymbol{X}$ and $\boldsymbol{Y}$ using the Gibbs sampler. Each component vector $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be generated by splitting them into lower dimension variables. In particular, if $\boldsymbol{Z}$ is a $2n$-dimensional vector, we can apply the Box-Muller method (see Section 5.4.3) to generate $n$ two-dimensional Gaussian RVs.

# 22 Solutions for Chapter 22: Filtering and Prediction of Random Processes

## 22.1 Conditional Expectation and MMSE Estimation

**22.1 The condition for MSE.** If the correlation matrix $\boldsymbol{r_X}$ is positive definite, hence is invertible, then

$$\mathcal{E}(\boldsymbol{a}) = R_{ss} - 2\boldsymbol{a}^\top R_{s\boldsymbol{x}} + \boldsymbol{a}^\top \boldsymbol{R_{xx}}\boldsymbol{a} = (\boldsymbol{a} - \boldsymbol{R}_{\boldsymbol{xx}}^{-1} R_{s\boldsymbol{x}})^\top \boldsymbol{R_{xx}}(\boldsymbol{a} - \boldsymbol{R}_{\boldsymbol{xx}}^{-1} R_{s\boldsymbol{x}}) + R_{ss} - R_{s\boldsymbol{x}}^\top \boldsymbol{R}_{\boldsymbol{xx}}^{-1} R_{s\boldsymbol{x}}.$$

Then it is clear that $\mathcal{E}(\boldsymbol{a})$ takes its minimum at $\boldsymbol{a} = \boldsymbol{R}_{\boldsymbol{xx}}^{-1} R_{s\boldsymbol{x}}$.

Alternatively, we differentiate $\mathcal{E}(\boldsymbol{a})$ and set it to zero:

$$\nabla_{\boldsymbol{a}}\mathcal{E}(\boldsymbol{a}) = \frac{\partial \mathcal{E}(\boldsymbol{a})}{\partial \boldsymbol{a}} = -2R_{s\boldsymbol{x}} + 2\boldsymbol{R_{xx}}\boldsymbol{a} = \boldsymbol{0},$$

Taking the second derivative, we obtain the Hessian matrix

$$H(\mathcal{E}) = \nabla_{\boldsymbol{a}}\nabla_{\boldsymbol{a}}^\top \mathcal{E}(\boldsymbol{a}) = \frac{\partial^2 \mathcal{E}(\boldsymbol{a})}{\partial \boldsymbol{a}\boldsymbol{a}^\top} = 2\boldsymbol{R_{xx}},$$

which is positive definite. Hence the stationary point $\boldsymbol{a}$ gives the minimum.

**22.2 Linear MMSE condition and the orthogonality.**
Since $\boldsymbol{R_{xx}} = E[\boldsymbol{X}\boldsymbol{X}^\top]$ and $R_{s\boldsymbol{x}} = E[S\boldsymbol{X}]$, (22.8) becomes

$$E[\boldsymbol{X}\boldsymbol{X}^\top]\boldsymbol{a} - E[S\boldsymbol{X}] = \boldsymbol{0}.$$

The first term can be written as

$$E[\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{a})] = E[\boldsymbol{X}(\boldsymbol{a}^\top \boldsymbol{X})] = E[\hat{S}\boldsymbol{X}],$$

Because $\hat{S} = \boldsymbol{a}^\top \boldsymbol{X} = \boldsymbol{X}^\top \boldsymbol{a}$ is a scalar. Thus, we have

$$E[\hat{S}\boldsymbol{X}] - E[S\boldsymbol{X}] = E[(\hat{S} - S)\boldsymbol{X}] = \boldsymbol{0},$$

which is equivalent to (22.12).

(22.18) can be written as

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{X}]\boldsymbol{a} - \mathrm{Cov}[S,\boldsymbol{X}] &= E\left[(\boldsymbol{X}-E[\boldsymbol{X}])(\boldsymbol{X}-E[\boldsymbol{X}])^\top\right]\boldsymbol{a} - E\left[(S-E[S])(\boldsymbol{X}-E[\boldsymbol{X}])\right] \\
&= E\left[\boldsymbol{a}^\top(\boldsymbol{X}-E[\boldsymbol{X}])(\boldsymbol{X}-E[\boldsymbol{X}])\right] - E\left[(S-E[S])(\boldsymbol{X}-E[\boldsymbol{X}])\right] \\
&= E\left[(\hat{S}-E[\hat{S}])(\boldsymbol{X}-E[\boldsymbol{X}])\right] - E\left[(S-E[S])(\boldsymbol{X}-E[\boldsymbol{X}])\right] \\
&= E\left[(\hat{S}-S-\{E[\hat{S}]-E[S]\})(\boldsymbol{X}-E[\boldsymbol{X}])\right] \\
&= \mathrm{Cov}[\hat{S}-S,\boldsymbol{X}].
\end{aligned}$$

So we have shown that (22.18) is equivalent to (22.21).

**22.3* Alternative proof of Lemma 22.1.**

The *law of iterated expectations* (or the *law of total expectation*) states: if $S$ is a RV such that $E[|S|]<\infty$, and $X$ is any RV, then

$$E[S] = E_x[E_{s|x}[S|X]], \tag{1}$$

where $E_{s|x}$ means the expectation with respect to the conditional probability of $S$ given $X$, and $E_x$ means the expectation with respect to the marginal probability of $X$.

(a) The proof of the above formula follows essentially the same step as in the proof of the lemma given in the text. You use the joint, conditional and marginal PDF of the RVs.

(b) Then by applying the above formula, we find

$$\begin{aligned}
\langle S-E[S|\boldsymbol{X}], h(\boldsymbol{X})\rangle &\triangleq E\left[(S-E[S|\boldsymbol{X}])h^*(\boldsymbol{X})\right] \\
&= E_x\left[\left(E_{s|x}(S-E[S|\boldsymbol{X}])\right)h^*(\boldsymbol{X})\right] = 0,
\end{aligned} \tag{2}$$

because the term in the parenthesis is zero for all $\boldsymbol{X}$:

$$E_{s|x}(S-E[S|\boldsymbol{X}]) = E_{s|x}(S-E[S|X]) = E_{s|x}[S|X] - E_{s|x}[S|X] = 0. \tag{3}$$

**22.4 Derivation of (22.18).**

$$\mathcal{E}(\boldsymbol{a}) = E\left[(S-a_0-\boldsymbol{a}^\top\boldsymbol{X})^2\right] = E\left[S^2+a_0^2+\boldsymbol{a}^\top\boldsymbol{R_{xx}}\boldsymbol{a}-2Sa_0+2a_0\boldsymbol{a}^\top\boldsymbol{X}-2S\boldsymbol{a}^\top\boldsymbol{X}\right].$$

Then its partial derivatives are

$$\begin{aligned}
\frac{\partial\mathcal{E}(\boldsymbol{a})}{\partial a_0} &= E\left[2a_0-2S+2\boldsymbol{a}^\top\boldsymbol{X}\right] = 0 \\
\frac{\partial\mathcal{E}(\boldsymbol{a})}{\partial\boldsymbol{a}} &= 2\boldsymbol{R_{xx}}\boldsymbol{a}+2a_0E[\boldsymbol{X}]-2R_{s\boldsymbol{x}} = \boldsymbol{0}.
\end{aligned}$$

Substituting

$$a_0E[\boldsymbol{X}] = E[S]E[\boldsymbol{X}] - \boldsymbol{a}^\top E[\boldsymbol{X}]E[\boldsymbol{X}] = E[S]E[\boldsymbol{X}] - E[\boldsymbol{X}]E[\boldsymbol{X}^\top]\boldsymbol{a}$$

into the second equation above, we find

$$(\boldsymbol{R_{xx}} - E[\boldsymbol{X}]E[\boldsymbol{X}^\top])\boldsymbol{a} = R_{s\boldsymbol{x}} - E[S]E[\boldsymbol{X}],$$

which implies (22.18).

**22.5 Alternative derivation of (22.18).**

(a) Since the linear estimate is now given by $\hat{S} = \tilde{\boldsymbol{\beta}}^\top \tilde{\boldsymbol{X}}$, we can write the MSE as

$$\mathcal{E}(\tilde{\boldsymbol{\beta}}) = E\left[(\hat{S} - S)^2\right] = E\left[(\tilde{\boldsymbol{\beta}}^\top \tilde{\boldsymbol{X}} - S)^2\right] = E[S^2] - 2\tilde{\boldsymbol{\beta}}^\top \boldsymbol{r}_{s\tilde{\boldsymbol{x}}} + \tilde{\boldsymbol{\beta}}^\top \boldsymbol{R}_{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}}\tilde{\boldsymbol{\beta}},$$

where

$$\boldsymbol{r}_{s\tilde{\boldsymbol{x}}} = E[S\tilde{\boldsymbol{X}}], \text{ and } \boldsymbol{R}_{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}} = E\left[\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^\top\right].$$

By differentiating $\mathcal{E}(\tilde{\boldsymbol{\beta}})$ by $\tilde{\boldsymbol{\beta}}$ and set it to zero, we obtain the result.

(b) Since we can write

$$\boldsymbol{R}_{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}} = \begin{bmatrix} 1 & E[\boldsymbol{X}^\top] \\ E[\boldsymbol{X}] & \boldsymbol{R}_{\boldsymbol{x}\boldsymbol{x}} \end{bmatrix}, \text{ and } \boldsymbol{R}_{s\tilde{\boldsymbol{x}}} = \begin{bmatrix} E[S] \\ R_{s\boldsymbol{x}} \end{bmatrix},$$

and substituting these into $\boldsymbol{R}_{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}}\tilde{\boldsymbol{\beta}} = \boldsymbol{r}_{s\tilde{\boldsymbol{x}}}$, and expanding it, we have

$$a_0 + \boldsymbol{a}^\top E[\boldsymbol{X}] = E[S], \text{ and } a_0 E[\boldsymbol{X}] + \boldsymbol{R}_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{a} = R_{s\boldsymbol{x}},$$

as expected. Then the rest of argument is the same as the solution to Problem 22.4.

**22.6 The MMSE estimate and the expectation.**

$$\mathcal{E} = E\left[(\hat{S} - S)^2\right].$$

Differentiate $\mathcal{E}$ with respect to $\hat{S}$ and set it to zero:

$$\frac{\partial \mathcal{E}}{\partial \hat{S}} = 2E[(\hat{S} - S)] = 0,$$

which leads to

$$\hat{S} = E[S].$$

**22.7 Example 22.1: Additive noise model.**

(a) Sum of the independent normal RVs $X = S + N$ is also normal. Hence $X$ is normally distributed according to $N(\mu_s + \mu_n, \sigma_s^2 + \sigma_n^2)$.

(b) By definition

$$\rho_{sx} = \frac{\mathrm{Cov}[S, X]}{\sigma_s \sigma_x}.$$

Since $\mathrm{Cov}[S, X] = E[(S - \mu_s)(X - \mu_x)] = E[(S - \mu_s)^2] = \sigma_s^2$, we have

$$\rho_{sx} = \frac{\sigma_s}{\sigma_x} = \sqrt{\frac{\sigma_s^2}{\sigma_s^2 + \sigma_n^2}}.$$

(c) From the result of Section 4.7.1 on "Bivariate normal distribution,",

$$E[U_2|U_1] = \rho U_1.$$

By substituting $U_i = \frac{X_i - \mu_i}{\sigma_i}, \;\; i = 1, 2$, we find

$$E\left[\frac{X_2 - \mu_2}{\sigma_2} | U_1\right] = \rho U_1.$$

Hence

$$E[X_2 - \mu_2 | X_1] = \rho \sigma_2 U_1.$$

Hence

$$E[X_2 | X_1] = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(X_1 - \mu_1).$$

Thus, for the problem at hand, we find

$$E[S|X] = \mu_s + \rho_{sx} \frac{\sigma_s}{\sigma_x}(X - \mu_s) = \mu_s + \left(\frac{\sigma_s}{\sigma_x}\right)^2 (X - \mu_s).$$

**22.8 Independent normal variables and their product.**

(a)

$$E[Z] = E[X]E[Y] = 1 \cdot (-1) = -1.$$

and

$$\text{Var}[Z] = E[Z^2] - (E[Z])^2 = E[X^2]E[Y^2] - 1 = 3 \times 3 - 1 = 8.$$

(b)

$$\begin{aligned}\text{Cov}[X, Z] &= E[XZ] - E[X]E[Z] = E[X^2]E[Y] + 1 \\ &= 3 \cdot (-1) + 1 = -2.\end{aligned}$$

(c)

$$\mathcal{E} = E[(a_0 + a_1 Z - X)^2].$$

$$\frac{\partial \mathcal{E}}{\partial a_0} = 0, \Longrightarrow a_0 + a_1 E[Z] - E[X] = 0, \Longrightarrow a_0 - a_1 - 1 = 0.$$

$$\frac{\partial \mathcal{E}}{\partial a_1} = 0, \Longrightarrow a_0 + a_1 E[Z^2] - E[XZ] = 0, \Longrightarrow a_0 + 9a_1 + 3 = 0.$$

Then, we have

$$a_0 = \frac{3}{5}, \;\; a_1 = -\frac{2}{5}.$$

Hence $\hat{X} = a_0 + a_1 Z = \frac{3-2Z}{5}$

$$\begin{aligned}\mathcal{E}_{\min} &= E[a_0^2 + a_1^2 Z^2 + X^2 - 2a_0 X - 2a_1 ZX + 2a_0 a_1 Z]\\ &= a_0^2 + 9a_1^2 + 3 - 2a_0 + 6a_1 - a_0 a_1\\ &= \frac{9}{25} + \frac{36}{25} + 3 - \frac{6}{5} - \frac{12}{5} + \frac{12}{25}\\ &= \frac{42}{25}\end{aligned}$$

(d) The MMSE estimate is the conditional expectation of $Z$ given $X$:

$$E[Z|X] = E[XY|X] = E[Y]X = -X.$$

Hence $\hat{Z} = \phi(X) = -X$.

**22.9 Correlation coefficient between the response variable and a predictor.**
The regression $M(X)$ is the projection of $Y$ on the space spanned by $X$, and $T(X)$ is another vector on the same plane. Then it is clear geometrically that the angle $\psi$ between the vectors $Y$ and $T(X)$ is equal to or greater than the angle $\theta$ between $Y$ and $M(X)$. Since the correlation coefficients are $\cos\psi$ and $\cos\theta$, respectively, it should follow that $|\rho_{T,Y}| = |\cos\psi| \leq |\cos\theta| = |\rho_{M,Y}|$.

Let $T(X)$ be adjusted by a scaling factor such that the vector $T(X)$ and $Y - T(X)$ are orthogonal (the correlation coefficient between $cT(x)$ and $Y$ is the same as the correlation coefficient between $T(X)$ and $Y$. Then

$$|\cos\psi| = \frac{\|T(X)\|}{\|Y\|} \quad \text{and} \quad |\cos\theta| = \frac{\|M(X)\|}{\|Y\|}. \tag{4}$$

Note that the vector $M(X) - T(X)$ is orthogonal to $Y - M(X)$. Thus

$$\|Y - T(X)\| \geq \|Y - M(X)\|, \tag{5}$$

because of the Pythagoras theorem:

$$\|Y - T(X)\| = \sqrt{\|Y - M(X)\|^2 + \|M(X) - T(X)\|^2} \geq \|Y - M(X)\|.$$

Then

$$\|T(X)\| \leq \|M(X)\|, \tag{6}$$

because of the Pythagoras theorem again:

$$\begin{aligned}\|T(X)\| &= \sqrt{\|Y\|^2 - \|Y - T(X)\|^2} = \sqrt{\|Y\|^2 - (\|Y - M(X)\|^2 + \|M(X) - T(X)\|^2)}\\ &\leq \sqrt{\|Y\|^2 - \|Y - M(X)\|^2} = \|M(X)\|.\end{aligned}$$

Then from (4) and (6) we have

$$|\cos\psi| \leq |\cos\theta|,$$

hence

$$|\rho_{Y,T}| \leq |\rho_{Y,S}|.$$

Thus (22.36) holds.

**Alternative proof:** An equivalent but somewhat more concise proof is as follows: According to equation (10.13), we can write

$$|\rho_{X,Y}| = \frac{|\langle X, Y\rangle|}{\|X\|\,\|Y\|}. \tag{7}$$

If $(Y - X) \perp X$, i.e $\langle X, Y - X\rangle = 0$, we have

$$\langle X, Y\rangle = \|X\|^2 \tag{8}$$

and equation (7) can be written as

$$|\rho_{X,Y}| = \frac{\|X\|}{\|Y\|}. \tag{9}$$

We will prove (22.36) using this formula.
By the orthogonality principle, $(Y - M(\boldsymbol{X})) \perp \boldsymbol{X}$. Thus, using equation (9), we can write

$$|\rho_{Y,M}| = \frac{\|M(\boldsymbol{X})\|}{\|Y\|}. \tag{10}$$

Let $T(\boldsymbol{X})$ be adjusted by a scaling factor such that the vector $T(\boldsymbol{X})$ and $Y - T(\boldsymbol{X})$ are orthogonal (the correlation coefficient between $cT(x)$ and $Y$ is the same as the correlation coefficient between $T(\boldsymbol{X})$ and $Y$.) In this case, we can write

$$|\rho_{Y,T}| = \frac{\|T(\boldsymbol{X})\|}{\|Y\|}. \tag{11}$$

Comparing (10) and (11), we conclude that to prove (22.36) we need to show that $\|M(\boldsymbol{X})\| \geq \|T(\boldsymbol{X})\|$.
By the Pythagoras theorem, we have

$$\|Y\|^2 = \|M(\boldsymbol{X})\|^2 + \|Y - M(\boldsymbol{X})\|^2,\ \|Y\|^2 = \|T(\boldsymbol{X})\|^2 + \|Y - T(\boldsymbol{X})\|^2,$$

and

$$\|Y - T(\boldsymbol{X})\|^2 = \|Y - M(\boldsymbol{X})\|^2 + \|T(\boldsymbol{X}) - M(\boldsymbol{X})\|^2.$$

From these equations we obtain

$$\|M(\boldsymbol{X})\|^2 = \|T(\boldsymbol{X})\|^2 + \|T(\boldsymbol{X}) - M(\boldsymbol{X})\|^2 \tag{12}$$

which proves that $\|M(\boldsymbol{X})\| \geq \|T(\boldsymbol{X})\|$.

**22.10 Analysis of variance (ANOVA) equation.** We can write

$$\begin{aligned} Y - E[Y] &= Y - M + M - E[Y] \\ &= (Y - M) + (M - E[M]) + (E[M] - E[Y]) \\ &= (Y - M) + (M - E[M]) \end{aligned} \tag{13}$$

where we use the law of total expectation (3.38):

$$E[M] = E[E[Y|X]] = E[Y].$$

The two components in (13) are orthogonal because

$$E[(Y-M)(M-E[M])] = E[YM] + E[ME[M]] - E[YE[M]] - E[M^2]$$
$$= (E[YM] - E[M^2]) + (E[YE[M]] - E[ME[M]) = 0 + 0,$$

where we use the following identities due to again the law of total expectation:

$$E[YM] = E_X[E_Y[Y|X]M] = E_X[M^2]$$
$$E[YE[M]] = E_X[E_Y[Y|X]E[M]] = E_X[ME[M]].$$

Thus, by taking the square and expectation of (13), we have

$$E[(Y-E[Y])^2] = E[(Y-M)^2] + E[(M-E[M])^2],$$

which is equivalent to (22.37).

**22.11 Optimal choice of regression coefficients.**

**Note:** There is a typo in (22.40). The last term (5th term) should be the Hermitian (or complex conjugate) of the 4th term, i.e. $\boldsymbol{c}_{\boldsymbol{xy}}^H \boldsymbol{\beta}^*$.
By setting the quadratic form

$$J = \boldsymbol{\beta}^\top \boldsymbol{C}_{\boldsymbol{xx}} \boldsymbol{\beta}^* - \boldsymbol{\beta}^\top \boldsymbol{c}_{\boldsymbol{x}y} - \boldsymbol{c}_{\boldsymbol{xy}}^H \boldsymbol{\beta}^*,$$

and taking the partial derivative of $\mathcal{E}$ with respect to $\boldsymbol{\beta}^\top$ and setting it $\mathbf{0}$, we find

$$\frac{\partial \mathcal{E}}{\partial \boldsymbol{\beta}^\top} = \frac{\partial J}{\partial \boldsymbol{\beta}^\top} = \boldsymbol{C}_{\boldsymbol{xx}} \boldsymbol{\beta}^* - \boldsymbol{c}_{\boldsymbol{x}y} = \mathbf{0}, \tag{14}$$

we find

$$\boldsymbol{C}_{\boldsymbol{xx}} \boldsymbol{\beta}^*_{\text{opt}} = \boldsymbol{c}_{\boldsymbol{x}y}.$$

Similarly, by differentiating $\mathcal{E}$ with respect to $b$,and setting it to zero, we have

$$\frac{\partial \mathcal{E}}{\partial b} = \frac{\partial b b^*}{\partial b} = b^* = 0,$$

Thus,

$$b_{\text{opt}} = 0.$$

It is not difficult to see that the last three terms of (22.40) are equivalent when $\boldsymbol{\beta} = \boldsymbol{\beta}_{\text{opt}}$, i.e.,

$$\boldsymbol{\beta}_{\text{opt}}^\top \boldsymbol{C}_{\boldsymbol{xx}} \boldsymbol{\beta}^*_{\text{opt}} = \boldsymbol{\beta}_{\text{opt}}^\top \boldsymbol{c}_{\boldsymbol{x}y} = \boldsymbol{c}_{\boldsymbol{xy}}^H \boldsymbol{\beta}^*_{\text{opt}} = \boldsymbol{c}_{bxy}^H \boldsymbol{C}_{\boldsymbol{xx}}^{-1} \boldsymbol{c}_{\boldsymbol{x}y}.$$

Thus, it follows that

$$\mathcal{E}_{\min} = \sigma_Y^2 - \boldsymbol{c}_{\boldsymbol{xy}}^H \boldsymbol{C}_{\boldsymbol{xx}}^{-1} \boldsymbol{c}_{\boldsymbol{x}y}.$$

**22.12 Equivalence of the regression and the solution from the orthogonality equation.**
If we apply the orthogonality principle, we find that the best estimate $T(\boldsymbol{x})$ should be such that $T(\boldsymbol{X})$ and $Y - T(\boldsymbol{X})$ should be orthogonal, i.e.,

$$\langle T(\boldsymbol{X}), Y - T(\boldsymbol{X}) \rangle = 0, \tag{15}$$

or

$$E[T(\boldsymbol{X})(Y - T(\boldsymbol{X}))^*] = 0.$$

By writing $T(\boldsymbol{X}) = \beta_0 + \beta^\top \boldsymbol{X}$, we have

$$E[(\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{X})(Y - \beta_0 - \boldsymbol{\beta}^\top \boldsymbol{X})^*] = 0,$$

or

$$E\left[\left\{\beta_0 + \boldsymbol{\beta}^\top(\boldsymbol{X} - E[\boldsymbol{X}]) + \boldsymbol{\beta}^\top E[\boldsymbol{X}]\right\}\left\{Y - E[Y] + E[Y] - \beta_0 - (\boldsymbol{X} - E[\boldsymbol{X}])^\top \boldsymbol{\beta} - E[\boldsymbol{X}]^\top \boldsymbol{\beta}\right\}^*\right] = 0,$$

By using $b$ defined by (23.41) we can rewrite the above as

$$-\beta_0 b^* + \boldsymbol{\beta}^\top(\boldsymbol{c}_{\boldsymbol{x}y} - \boldsymbol{C}_{\boldsymbol{x}\boldsymbol{x}}\beta^*) - \boldsymbol{\beta}^\top E[\boldsymbol{X}]b^* = 0,$$

which can be arranged as

$$\boldsymbol{\beta}^\top(\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{\beta}^* - \boldsymbol{c}_{\boldsymbol{x}y}) - (\beta_0 + \boldsymbol{\beta}^\top E[\boldsymbol{X}])b^* = 0,$$

or

$$\boldsymbol{\beta}^\top(\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{\beta}^* - \boldsymbol{c}_{\boldsymbol{x}y}) - E[T(\boldsymbol{X})]b^* = 0.$$

In order for the above equation to hold for any $\boldsymbol{X}$, it is necessary and sufficient that

$$\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{\beta}^* - \boldsymbol{c}_{\boldsymbol{x}y} = 0 \text{ and } b = 0,$$

Hence, the optimal $\boldsymbol{\beta}$ and $b$ should satisfy

$$\boldsymbol{C}_{\boldsymbol{x}\boldsymbol{x}}\boldsymbol{\beta}^*_{\text{opt}} = \boldsymbol{c}_{\boldsymbol{x}y}, \text{ and } b_{\text{opt}} = 0,$$

which imply

$$\beta_{0,\text{opt}} = E[Y] - -\boldsymbol{\beta}^\top_{\text{opt}} E[\boldsymbol{X}].$$

Thus, the set of equations in (22.42) have been derived. Conversely, if the set of equations in (22.42) hold, then we can derive the orthogonality condition (15).

**22.13* Regression coefficient estimates.**

**Note:** There is a typo in (22.57) and (22.61). (22.57) should be

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top\right]^{-1} \sum_{j=1}^n (\boldsymbol{x}_j - \overline{\boldsymbol{x}}) y_j. \tag{16}$$

and the second equation of (22.61) should be

$$\text{Var}[\hat{\beta}_0] = \frac{\sigma_\epsilon^2}{n}\left(1 + n\overline{\boldsymbol{x}}^\top \left[\sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top\right]^{-1} \overline{\boldsymbol{x}}\right). \tag{17}$$

We first derive (16), i.e., the correct expression of (22.57). Let

$$Q = \sum_{j=1}^n (y_j - \beta_0 - \boldsymbol{\beta}^\top \boldsymbol{x}_j)^2.$$

Differentiate it with respect to $\boldsymbol{\beta}^\top$ and set it to $\mathbf{0}$:

$$\frac{\partial Q}{\partial \boldsymbol{\beta}^\top} = -2\sum_{j=1}^{n}(y_j - \beta_0 - \boldsymbol{\beta}^\top \boldsymbol{x}_j)\boldsymbol{x}_j = \mathbf{0}, \tag{18}$$

Similarly, by differentiating $Q$ with respect to $\beta_0$ and setting it to zero, we have

$$\frac{\partial Q}{\partial \beta_0} = -2\sum_{j=1}^{n}(y_j - \beta_0 - \boldsymbol{\beta}^\top \boldsymbol{x}_j) = 0,$$

from which we find the optimal $\hat{\beta}_0$ should satisfy

$$\sum_{j=1}^{n} y_j - n\hat{\beta}_0 - \hat{\boldsymbol{\beta}}^\top \sum_{j=1}^{n} \boldsymbol{x}_j = \mathbf{0},$$

hence,

$$\hat{\beta}_0 = \frac{\sum_{j=1}^{n} y_j}{n} - \hat{\boldsymbol{\beta}}^\top \frac{\sum_{j=1}^{n} \boldsymbol{x}_j}{n} = \overline{y} - \hat{\boldsymbol{\beta}}^\top \overline{\boldsymbol{x}}.$$

Substituting $\hat{\beta}_0$ into (18), we obtain

$$\sum_{j=1}^{n}(y_j - \overline{y} + \hat{\boldsymbol{\beta}}^\top \overline{\boldsymbol{x}} - \hat{\boldsymbol{\beta}}^\top \boldsymbol{x}_j)\boldsymbol{x}_j = \sum_{j=1}^{n}[(y_j - \overline{y}) - \hat{\boldsymbol{\beta}}^\top (\boldsymbol{x}_j - \overline{\boldsymbol{x}})]\boldsymbol{x}_j = 0. \tag{19}$$

Since

$$\sum_{j=1}^{n}(y_j - \overline{y}) = 0 \quad \text{and} \quad \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}}) = 0,$$

we can rewrite equation (19) as

$$\sum_{j=1}^{n}[(y_j - \overline{y}) - \hat{\boldsymbol{\beta}}^\top (\boldsymbol{x}_j - \overline{\boldsymbol{x}})](\boldsymbol{x}_j - \overline{\boldsymbol{x}}) = \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}})[(y_j - \overline{y}) - (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \hat{\boldsymbol{\beta}}] = 0 \tag{20}$$

or

$$\boldsymbol{\Sigma}_{\boldsymbol{x}} \hat{\boldsymbol{\beta}} = \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}})(y_j - \overline{y}), \tag{21}$$

where we denoted as

$$\boldsymbol{\Sigma}_{\boldsymbol{x}} \triangleq \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top.$$

Therefore,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}})(y_j - \overline{y}) = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}}) y_j \tag{22}$$

which is the corrected version of (22-57).

Now we proceed to derive (22.62) and then (22.61). From (22.47) we have

$$E[y_j] = f(\boldsymbol{x}_j) = \beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}_j$$

where we assume that $\epsilon_j$ represent noise with zero mean, i.e., $E[\epsilon_j] = 0$. It follows from this equation that

$$E[\overline{y}] = \beta_0 + \boldsymbol{\beta}^\top \overline{x}.$$

Thus,

$$E[y_j - \overline{y}] = \boldsymbol{\beta}^\top (\boldsymbol{x}_j - \overline{\boldsymbol{x}}) = (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \boldsymbol{\beta}. \tag{23}$$

Taking the expectation of (22) and using (23), we obtain

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}] &= \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}}) E[y_j - \overline{y}] \\ &= \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \left( \sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \right) \boldsymbol{\beta} = \boldsymbol{\beta}. \end{aligned} \tag{24}$$

In order to derive the first equation of (22.62), note that

$$y_j = E[y_j] + \epsilon_j.$$

Similar to the derivation of (24), we obtain

$$\hat{\boldsymbol{\beta}} = E[\hat{\boldsymbol{\beta}}] + \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}}) \epsilon_j = \boldsymbol{\beta} + \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}}) \epsilon_j, \tag{25}$$

from which the first equation of (22.62) readily follows. From the last equation we have

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}}) \epsilon_j,$$

Thus, the variance of $\hat{\boldsymbol{\beta}}$ is computed as

$$\begin{aligned} \mathrm{Var}[\hat{\boldsymbol{\beta}}] &= \sum_{j=1}^{n} \sum_{k=1}^{n} E[\epsilon_j \epsilon_k] \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_k - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}^{-1} \\ &= \sigma_\epsilon^2 \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}, \end{aligned}$$

where we use the property that the noise variables $\epsilon_j$'s are mutually independent, i.e., $E[\epsilon_j \epsilon_k] = \sigma_\epsilon^2 \delta_{jk}$.
The first equation in (22.61) can be obtained by taking the expectation of (22.58):

$$E[\hat{\beta}_0] = E[\overline{y}] - E[\hat{\boldsymbol{\beta}}^\top \overline{\boldsymbol{x}}] = \beta_0 + \boldsymbol{\beta}^\top \overline{\boldsymbol{x}} - \boldsymbol{\beta}^\top \overline{\boldsymbol{x}} = \beta_0. \tag{26}$$

In order to compute the variance of the estimate $\hat{\beta}_0$, we write from (22.58)

$$\hat{\beta}_0 = \overline{y} - \hat{\boldsymbol{\beta}}^\top \overline{\boldsymbol{x}} = \beta_0 + \boldsymbol{\beta}^\top \overline{\boldsymbol{x}} + \frac{1}{n} \sum_{j=1}^{n} \epsilon_j - \boldsymbol{\beta}^\top \overline{\boldsymbol{x}} + \sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \epsilon_j$$

where we used equation (25) for $\hat{\boldsymbol{\beta}}$. Thus,

$$\hat{\beta}_0 = \beta_0 + \sum_{j=1}^{n} \left( \frac{1}{n} - (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \right) \epsilon_j \tag{27}$$

Therefore,

$$\begin{aligned}
\operatorname{Var}[\hat{\boldsymbol{\beta}}_0] &= E\left[ \epsilon_j \epsilon_k \sum_{j=1}^{n} \left( \frac{1}{n} - (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \right) \sum_{k=1}^{n} \left( \frac{1}{n} - (\boldsymbol{x}_k - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \right) \right] \\
&= \frac{1}{n^2} \sum_j \sum_k \sigma_\epsilon^2 \delta_{jk} + \sum_j \sum_k \sigma_\epsilon^2 \delta_{jk} \overline{\boldsymbol{x}}^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_k - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \\
&- \frac{1}{n} \sum_j \sum_k \sigma_\epsilon^2 \delta_{jk} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} - \frac{1}{n} \sum_j \sum_k \sigma_\epsilon^2 \delta_{jk} (\boldsymbol{x}_k - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \\
&= \frac{\sigma_\epsilon^2}{n} + \sigma_\epsilon^2 \overline{\boldsymbol{x}}^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} - 0 - 0 \\
&= \frac{\sigma_\epsilon^2}{n} \left( 1 + n \overline{\boldsymbol{x}}^\top \left[ \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top \right]^{-1} \overline{\boldsymbol{x}} \right).
\end{aligned} \tag{28}$$

## 22.2 Linear Smoothing and Prediction: Wiener Filter Theory

**22.14* An alternative expression for (22.74).** If we define the output of the linear system as

$$Y_t = \sum_{i=0}^{n} h^*[i] X_{t-k},$$

Then (22.73) should replaced by

$$R_{yy}[d] = \sum_{k=0}^{n} \sum_{j=0}^{n} h^*[k] h[j] R_{xx}[d + j - k], \tag{29}$$

which in matrix form becomes

$$R_{yy}[d] = \boldsymbol{h}^H \boldsymbol{R}_{xx}[d] \boldsymbol{h},$$

where $\boldsymbol{h}^H = (h^*[0], h^*[1], \ldots, h^*[n])$.

**22.15 Alternative derivation of Wiener-Hopf equations (22.89) and (22.90).**

(a) The prediction error $e_{t+p}$ is orthogonal to the linear subspace $L_X$, if it is orthogonal to the input signal $X_{t-j}$ for all $j \geq 0$:

$$\langle S_{t+p} - \sum_{k=0}^{\infty} h[k] X_{t-k}, X_{t-j} \rangle = 0, \ \ j \geq 0,$$

or equivalently

$$E\left[\left(S_{t+p} - \sum_{k=0}^{\infty} h[k]X_{t-k}\right)\overline{X}_{t-j}\right] = 0,\ \ j \geq 0,$$

which can be reduced to

$$R_{sx}[j+p] = \sum_{k=0}^{\infty} h[k]R_{xx}[j-k],$$

which is equivalent to (22.89).

(b) The continuous-time case proceeds exactly in the same fashion. The prediction error

$$e(t+\lambda) = S(t+\lambda) - \hat{S}(t+\lambda) = S(t+\lambda) - \int_0^{\infty} h(u)X(t-u)\,du$$

is orthogonal to the subspace $L_X$ spanned by the past input $\{X(t-v),\ v \geq 0\}$, if

$$\langle S(t+\lambda) - \int_0^{\infty} h(u)X(t-u)\,du, X(t-v)\rangle = 0,\ \ v \geq 0,$$

or equivalently

$$E\left[\left(S(t+\lambda) - \int_0^{\infty} h(u)X(t-u)\,du\right)\overline{X}(t-v)\right] = 0,\ \ v \geq 0,$$

which reduces to

$$R_{sx}(v+\lambda) = \int_0^{\infty} h(u)R_{xx}(v-u)\,dv.$$

This is equivalent to the Wiener-Hopf integral equation (22.90).

**22.16 Derivation of (22.103).** Since $Z_t$ is WSS, its autocorrelation function is

$$\begin{aligned} R_{zz}[k] &= E[Z_k\overline{Z}_0] = E\left[\sum_j g_s[k-j]W_j \sum_i \overline{g}_s[-i]\overline{W}_i\right] \\ &= \sum_j\sum_i g_s[k-j]E[W_j\overline{W}_i]\overline{g}_s[-i] = \sum_j\sum_i g_s[k-j]R_{ww}[j-i]\overline{g}_s[-i] \end{aligned}$$

Then taking the $Z$-transform of the above,

$$\begin{aligned} P_{zz}(z) &= \sum_k R_{zz}[k]z^{-k} = \sum_j\sum_k\sum_i g_s[k-j]z^{-(k-j)}R_{ww}[j-i]z^{-(j-i)}\overline{g}_s[-i]z^{-i} \\ &= G_s(z)R_{ww}(z)\overline{G}_s(z^{-1}). \end{aligned}$$

**22.17 Physically unrealizable filter.**

The second term can be expanded as

$$\frac{z}{1-\overline{\gamma}z} = \sum_{j=0}^{\infty} \overline{\gamma}^j z^{j+1}.$$

Thus, it has non-zero terms only in the negative time domain, i.e.,

$$\text{The } k\text{-th term} = \begin{cases} 0, & k \geq 0 \\ \overline{\gamma}^{-k-1}, & k \leq -1. \end{cases}$$

Hence, this is an anti-causal term or physically unrealizable.

**22.18 Signal power $\sigma_s^2$.**

$$\begin{aligned}
\sigma_s^2 = R_{ss}[0] &= \frac{1}{2\pi i} \oint_C \frac{P_{ss}(z)}{z}\, dz = \frac{|A|^2}{2\pi i} \oint_C \frac{1}{(1-\alpha z^{-1})(1-\overline{\alpha} z) z}\, dz \\
&= \frac{|A|^2}{2\pi i} \oint_C \left( \frac{1}{(1-|\alpha|^2)(z-\alpha)} + \frac{\overline{\alpha}}{(1-|\alpha|^2)(1-\overline{\alpha} z)} \right) dz \\
&= \frac{|A|^2}{1-|\alpha|^2},
\end{aligned}$$

where the contour $C$ is the unit circle $|z| = 1$, and we used the residue theorem to evaluate the contour integration.

Alternatively

$$\begin{aligned}
\sigma_s^2 &= \frac{1}{2\pi} \int_0^{2\pi} P_s(\omega)\, d\omega \\
&= \frac{1}{2\pi} \int_0^{2\pi} \frac{|A|^2}{1 + |\alpha|^2 - 2\Re\{\alpha\} \cos\omega}\, d\omega, \qquad = \frac{|A|^2}{1-|\alpha|^2}.
\end{aligned}$$

**22.19 Continuous-time Gauss-Markov process.**

(a)

$$G_s(f) = \frac{1}{1 + i a^{-1} f}, \quad -\infty < f < \infty.$$

(b) The impulse response function $g(t)$ is

$$g_s(t) = \sum_{-\infty}^{\infty} G_s(f) e^{i2\pi f t}\, df = \int_{-\infty}^{\infty} \frac{e^{i2\pi f t}}{1 + i a^{-1} f}\, df. \tag{30}$$

Now recall the residue theorem in complex analysis, which we already used in Section 8.2, where we dealt with characteristic functions.

**Theorem 22.1 (The Residue Theorem).** *For any closed contour $C$ that encloses at most a finite number of singularities $z_1, z_2, \ldots, z_n$ of a function $f(z)$ continuous on $C$, we have*

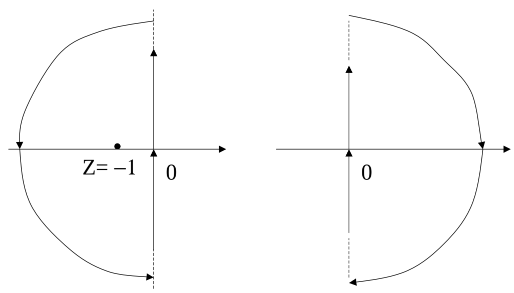
$$\boxed{\oint_C f(z)\, dz = 2\pi i \left\{ \sum_{k=1}^{n} \mathrm{Res} f(z_k) \right\}.} \tag{31}$$

□

By setting

$$f_s(z) = \frac{e^{2\pi z t}}{1 + a^{-1} z}, \tag{32}$$

we consider the line integration of $f_s(z)$ along the contour $C$ shown in Figure 22.1 (a). The




(a) Contour for t > 0. (b) Contour for t < 0.

**Figure 22.1** (a) For $t > 0$, the function $f_s(z) = \frac{e^{2\pi z t}}{1+a^{-1}z}$ has $z = -a$ as a pole inside the contour $C$; (b) For $t < 0$, the function $f_s(z)$ is analytic inside the contour.

line integration along the semi-circle in $Z < 0$ vanishes for $t \geq 0$, and the residue at $z = -a$ is $ae^{-2\pi at}$. Thus, we have

$$I = \int_{-\infty}^{\infty} \frac{e^{2\pi iyt}}{1 + ia^{-1}y}\, idy = 2\pi iae^{-2\pi t}, \;\; t \geq 0.$$

By setting $y = f$ in the last expression, we find that the integration of (30) is equal to $I/i$ for $t \geq 0$. For $t < 0$, we take the contour integration of Figure 22.1 (b). Since there is no pole inside of this contour, the integral $I = 0$ for $t < 0$. Thus, we have obtained

$$g_s(t) = \begin{cases} 0, & t < 0, \\ 2\pi ae^{-2\pi at}, & t \geq 0. \end{cases}$$

(c) Hence, we find

$$f_s(t) = \begin{cases} 0, & t < 0, \\ 2\pi ae^{-2\pi a(t+\lambda)}, & t \geq 0. \end{cases}$$

By taking the Fourier transform, we obtain the transfer function

$$F_s(f) = 2\pi a \int_0^{\infty} e^{-2\pi a(t+\lambda)} e^{-2\pi ift}\, dt = \frac{e^{-2\pi a\lambda}}{1 + ia^{-1}f}.$$

Then combining this and $G_s^{-1}(f) = 1 + ia^{-1}f$, we find that the optimal pure predictor is given by

$$H_{\mathrm{opt}}(f) = \frac{F_s(f)}{G_s(f)} = e^{-2\pi\lambda} \leq 1. \tag{33}$$

Thus, the optimal prediction of $S(t + \lambda)$ is simply given by attenuating $S(t)$ by factor $e^{-2\pi a\lambda}$:

$$\hat{S}(t + p) = e^{-2\pi a\lambda} S(t).$$

By taking the Fourier transform, we obtain the transfer function

$$F_s(f) = 2\pi \int_0^\infty e^{-2\pi_{t+p}} z^{-t} = \frac{e^{-2\pi\lambda}}{1+if}.$$

Then combining this and $G_s^{-1}(f) = 1 + ia^{-1}f$, we find that the optimal pure predictor is given by

$$H_{\text{opt}}(f) = \frac{F_s(f)}{G_S(f)} = e^{-2\pi a\lambda} \leq 1. \tag{34}$$

(d) Thus, the optimal prediction of $S(t+\lambda)$ is simply given by attenuating $S(t)$ by factor $e^{-2\pi\lambda}$:

$$\hat{S}(t+\lambda) = e^{-2\pi a\lambda} S(t).$$

(e) The conditional distribution of $S(t+\lambda)$ given $S_t$ is Gaussian with mean

$$E[S(t+\lambda)|S(t)] = \rho S(t) = e^{-2\pi a\lambda} S(t), \;\; \lambda \geq t_0. \tag{35}$$

and variance

$$\text{Var}[S(t+\lambda)|S(t)] = \sigma_s^2(1-\rho^2) = \sigma_s^2 \left(1 - e^{-4\pi a\lambda}\right), \;\; \lambda \geq 0, \tag{36}$$

where $\sigma_s^2 = \int_{-\infty}^{\infty} P_{ss}(f)\, df = \pi a$.

**22.20 Optimum smoothing filter for uncorrelated signal and noise.**

(a) The first term in the square bracket of (22.204) is causal, and a portion of the second term contributes to the causal part, i.e.,

$$\left[\frac{az^{-d}}{1-\alpha z^{-1}} + \frac{bz^{-d+1}}{1-\overline{\gamma}z}\right]_+ = \frac{az^{-d}}{1-\alpha z^{-1}} + \left[\frac{bz^{-d+1}}{1-\overline{\gamma}z}\right]_+.$$

and

$$\left[\frac{bz^{-d+1}}{1-\overline{\gamma}z}\right]_+ = b\sum_{j=0}^{\infty} \overline{\gamma}^j z^{-(d-j-1)} u[d-j-1] = b\sum_{j=0}^{d-1} \overline{\gamma}^j z^{-(d-j-1)}$$

Hence,

$$H_{\text{opt}}(z) = \frac{|A|^2(1-\alpha z^{-1})(1-\beta z^{-1})}{|C|^2(1-\gamma z^{-1})} \left[\frac{az^{-d}}{1-\alpha z^{-1}} + b\sum_{j=0}^{d-1} \overline{\gamma}^j z^{-(d-j-1)}\right]$$

(b) Set $\beta = 0$:

$$a = \frac{1}{1-\alpha\overline{\gamma}}, \;\; b = \frac{\overline{\gamma}}{1-\alpha\overline{\gamma}}.$$

Thus,

$$H_{\text{opt}}(z) = \frac{|A|^2}{|C|^2(1-\alpha\overline{\gamma})} \frac{\left[z^{-d} + (1-\alpha z^{-1}) \sum_{j=0}^{d-1} \overline{\gamma}^{j+1} z^{-d+j+1}\right]}{1-\gamma z^{-1}}.$$

Hence

$$h_{\rm opt}[k] = \frac{|A|^2\gamma^{k-d}}{|C|^2(1-\alpha\overline{\gamma})}\left[1 + \sum_{j=\max(0,d-k+1)}^{d-1} |\gamma|^{2(j+1)} - \alpha\gamma \sum_{j=\max(0,d-k)}^{d-1} |\gamma|^{2j}\right]$$

For $d = 0$, the the second and third terms in the square bracket become zero, hence

$$h_{\rm opt}[k] = \frac{|A|^2\gamma^{k}}{|C|^2(1-\alpha\overline{\gamma})},$$

which is equivalent to (22.155), as expected.

(c) In the absence of noise, we know the exact value of the signal $S_{t-d}$ by the time we have $S_t$. Hence the smoothed output is just $S_t$ itself, i.e., $\hat{S}_{t-d} = S_{t-d}$, and the estimation error is always zero.

To verify this observation, set $B = 0$ in the result of (a). Then

$$P_{xx}(z) = P_{ss}(z) = \frac{|A|^2}{(1-\alpha z^{-1})(1-\overline{\alpha}z)} = G_s(z)\overline{G}_s(z^{-1})$$

Thus

$$\frac{P_{ss}(z)z^{-d}}{\overline{G}}_s(z^{-1}) = G_s(z)z^{-d},$$

and

$$\left[\frac{P_{ss}(z)z^{-d}}{\overline{G}}_s(z^{-1})\right]_+ = G_s(z)z^{-d},$$

and

$$H_{\rm opt}(z) = \frac{1}{G_s(z)}G_s(z)z^{-d} = z^{-d}.$$

Hence

$$h_{\rm opt}[k] = \delta_{k,d} = \begin{cases} 1, & k = d \\ 0, & k \neq d. \end{cases}$$

Therefore

$$\hat{S}_{t-d} = \sum_{k=0}^{\infty} h_{\rm opt}[k]X_{t-k} = X_{t-d} = S_{t-d}.$$

**22.21 Conservation of inner-products and expressions for the minimum mean square error.**

(a)

$$\frac{1}{2\pi i}\oint G^{(1)}(z)\overline{G}^{(2)}(z^{-1})\frac{dz}{z} = \sum_k \sum_j g^{(1)}[k]\overline{g}^{(2)}[j]\frac{1}{2\pi i}\oint z^{j-k}\frac{dz}{z}$$
$$= \sum_k \sum_j g^{(1)}[k]\overline{g}^{(2)}[j]\delta_{j,k} = \sum_k \sum_j g^{(1)}[k]\overline{g}^{(2)}[k],$$

where we used $\oint z^{n-1}\,dz = 2\pi i \delta_{n,0}$. (This formula can be proved by writing $z = e^{i\theta}$, and changing $\oint \; dz$ to $\int_0^{2\pi} \; d\theta$.)

(b) The autocorrelation and the power spectrum are related by

$$P_{ss}(z) = \sum_{k=-\infty}^{\infty} R_{ss}[k] z^{-k}, \quad \text{and} \quad R_{ss}[k] = \frac{1}{2\pi i} \oint P_{ss}(z) z^{k-1}\,dz.$$

Hence by setting $k = 0$, we have

$$R_{ss}[0] = \frac{1}{2\pi i} \oint P_{ss}(z) \frac{dz}{z}.$$

Similarly, the second term of (22.161) is (by dropping the subscript $(\mathrm{opt})$ of $h_{\mathrm{opt}}$)

$$\begin{aligned}
\sum_k \sum_j h[k] R_{xx}[j-k] \overline{h}[j] &= \sum_k \sum_j h[k] \frac{1}{2\pi i} \oint P_{xx}(z) z^{j-k} \frac{dz}{z} \overline{h}[j] \\
&= \frac{1}{2\pi i} \oint \left( \sum_k h[k] z^{-k} \right) P_{xx}(z) \left( \sum_j \overline{h}[j] z^j \right) \frac{dz}{z} \\
&= \frac{1}{2\pi i} \oint H(z) P_{xx}(z) \overline{H}(z^{-1}) \frac{dz}{z}.
\end{aligned}$$

## 22.3 Kalman Filter

**22.22 Derivation of the propagation step.** We replace the forward variables in both sides of (22.168) by the PDF of (22.167). Then by writing $\boldsymbol{y}_0^t = (\boldsymbol{y}_0^{t-1}, \boldsymbol{y}_t)$ in the argument of the LHS PDF of (22.168) and dividing by $f_{\boldsymbol{Y}_0^{t-1}}(\boldsymbol{y}_0^{t-1})$, we have

$$f_{\boldsymbol{S}_t, \boldsymbol{Y}_t | \boldsymbol{Y}_0^{t-1}}(\boldsymbol{s}_t, \boldsymbol{y}_t | \boldsymbol{y}_0^{t-1}) = \int_{\mathcal{S}} f_{\boldsymbol{S}_t, \boldsymbol{Y}_t | \boldsymbol{S}_{t-1}}(\boldsymbol{s}_t, \boldsymbol{y}_t | \boldsymbol{s}_{t-1}) f_{\boldsymbol{S}_{t-1} | \boldsymbol{Y}_0^{t-1}}(\boldsymbol{s}_{t-1} | \boldsymbol{y}_0^{t-1})\, d\boldsymbol{s}_{t-1}, \tag{37}$$

where we used the following property of the HMM:

$$f_{\boldsymbol{S}_t, \boldsymbol{Y}_t | \boldsymbol{S}_{t-1}, \boldsymbol{Y}_0^{t-1}}(\boldsymbol{s}_t, \boldsymbol{y}_t | \boldsymbol{s}_{t-1}, \boldsymbol{y}_0^{t-1}) = f_{\boldsymbol{S}_t, \boldsymbol{Y}_t | \boldsymbol{S}_{t-1}}(\boldsymbol{s}_t, \boldsymbol{y}_t | \boldsymbol{s}_{t-1}).$$

Then by integrating (37) with respect to $\boldsymbol{y}_t$, we obtain (22.170).

**22.23 Derivation of the update step.**

$$\begin{aligned}
&f_{\boldsymbol{S}|\boldsymbol{Y}_0^t}(\boldsymbol{s}_t | \boldsymbol{y}_0^t) f_{\boldsymbol{Y}_t | \boldsymbol{Y}_0^{t-1}}(\boldsymbol{y}_t | \boldsymbol{y}_0^{t-1}) = f_{\boldsymbol{S}|\boldsymbol{Y}_0^t}(\boldsymbol{s}_t | \boldsymbol{y}_0^{t-1}, \boldsymbol{y}_t) f_{\boldsymbol{Y}_t | \boldsymbol{Y}_0^{t-1}}(\boldsymbol{y}_t | \boldsymbol{y}_0^{t-1}) \\
&\quad = f_{\boldsymbol{S}_t, \boldsymbol{Y}_t | \boldsymbol{Y}_0^{t-1}}(\boldsymbol{s}_t, \boldsymbol{y}_t | \boldsymbol{y}_0^{t-1}) = f_{\boldsymbol{S}_t | \boldsymbol{Y}_0^{t-1}}(\boldsymbol{s}_t | \boldsymbol{y}_0^{t-1}) f_{\boldsymbol{Y}_t | \boldsymbol{S}_t, \boldsymbol{Y}_0^{t-1}}(\boldsymbol{y}_t | \boldsymbol{s}_t, \boldsymbol{y}_0^{t-1}) \\
&\quad = f_{\boldsymbol{S}_t | \boldsymbol{Y}_0^{t-1}}(\boldsymbol{s}_t | \boldsymbol{y}_0^{t-1}) f_{\boldsymbol{Y}_t | \boldsymbol{S}_t}(\boldsymbol{y}_t | \boldsymbol{s}_t),
\end{aligned}$$

which leads to (22.171).

**22.24 Derivation of the mean and covariance matrix (22.174).**

$$
\begin{aligned}
E[\boldsymbol{Y}_0] &= \boldsymbol{B}_0 E[\boldsymbol{S}_0] + E[\boldsymbol{W}_0] = \boldsymbol{B}_0 \hat{\boldsymbol{s}}_0, \\
\mathbf{Var}[\boldsymbol{Y}_0] &= \mathbf{Var}[\boldsymbol{B}_0 \boldsymbol{Y}_0] + \mathbf{Var}[\boldsymbol{W}_0] = \boldsymbol{B}_0 \mathbf{Var}[\boldsymbol{S}_0] \boldsymbol{B}_0^\top + \boldsymbol{C}_{\boldsymbol{u}_0} = \boldsymbol{B}_0 \boldsymbol{P}_0 \boldsymbol{B}_0^\top + \boldsymbol{C}_{\boldsymbol{u}_0}, \\
\mathbf{Cov}[\boldsymbol{Y}_0, \boldsymbol{S}_0] &= \mathbf{Cov}[\boldsymbol{B}_0 \boldsymbol{S}_0 + \boldsymbol{W}_0, \boldsymbol{S}_0] = \mathbf{Cov}[\boldsymbol{B}_0 \boldsymbol{S}_0, \boldsymbol{S}_0] = \boldsymbol{B}_0 \boldsymbol{P}_0.
\end{aligned}
$$

Hence we obtain (22.174).

**22.25 Kalman gain for scalar variables $S_0$ and $Y_0$.** It is simply given by

$$
K_0 = \rho_{s_0,s_0} \left( \frac{\sigma_{s_0}}{\sigma_{y_0}} \right),
$$

where $\sigma_{s_0}$ and $\sigma_{y_0}$ are the standard deviations of $S_0$ and $Y_0$, respectively and $\rho_{s_0,y_0}$ is the **correlation** coefficient between the two RVs.
See Section 4.7.1 of Chapter 4 and Example 22.1 and Problem 22.7 in Section 22.1.3).

**22.26 Derivation of the Kalman's predicted estimate.** Since $\boldsymbol{S}_1 = \boldsymbol{A}_1 \boldsymbol{S}_0 + \boldsymbol{W}_1$,

$$
\hat{\boldsymbol{s}}_{1|0} = E[\boldsymbol{S}_1 | \boldsymbol{y}_0] = \boldsymbol{A}_1 E[\boldsymbol{S}_0 | \boldsymbol{y}_0] + E[\boldsymbol{W}_1 | \boldsymbol{y}_0] = \boldsymbol{A}_1 \hat{\boldsymbol{s}}_{0|0} + 0.
$$

Similarly,

$$
\boldsymbol{P}_{1|0} = \mathbf{Var}[\boldsymbol{S}_1 | \boldsymbol{y}_0] = \boldsymbol{A}_1 \mathbf{Var}[\boldsymbol{S}_0 | \boldsymbol{y}_0] \boldsymbol{A}_1^\top + \boldsymbol{C}_{\boldsymbol{u}_1}.
$$

Since $\mathbf{Var}[\boldsymbol{S}_0 | \boldsymbol{y}_0] = \boldsymbol{P}_{0|0}$ by definition, we have (22.180).

# 23 Solutions for Chapter 23 Queuing and Loss Models

## 23.1 Introduction

## 23.2 Little's Formula

**23.1 Little's formula [197].**

(a) False : Little's formula holds for arbitrary arrival processes.
(b) False : For the same reason as part (a).
(c) True: Little's formula holds for arbitrary work-conserving disciplines.
(d) True: For the same reason as part (a).

**23.2 Little's formula for multiple type customers [197].** Little's law can be be generalized to

$$\overline{Q}_r = \lambda_r E[W_r], \tag{1}$$

where $\overline{Q}_r$ is the mean number of type $r$ customers in queue and $E[W_r]$ is the mean waiting time for type $r$ customers in queue, for $r = 1, \ldots, R$. The average queue size for a type $r$ customer is given by (1) for the FCFS queue discipline or any other work-conserving discipline.

**23.3 Distributions seen by arrivals and departures [197].** In the interval $[0, T]$, for each arrival which causes $Q(t)$ to increase from $n$ to $n+1$ ($n = 0, 1, \cdots$), there must be a corresponding departure that causes $Q(t)$ to decrease from $n+1$ to $n$ (since $Q(0) = Q(T) = 0$). This implies that the average queue size seen by an arrival is the same as that seen by a departure in the interval $[0, T]$.

## 23.3 Queueing Models

**23.4 Choice of a work unit in a queueing model.**

(a)

$$\lambda = \frac{1}{E[T]} \text{ [sec}^{-1}\text{]},$$

and

$$\mu = \frac{B}{L} \text{ [sec}^{-1}\text{]}.$$

(b) The interarrival times $T_n$'s are i.i.d. exponential random variables. The packet lengths $L_n$'s are geometrically distributed.

**23.5 State classification in the M/M/1 queue.**

Consider an embedded DTMC $\{X_n\}$ of the CTMC $N(t)$, by ignoring the sojourn time. Then $\{X_n\}$ is a random walk with $P(X_{n+1} = i+1|X_n = i) = p$, and $P(X_{n+1} = i-1|X_n = i) = q = 1-p$.

(i) If $0 < \rho < 1$, then $p < q$, then all states are positive recurrent with respect to the embedded Markov chain $\{X_n\}$. Then the stationary distribution $\tilde{\pi}_i = (1-r)r^i, \quad i = 0, 1, 2, \ldots$, where $r = \frac{p}{q}$. Thus, the mean recurrence time $\mu_{ii}$ is given by (see (15.62) of Section 15.3.1)

$$\mu_{ii} = \frac{1}{\tilde{\pi}_i} < \infty, \text{ for all } i = 0, 1, 2, \ldots.$$

Hence all states are positively recurrent. Then by definition all states are ergodic with respect to the CTMC $N(t)$ (see e.g. [203] p. 184).

(ii) If $\rho = 1$, then $p = q = 0.5$, then there is no bona fide probability distribution, since $\tilde{\pi}_i = 0$, for all $i = 0, 1, 2, \ldots$. Then the mean recurrence time

$$\mu_{ii} = \frac{1}{\tilde{\pi}_i} = \infty.$$

Hence all states are null-recurrent with respect to the EMC $\{X_n\}$. By definition all states of the CTMC are null-recurrent (see e.g. [203] p. 184).

(iii) Finally if $\rho = 0$, $p = 0$ and $q = 1$. Then state 0 is an absorbing state, and all other states are transient states.

**23.6 $L(t)$ in the M/M/1 queue analysis.**

$L(t)$ is not a BD process, since we cannot define $P(L(t+h) = 1|L(t) = 0)$ and $P(L(t+h) = 0|L(t)) = 0$, since transitions to state 0 or 1 depends on whether $N(t) = 0$ (the server is idle) or $N(t) = 1$ (i.e., the server is busy). In other words,

$$P(L(t+h) = 1|L(t) = 0, N(t) = 1)) = \lambda h + o(h)$$
$$P(L(t+h) = 0|L(t) = 0, N(t) = 0)) = 1 + o(h)$$

Thus, we cannot define $P_{m,n}(h)$ of (14.41) for $L(t)$.

**23.7 Mean and variance in the M/M/1 queue.**

$$\begin{aligned}\overline{N} &= \sum_{n=0}^{\infty} n(1-\rho)\rho^n = \rho(1-\rho)\sum_{n=0}^{\infty} n\rho^{n-1} \\ &= \rho(1-\rho)\frac{d}{d\rho}\rho^n = \rho(1-\rho)\frac{d}{d\rho}(1-\rho)^{-1} \\ &\frac{\rho}{1-\rho}.\end{aligned}$$

Similarly, use the formula

$$\sum_{n=0}^{\infty} n(n-1)\rho^{n-2} = \frac{d^2}{d\rho^2}\left(\frac{1}{1-\rho}\right) = \frac{2}{(1-\rho)^3}.$$

Thus,

$$\sigma_N^2 = E[N^2] - \overline{N}^2 = \sum_{n=0}^{\infty} n^2(1-\rho)\rho^n - \left(\frac{\rho}{1-\rho}\right)^2$$

$$(1-\rho)\left[\rho^2 \sum_{n=0}^{\infty} n(n-1)\rho^{n-2} + \rho \sum_{n=0}^{\infty} n\rho^{n-1}\right] - \left(\frac{\rho}{1-\rho}\right)^2$$

$$= \frac{\rho}{(1-\rho)^2}.$$

Similarly

$$\overline{Q} = \overline{N} - (1-\pi_0) = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}.$$

We write

$$(n-1)^2 = n(n-1) - n + 1.$$

Then by changing $\sum_{n=1}^{\infty}$ to $\sum_{n=0}^{\infty}$ and subtracting $n=0$ term, we have

$$\begin{aligned}
\sum_{n=0}^{\infty}(n-1)^2\pi_n &= (1-\rho)\rho^2 \sum_{n=0}^{\infty} n(n-1)\rho^{n-2} - \pi_0 - (1-\rho)\rho \sum_{n=0}^{\infty} n\rho^{n-1} + 1 \\
&= \frac{2\rho^2}{(1-\rho)^2} - (1-\rho) - \frac{\rho}{1-\rho} \\
&= \frac{\rho^3 + \rho^2 - \rho^4}{(1-\rho)^2} \\
&= \frac{\rho^2(1+\rho-\rho^2)}{(1-\rho)^2}
\end{aligned}$$

**23.8* Derivation of the waiting time distribution (23.37).**

From (23.36) we have

$$\begin{aligned}
F_W(x) &= 1 - \rho + (1-\rho)\sum_{n=1}^{\infty}\rho^n - (1-\rho)e^{-\mu x}\sum_{n=1}^{\infty}\sum_{j=0}^{n-1}\rho^n \frac{(\mu x)^j}{j!} \\
&= 1 - (1-\rho)e^{-\mu x}\sum_{j=0}^{\infty}\sum_{n=j+1}^{\infty}\rho^n \frac{(\mu x)^j}{j!} \\
&= 1 - (1-\rho)e^{-\mu x}\sum_{j=0}^{\infty}\frac{\rho^{j+1}}{1-\rho}\frac{(\mu x)^j}{j!} = 1 - \rho e^{-\mu x}\sum_{j=0}^{\infty}\frac{(\rho\mu x)^j}{j!} \\
&= 1 - \rho e^{-\mu(1-\rho)x}.
\end{aligned}$$

**23.9 Cumulative Poisson distribution [203]**

(a) The left-hand side (LHS) is given by

$$\begin{aligned}
\text{LHS} &= \sum_{j=0}^{k} P(k-j;a_1)Q(j;a_2) = \sum_{j=0}^{k} \frac{a_1^{k-j}}{(k-j)!} e^{-a_1} \sum_{i=0}^{j} \frac{a_2^i}{i!} e^{-a_2} \\
&= \sum_{l=0}^{k} \sum_{i=0}^{l} \frac{a_1^{l-i}}{(l-i)!} \frac{a_2^i}{i!} e^{-(a_1+a_2)} = \sum_{l=0}^{k} \frac{(a_1+a_2)^l}{l!} e^{-(a_1+a_2)} \\
&= Q(k;a_1+a_2) = \text{RHS},
\end{aligned} \tag{2}$$

where the variable substitution $l = i + k - j$ is made in (48).

(b) We can write

$$P(k;y) = \frac{d}{dy}\left[e^{-y}\right]\frac{y^k}{k!}.$$

Using integration by parts, we can write

$$\begin{aligned}
\int_a^\infty P(k;y)dy &= e^{-y}\frac{y^k}{k!}\Big|_{y=a}^{\infty} - \int_a^\infty e^{-y}\frac{y^{k-1}}{(k-1)!}dy \\
&= P(k;a) - \int_a^\infty P(k-1;y)dy.
\end{aligned}$$

By applying this recursive relation repeatedly, we find

$$\int_a^\infty P(k;y)dy = P(k;a) + P(k-1;a) + \cdots + P(0;a) = Q(k;a).$$

We can write

$$\int_a^\infty P(k;y)dy = \int_a^\infty e^{-y}\frac{y^k}{k!}dy = \frac{1}{k!}\int_a^\infty e^{-y}y^k dy = \frac{\Gamma(k+1;a)}{k!}.$$

(c) The identity to be proven is equivalent to

$$(k+a+1)Q(k;a) = aQ(k-1;a) + (k+1)Q(k+1;a). \tag{3}$$

The LHS of (3) can be written as

$$\text{LHS} = aQ(k;a) + (k+1)Q(k;a). \tag{4}$$

We can also write the following relations:

$$Q(k;a) = Q(k-1;a) + P(k;a) \tag{5}$$

$$Q(k;a) = Q(k+1;a) - P(k+1;a) \tag{6}$$

Substituting (5) and (6) into (4), we obtain

$$\begin{aligned}
\text{LHS} &= a[Q(k-1;a) + P(k;a)] + (k+1)[Q(k+1;a) - P(k+1;a)] \\
&= [aP(k;a) - (k+1)P(k+1;a)] \\
&\quad + [aQ(k-1;a) + (k+1)Q(k+1;a)]
\end{aligned} \tag{7}$$

It is easy to verify that the first term in square brackets in (7) is zero. Hence, (3) is established.

(d) The result in part (c) can be rewritten in the form (3). Substituting $k$ for $k+1$ in (3), we have

$$(k+a)Q(k-1;a) = aQ(k-2;a) + kQ(k;a) \tag{8}$$

Rearranging terms, we can write

$$kQ(k;a) - aQ(k-1;a) = kQ(k-1;a) - aQ(k-2;a). \tag{9}$$

Using the relation

$$kQ(k-1;a) = Q(k-1;a) + (k-1)Q(k-1;a) \tag{10}$$

in (9), we obtain

$$kQ(k;a) - aQ(k-1;a) \tag{11}$$

$$= Q(k-1;a) + [(k-1)Q(k-1;a) - aQ(k-2;a)]. \tag{12}$$

Applying the above recursive relation repeatedly, we obtain

$$\begin{aligned}
&kQ(k;a) - aQ(k-1;a) \\
&= Q(k-1;a) + Q(k-2;a) + [(k-2)Q(k-2;a) - aQ(k-3;a)] \\
&= Q(k-1;a) + Q(k-2;a) + Q(k-3;a) + \cdots + \\
&Q(1;a) + [Q(1;a) - aQ(0;a)] \\
&= \sum_{j=0}^{k-1} Q(j;a)
\end{aligned}$$

**23.10* Time-dependent solution for a certain BD process**: Consider a BD process with $\lambda_n = \lambda$ for all $n \geq 0$, and $\mu_n = n\mu$ for all $n \geq 1$. This process represents the $M/M/\infty$ queue. Find the partial differential equation that $G(z,t)$ must satisfy. Show that the solution to this equation is

$$G(z,t) = \exp\left\{\frac{\lambda}{\mu}(1-e^{-\mu t})(z-1)\right\}.$$

Show that the solution for $p_n(t)$ is given as

$$p_n(t) = \frac{(\frac{\lambda}{\mu}(1-e^{-\mu t}))^j}{j!} \exp\left\{-\frac{\lambda}{\mu}(1-e^{-\mu t})\right\},\ 0 \leq n < \infty. \tag{13}$$

**23.11 Time-dependent solution for M/G/$\infty$ is also Poisson [203, 249].**

(a) Let $U$ be the arrival epoch. For given $U = u(<t)$, the (conditional) probability that this customer is still in service at time $t$ is $P(S > t-u) = 1 - F_S(y-u)$. By averaging this over the distribution of $U$, we find the probability that the customer is still in service at time $t$ is given by (23.145).

(b)

$$\binom{N}{n} p^n(t)(1-p(t))^{N-n},\ \ 0 \leq n \leq N.$$

(c)

$$\pi_n(t) = \sum_{N=n}^{\infty} \frac{(\lambda t)^N}{n!} e^{-\lambda t} \binom{N}{n} p^n(t)(1-p(t))^{N-n}$$
$$= \frac{(\lambda t p(t))^n}{n!} e^{-\lambda t p(t)}.$$

By substituting $p(t)$ of (23.145), we obtain (23.47).

**23.12 The departure process of M/G/$\infty$ is Poisson.**

(a) The probability that a randomly picked customer is no longer in service at time $t$ is $1 - p(t)$. Then by applying the arguments given in the solution to Problem 23.11 (b) and (c), the probability that there are $m$ customers who have received service and departed by time $t$ is found to be

$$\frac{(\lambda t(1-p(t)))^m}{m!} e^{-\lambda t(1-p(t))}.$$

(b) Let $A(t)$ denote the number of arrivals in $(0,t)$. Conditioned on the event $\{A(t+s)=n\}$, let $E_1$ be the event that a given customer out of the $n$ departs in $(0,t)$ and let $E_2$ be the event that a given customer departs in $(t,t+s)$. Let $p_1$ and $p_2$ denote the probabilities of events $E_1$ and $E_2$, respectively.
Since the M/G/$\infty$ queue has an infinite number of servers with independent service times, all events associated with a given customer are independent of events associated with all other customers. Hence, the events $E_1$ and $E_2$ for one customer are independent of $E_1$ and $E_2$ for another customer. Let $D_1$ and $D_2$ represent the number of of departures in $(0,t)$ and $(t,t+s)$, respectively. By the aforementioned independence property, we find that the joint distribution of $D_1$ and $D_2$ conditioned on $A(t+s)=n$ has a multinomial distribution:

$$P[D_1=n_1, D_2=n_2 \mid A(t+s)=n] = \frac{n!}{n_1!n_2!(n-n_1-n_2)!} p_1^{n_1} p_2^{n_2} (1-p_1-p_2)^{n-n_1-n_2}. \tag{14}$$

Using a similar approach to the solution of Problem 23.11 (b) and (c), we obtain

$$P[D_1=n_1, D_2=n_2]$$
$$= \sum_{n=n_1+n_2} P[D_1=n_1, D_2=n_2 \mid A(t+s)=n] \cdot P[A(t)=n]$$
$$= \sum_{n=n_1+n_2} \frac{n!}{n_1!n_2!(n-n_1-n_2)!} p_1^{n_1} p_2^{n_2} (1-p_1-p_2)^{n-n_1-n_2} \cdot \frac{[\lambda(t+s)]^n}{n!} e^{-\lambda(t+s)} \tag{15}$$
$$= \frac{[\lambda p_1(t+s)]^{n_1}}{n_1!} \cdot \frac{[\lambda p_2(t+s)]^{n_2}}{n_2!}. \tag{16}$$

Equation (16) establishes that $D_1$ and $D_2$ are independent, which was to be shown. Moreover, $D_1$ is Poisson distributed with parameter $p_1\lambda(t+s)$ and $D_2$ is Poisson distributed with parameter $p_2\lambda(t+s)$.

(c) Applying the uniformity of Poisson arrivals in the interval $(0, t+s)$, the probability $p_2$ can be expressed as follows:

$$\begin{aligned} p_2 &= P[\text{customer arriving in } (0,t+s) \text{ departs in } (t,t+s)] \\ &= \frac{t}{t+s} \cdot P[\text{departs in } (t,t+s) \mid \text{arrives in } (0,t)] \\ &\quad + \frac{s}{t+s} P[\text{departs in } (t,t+s) \mid \text{arrives in } (t,t+s)]. \end{aligned} \tag{17}$$

We have

$$\begin{aligned} &P[\text{departs in } (t,t+s) \mid \text{arrives in } (0,t)] \\ &\quad = \frac{1}{t}\int_0^t [F_S(t+s-u) - F_S(t-u)]\,du \\ &\quad = \frac{1}{t}\int_s^{t+s} F_S(x)\,dx - \frac{1}{t}\int_0^t F_S(x)\,dx \end{aligned} \tag{18}$$

and

$$P[\text{departs in } (t,t+s) \mid \text{arrives in } (t,t+s)] \tag{19}$$

$$\begin{aligned} &\quad = \frac{1}{s}\int_t^{t+s} F_S(t+s-u)\,du \\ &\quad = \frac{1}{s}\int_0^s F_S(x)\,dx. \end{aligned} \tag{20}$$

Substituting (18) and (20) into (17), we have

$$p_2 = \frac{1}{t+s}\left[\int_s^{t+s} F_S(x)\,dx - \int_0^t F_S(x)\,dx\right] + \frac{1}{t+s}\int_0^s F_S(x)\,dx \tag{21}$$

$$= \frac{1}{t+s}\int_t^{t+s} F_S(x)\,dx. \tag{22}$$

Applying (22) in (16), we have

$$P[D_2 = n] = e^{-\lambda \int_t^{t+s} F_S(x)\,dx} \cdot \frac{\left[\lambda \int_t^{t+s} F_S(x)\,dx\right]^n}{n!}. \tag{23}$$

Setting $s = h$ and $n = 1$, we obtain

$$\begin{aligned} P[\text{one departure in } (t,t+h)] &= e^{-\lambda F_s(t)h} \lambda F_S(t)h = \lambda F_S(t)h + o(h) \\ &\xrightarrow{t\to\infty} \lambda h + o(h). \end{aligned} \tag{24}$$

Thus, in steady-state the departure process is a homogeneous Poisson process with rate $\lambda$.

**23.13 Mean queue length and mean waiting time.**

(a)

$$
\begin{aligned}
\bar{Q} &= \sum_{n=m}^{\infty}(n-m)p_n = \sum_{n=m}^{\infty}\rho^{n-m}p_m = p_m\sum_{k=0}^{\infty}k\rho^k = p_m\rho\sum_{k=0}^{\infty}k\rho^{k-1} \\
&= \rho p_m\sum_{k=0}^{\infty}\frac{d\rho^k}{d\rho} = \rho p_m\frac{d}{d\rho}\left[\sum_{k=0}^{\infty}\rho^k\right] = \rho p_m\frac{d}{d\rho}\left[\frac{1}{1-\rho}\right] \\
&= \rho p_m\frac{1}{(1-\rho)2} = \frac{\rho}{1-\rho}\cdot\frac{p_m}{1-\rho} = \frac{\rho C(m,a)}{1-\rho},
\end{aligned}
$$

since $C(m,a) = p_m/(1-\rho)$.

(b) Applying Little's formula:

$$
\bar{W} = \frac{\bar{Q}}{\lambda} = \frac{(\rho/\lambda)C(m,a)}{1-\rho} = \frac{C(m,a)}{m\mu(1-\rho)}.
$$

**23.14 Relations between the two Erlang formulas.**

From (23.50), (23.52), and (23.53), we have

$$
C(m,a) = \frac{\frac{a^m}{m!}\frac{m}{m-a}}{\sum_{n=0}^{m-1}\frac{a^n}{n!}+\frac{a^m}{m!}\frac{1}{1-\rho}}. \tag{25}
$$

Dividing the numerator and denominator by $\sum_{n=0}^{m}\frac{a^m}{m!}$ and using $\rho = \frac{a}{m}$ and the definition of $B(m,a)$, we obtain

$$
C(m,a) = \frac{B(m,a)\frac{m}{m-a}}{1-B(m,a)+B(m,a)\frac{m}{m-a}} = \frac{mB(m,a)}{m-a(1-B(m,a))}. \tag{26}
$$

**23.15* Waiting time distribution in the M/M/$m$ queue.**

(a)

$$
\begin{aligned}
F_W^c(x)\sum_{n=0}^{\infty}a_{m+n}F_W^c(x|m+n) &= \sum_{n=0}^{\infty}\pi_{m+n}F_W^c(x|m+n) \\
&= F_W^c(0)(1-\rho)\sum_{n=0}^{\infty}\rho^nF_W^c(x|m+n),
\end{aligned} \tag{27}
$$

where we used $\pi_{m+n} = \rho^n\pi_m$ from (23.52) and $\pi_m = (1-\rho)F_W^c(0)$ from (23.53). Note that the formula (27) holds under any work-conserving queue discipline, although the actual functional forms of $F_W^c(x)$ and $F_W^c(x|m+n)$ will depend on the specific queue discipline.

(b) Let $T_i$ be the interval between the $(i-1)$st service completion and the $i$th completion, $i = 2,3,\ldots,n+1$, as illustrated in Figure 23.1.

In this scenario, all $m$ exponential servers are busy. Let $X_j$, $1\le j\le m$ be the interval from an arbitrarily chosen instant until the completion of a job in service at the $j$th server. The $X_j$'s are independent and identically distributed with complementary distribution function

$$
F_{X_j}^c(t) = P\{X_j \ge t\}e^{-\mu t}.
$$

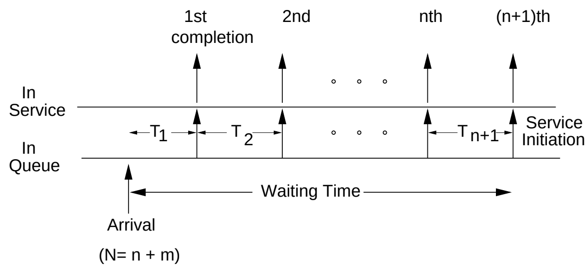



**Figure 23.1** Relationship between the waiting time and service completion intervals $T_i$'s in M/M/$m$.

The distribution of $T_i$, $1 \le i \le n+1$ is equivalent to that of the random variable $T$ defined by

$$T \triangleq \min\{X_1, X_2, \ldots, X_m\}.$$

The complementary distribution function of $T$ satisfies

$$\begin{aligned} F_T^c(t) &= P\{T \ge t\} = P\{X_j \ge t : \forall j,\ 1 \le j \le m\} \\ &= \prod_{j=1}^m P\{X_j \ge t\} = \prod_{j=1}^m e^{-\mu t} = e^{-m\mu t}. \end{aligned}$$

Therefore, $T_i$, $1 \le i \le n+1$ are exponentially distributed with parameter $m\mu$. Further, it should be clear that the $T_i$'s are independent.

(c) The waiting time of the customer in question is $T_1 + T_2 + \cdots + T_{n+1}$. Following the arguments that led to (**??**)(Note that in the waiting time analysis for M/M/1, we assumed that $n-1$ customers were in queue, whereas here we assume $n$ customers in queue.), we obtain (23.148), which is an $(n+1)$-stage Erlangian distribution.

(d) Substitution of (23.148) into (27) gives

$$F_W^c(x) = F_W^c(0) e^{-m\mu x}(1-\rho) \sum_{n=0}^{\infty} \rho^n \sum_{j=0}^{n} \frac{(m\mu x)^j}{j!}. \tag{28}$$

The double summation over $(n, j)$ can be rewritten as a double summation over $(k, j)$, where $k = n - j$, as follows:

$$\begin{aligned} \sum_{n=0}^{\infty} \sum_{j=0}^{n} \rho^n \frac{(m\mu x)^j}{j!} &= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \rho^{k+j} \frac{(m\mu x)^j}{j!} \\ &= \sum_{k=0}^{\infty} \rho^k e^{\rho m\mu x} = \frac{1}{1-\rho} e^{\rho m\mu x} \end{aligned} \tag{29}$$

Using (29)in (28) yields where we interchanged the order of summation and using the formula for a geometric series, we obtain

$$F_W^c(x) = F_W^c(0) e^{-m\mu(1-\rho)x} \tag{30}$$

or (23.149).

## 23.16 Normalization constant of (23.62).

For any B-D process, the normalization constant is given by:

$$G = \sum_{n=0}^{\infty} \prod_{j=0}^{n-1} \frac{\lambda(j)}{\mu(j+1)}, \tag{31}$$

where the $n = 0$ term in the sum is the constant 1. For the $M(K)/M/m$ queue, the birth and death rates are given by:

$$\lambda(n) = (K - n)\nu, \tag{32}$$

$$\mu(n) = \min\{m, n\}\mu, \quad n = 0, 1, \cdots, K. \tag{33}$$

For $0 \leq n < m$ we have:

$$\prod_{j=0}^{n-1} \frac{\lambda(j)}{\mu(j+1)} = \prod_{j=0}^{n-1} \frac{(K-j)\nu}{(j+1)\mu} = \binom{K}{n} r^n \tag{34}$$

For $m \leq n \leq K$ we have:

$$\prod_{j=0}^{n-1} \frac{\lambda(j)}{\mu(j+1)} = \prod_{j=0}^{m-1} \frac{(K-j)\nu}{(j+1)\mu} \prod_{j=m}^{n-1} \frac{(K-j)\nu}{m\mu} = \frac{r^n \prod_{j=0}^{n-1}(K-j)}{m! m^{n-m}}.$$

But:

$$\prod_{j=0}^{n-1}(K-j) = \frac{K!}{(K-n)!}.$$

Therefore, the normalization constant is given by

$$G(K) = \sum_{n=0}^{m-1} \binom{K}{n} r^n + \sum_{n=m}^{K} \frac{K! r^n}{(K-n)! m! m^{n-m}}. \tag{35}$$

**23.17 Steady-state distribution of the number of customers in the system G($K$)/M/$m$.**
For any B-D process:

$$\pi_n = \prod_{j=0}^{n-1} \frac{\lambda(j)}{\mu(j+1)} \pi_0. \tag{36}$$

using the results in the previous problem, we readily find that when $0 \leq n \leq m$ in the $M(K)/M/m$ queue, we have

$$\pi_n(K) = \binom{K}{n} r^n \pi_0(K), \tag{37}$$

When $m \leq n \leq K$, we have

$$\begin{aligned} \pi_n(K) &= \frac{K! r^n}{(K-n)! m! m^{n-m}} \pi_0(K) = \frac{\frac{(mr^{-1})^{K-n}}{(K-n)!}}{\frac{(mr^{-1})^K}{K!}} \frac{m^m}{m!} \pi_0(K) \\ &= \frac{K!}{(K-n)! m!} m^{m-n} r^n \pi_0(K). \end{aligned}$$

**23.18 Derivation of (23.69).** This is a special case of the more general result discussed in the next problem. But let us derive the same result for this specific system G($K$)/M/$m$.

By substituting (23.64) into (23.68), we obtain

$$a_n(K) = \frac{(K-n)\pi_n(K)}{S}, \tag{38}$$

where

$$\begin{aligned}
S &= \sum_{i=0}^{m-1}(K-i)\binom{K}{i}r^i\pi_0(K) + \sum_{i=m}^{K}(K-i)\frac{K!}{(K-i)!m!}m^{m-i}r^i\pi_0(K) \\
&= \left[\sum_{i=0}^{m-1}\frac{K(K-1)!}{i!(K-i-1)!}r^i + \sum_{i=m}^{K-1}\frac{K(K-1)!}{(K-1-i)!m!}m^{m-i}r^i\right]\pi_0(K) \\
&= K\left[\sum_{i=0}^{m-1}\binom{K-1}{i}r^i + \sum_{i=m}^{K-1}\frac{(K-1)!}{(K-1-i)!m!}m^{m-i}r^i\right]\pi_0(K) \\
&= KG(K-1)\pi_0(K) = K\frac{\pi_0(K)}{\pi_0(K-1)},
\end{aligned}$$

where in deriving the second line, we used the fact $K-i=0$ for $i=K$, so we changed the summation from $i=0$ to $i=K-1$. Hence,

$$\begin{aligned}
a_n(K) &= \frac{(K-n)\binom{K}{n}r^n\pi_0(K)}{S} = \frac{K\binom{K-1}{n}r^n\pi_0(K)}{K\pi_0(K)\pi_0^{-1}(K-1)} \\
&= \binom{K-1}{n}r^n\pi_0(K-1) = \pi_n(K-1), \quad 0 \le n \le m.
\end{aligned}$$

Similarly, for $m \le n \le K$,

$$\begin{aligned}
a_n(K) &= \frac{\frac{(K-n)K!}{(K-n)!m!}m^{m-n}r^n\pi_0(K)}{S} = \frac{K\frac{(K-1)!}{(K-1-n)!m!}m^{m-n}r^n\pi_0(K)}{S} \\
&= \frac{(K-1)!}{(K-1-n)!m!}m^{m-n}r^n\pi_0(K-1) = \pi_n(K-1), \quad m \le n \le K.
\end{aligned}$$

Thus, we have proved (23.69).

**23.19 The arrival theorem.**

The assumptions made for the queue-dependent arrival rates $\lambda_j = f(K-j)$ and queue-dependent service rates $\mu_j$'s lead to a *closed* cyclic system, with $K$ customers circulating in the system. From the BD process model, we have

$$\pi_0^{-1}(K) = G(K) = \sum_{n=0}^{K}\prod_{i=0}^{n-1}\frac{f(K-i)}{\mu_{i+1}}$$

$$\pi_n(K) = \pi_0(K)\prod_{i=0}^{n-1}\frac{f(K-i)}{\mu_{i+1}}, \quad 0 \le n \le K.$$

Then from (23.68)

$$a_n(K) = \frac{f(K-n)\pi_n(K)}{S},$$

where

$$
\begin{aligned}
S &= \sum_{n=0}^{K} f(K-n)\pi_n(K) = \left[\sum_{n=0}^{K} \frac{f(K-n)f(K)f(K-1)\cdots f(K-n+1)}{\prod_{i=0}^{n-1}\mu_{i+1}}\right]\pi_0(K) \\
&= f(K)\left[\sum_{n=0}^{K-1} \frac{f(K-1)f(K-2)\cdots f(K-n)}{\prod_{i=0}^{n-1}\mu_{i+1}}\right]\pi_0(K) \\
&= f(K)\pi_0^{-1}(K-1)\pi_0(K).
\end{aligned}
$$

Here we used the assumption $f(0)=0$ to change the sum limit to $K-1$ in the second line. Note that the denominator of $a_n(K)$ can be rearranged as

$$
\begin{aligned}
f(K-n)\pi_n(K) &= f(K-n)\pi_0(K)\prod_{i=0}^{n-1}\frac{f(K-i)}{\mu_{i+1}} \\
&= \pi_0(K)f(K)\prod_{i=0}^{n-1}\frac{f(K-1-i)}{\mu_{i+1}} \\
&= \pi_0(K)f(K)\pi_0^{-1}(K-1)\pi_n(K-1).
\end{aligned}
$$

Hence,

$$
a_n(K) = \frac{\pi_0(K)f(K)\pi_0^{-1}(K-1)\pi_n(K-1)}{f(K)\pi_0^{-1}(K-1)\pi_0(K)} = \pi_n(K-1).
$$

**23.20 Derivation of the distribution (23.66).**

$$
\begin{aligned}
\pi_n(K) &= \frac{\frac{r^n}{(K-n)!}}{\sum_{i=0}^{K}\frac{r^i}{(K-i)!}} \\
&= \frac{\frac{r^{-(K-n)}}{(K-n)!}e^{-r^{-1}}}{\sum_{i=0}^{K}\frac{r^{-(K-i)}}{(K-i)!}e^{-r^{-1}}} \\
&= \frac{P(K-n;r^{-1})}{Q(K;r^{-1})}.
\end{aligned}
$$

**23.21 Limit of the machine repairman model.** We should go back to the original distribution that has led to (23.65), i.e.,

$$
\pi_n(K) = \frac{\frac{r^nK!}{(K-n)!}}{\sum_{i=0}^{K}\frac{r^iK!}{(K-i)!}}
$$

Then the numerator can be written as

$$
\frac{r^nK!}{(K-n)!} = K(K-1)\cdots(K-n+1)r^n = KR(K-1)r\cdots(K-n+1)r \to \rho^n, \ \text{as} \ K\to\infty,
$$

because for any finite $k$, $(K-k)r = KR\left(1-\frac{k}{K}\right) \to Kr = \lambda$, as $K\to\infty$.

Thus the denominator

$$\sum_{i=0}^{K} \frac{r^i K!}{(K-i)!} \to \sum_{i=0}^{K} \rho^i = (1-\rho)^{-1}.$$

Thus,

$$\pi_n(K) \to \rho^n(1-\rho), \quad \text{as } K \to \infty.$$

**23.22* The waiting time distribution in M($K$)/K/$m$.**
If an arriving customer finds only $n \le m-1$ customers in the system, it gets immediate service without waiting, i.e.,

$$F_W^c(x|n) = P[W > x | N = n] = 0, \quad 0 \le n \le m-1, \quad x \ge 0. \tag{39}$$

On the other hand, when $n \ge m$, the waiting time is given by:

$$W = R_1 + S_2 + \cdots + S_{n-m+1}, \tag{40}$$

where $R_1$ represents the residual time until the next service completion. The random variables $S_2, \cdots, S_{n-m+1}$ represent the subsequent inter-service times. Note that after $n-m+1$ service completions, the $n$th call enters service. By the memoryless property of the exponential distribution, the remaining time in service of a call currently in service is exponentially distributed. Then the time between service completions (while all servers are busy) is the minimum of $m$ i.i.d. exponentially distributed random variables with parameter $\mu$. Hence, the time between service completions is exponentially distributed with parameter $m\mu$. Therefore, $R_1$, $S_2, \ldots, S_{n-m+1}$ are i.i.d. and exponentially distributed with parameter $m\mu$. Then $W$ has an $n-m+1$-stage Erlangian distribution, i.e.,

$$F_W^c(x|n) = e^{-\mu x} \sum_{j=0}^{n-m} \frac{(\mu x)^j}{j!}, \quad m \le n \le K. \tag{41}$$

Alternatively, we can write

$$F_W^c(x|n+m) = e^{-\mu x} \sum_{j=0}^{n} \frac{(\mu x)^j}{j!} = Q(n; m\mu x), \ 0 \le n \le K-m, \tag{42}$$

which is (23.71). Therefore,

$$F_W^c(x) = \sum_{n=0}^{K-m} a_n(K) F_W^c(x|n+m). \tag{43}$$

Since $a_n(K) = \pi_n(K-1)$ for the $M(K)/M/m$ system, we have:

$$F_W^c(x) = \sum_{n=0}^{K-m} \pi_{n+m}(K-1) F_W^c(x|n+m), \tag{44}$$

which is (23.70), where $F_W^c(x|n+m)$ is given by (42). We note that since $\pi_K(K-1)=0$, the upper limit of the summation in (23.70) can be replaced by $K-m-1$, i.e.,

$$F_W^c(x) = \sum_{n=0}^{K-m-1} \pi_{n+m}(K-1) F_W^c(x|n+m). \tag{23.70'}$$

**23.23 The waiting time distribution in M($K$)/K/$m$ – continued.**

Applying the distribution of the number in system for a G($K$)/M/1 given in (23.64), we have

$$\begin{aligned}\pi_{n+m}(K-1) &= \frac{(K-1)!}{(K-n-m-1)!m!} m^{-n} r^{n+m} \pi_0(K-1)\\ &= \frac{(K-1)!}{(K-n-m-1)!} \frac{m^m}{m!} (mr^{-1})^{-(n+m)} \pi_0(K-1).\end{aligned} \tag{45}$$

Substituting this into (23.70'), we have

$$F_W^c(x) = \frac{m^m}{m!} \frac{\pi_0(K-1)}{P(K-1;mr^{-1})} \sum_{n=0}^{K-m-1} P(K-m-n-1;mr^{-1}) Q(n;\mu x). \tag{46}$$

To simplify this expression further, we first establish the following identity:

$$\sum_{j=0}^{k} P(k-j;a_1) Q(j;a_2) = Q(k;a_1+a_2). \tag{47}$$

The left-hand side (LHS) is given by

$$\begin{aligned}\text{LHS} &= \sum_{j=0}^{k} P(k-j;a_1) Q(j;a_2) = \sum_{j=0}^{k} \frac{a_1^{k-j}}{(k-j)!} e^{-a_1} \sum_{i=0}^{j} \frac{a_2^i}{i!} e^{-a_2}\\ &= \sum_{l=0}^{k} \sum_{i=0}^{l} \frac{a_1^{l-i}}{(l-i)!} \frac{a_2^i}{i!} e^{-(a_1+a_2)} = \sum_{l=0}^{k} \frac{(a_1+a_2)^l}{l!} e^{-(a_1+a_2)}\\ &= Q(k;a_1+a_2) = \text{RHS},\end{aligned} \tag{48}$$

where the variable substitution $l = i+k-j$ is made in (48). Now applying (47) in (46), we have

$$F_W^c(x) = \frac{m^m}{m!} \pi_0(K-1) \frac{Q(K-m-1;m(\mu x + r^{-1}))}{P(K-1;mr^{-1})},$$

from which (23.72) follows.

To derive (23.73), we first establish the following identity:

$$\frac{\partial}{\partial \lambda} Q(k;\lambda) = -P(k;\lambda). \tag{49}$$

The LHS is given by

$$\text{LHS} = \sum_{j=0}^{k} \frac{\partial}{\partial \lambda}\left[\frac{\lambda^j}{j!}e^{-\lambda}\right] = e^{-\lambda}\left[\sum_{j=0}^{k-1}\frac{\lambda^j}{j!} - \sum_{j=0}^{k}\frac{\lambda^j}{j!}\right]$$
$$= -e^{-\lambda}\frac{\lambda^k}{k!} = -P(k;\lambda) = \text{RHS}.$$

We also note that $F_W^c(0)$ has a jump discontinuity at $x = 0$, since there is a non-zero probability that an arriving customer does not have to wait. Setting $x = 0$ in (23.72), we have

$$F_W^c(0) = \frac{m^m}{m!}\frac{\pi_0(K-1)}{P(K-1;mr^{-1})}\frac{Q(K-m-1;mr^{-1})}{P(K-1;mr^{-1})}. \tag{50}$$

Hence,

$$F_W(0) = 1 - \frac{m^m}{m!}\frac{\pi_0(K-1)}{P(K-1;mr^{-1})}\frac{Q(K-m-1;mr^{-1})}{P(K-1;mr^{-1})}. \tag{51}$$

Next, we differentiate (23.72) with respect to $x$ for $x > 0$ and apply (49) to obtain

$$f_W(x) = -\frac{\partial F_W^c(x)}{\partial x} \tag{52}$$
$$= \frac{m^{m+1}\mu}{m!}\pi_0(K-1)\frac{P(K-m-1;m\mu(x+\nu^{-1}))}{P(K-1;mr^{-1})}. \tag{53}$$

Hence, for $x \geq 0$,

$$f_W(x) = F_W(0)\delta(x) + \frac{m^{m+1}\mu}{m!}\pi_0(K-1)\frac{P(K-m-1;m\mu(x+\nu^{-1}))}{P(K-1;mr^{-1})}, \tag{54}$$

where $F_W(0)$ is given by (51).

**23.24 PASTA (Poisson arrivals see time averages) in an M/G/1 queue.**
Since $N(t)$ is an ergodic process in a stable M/G/1 queue $\{\pi_n\}$ should also be equivalent to the long-run time average distribution.

(a) The number of arrivals in the interval $(0, T]$ is $A(T)$ is Poisson distributed with mean $E[A(T)] = \lambda T$. Let $\tau_n(T)$ be the sum of the interval during which there are $n$ customers in the system is. Then,

$$E[\tau_n(T)] = \pi_n T, \text{ and } \lim_{T\to\infty}\frac{\tau_n(T)}{T} \overset{w.p.1}{\to} \pi_n.$$

Let $A_n(T)$ be the number of customers who find exactly $n$ customers upon arrival. Then for given $\tau_n(T)$, $A_n(T)$ is Poisson distributed with mean $\lambda\tau_n(T)$.

(b) Taking the ratio of $A_n(T)$ over $A(T)$, we have

$$a_n = \lim_{T\to\infty}\frac{A_n(T)}{A(T)} = \lim_{T\to\infty}\frac{\lambda\tau_n(T)}{\lambda T} = \lim_{T\to\infty}\frac{\tau_n(T)}{T} = \pi_n.$$

**23.25 Derivation of $\pi_0$.** By setting $z = 1$ in the expression for $P(z)$,

$$1 = \frac{\pi_0 f_S^*(0)}{1 + \lambda f_S^{*\prime}(0)} = \frac{\pi_0}{1 - \lambda E[S]}, \tag{55}$$

where we used the relation $f_S^*(0) = \int_0^\infty f_S(t)\,dt = 1$. The first derivative of $f_S^*(s)$

$$f_S^{*\prime}(s) = \int_0^\infty (-t) f_S(t) e^{-st}\,dt \tag{56}$$

evaluated at $s = 0$ is $-E[S]$. From (56) we find (23.83).

**23.26 M/G/1 with hyper-exponential service time [203].**

The mean service time is given by $1/\mu = p_1/\mu_1 + p_2/\mu_2$. The LT of the pdf $f_S(t) = (d/dt)F_S(t)$ is

$$f_S^*(s) = \frac{p_1\mu_1}{\mu_1 + s} + \frac{p_2\mu_2}{\mu_2 + s}.$$

Then from (23.82), we have

$$A(z) = f_S^*(\lambda - \lambda z) = \frac{p_1}{1 + \rho_1(1-z)} + \frac{p_2}{1 + \rho_2(1-z)},$$

where

$$\rho_i = \lambda/\mu_i, \;\; i = 1, 2.$$

Therefore, the PGF of the probability distribution $\{\pi_n\}$ is

$$P(z) = \frac{(1-\rho)\,[1 + (\rho_1 + \rho_2 - \rho)(1-z)]}{\rho_1\rho_2 z^2 - (\rho_1 + \rho_2 + \rho_1\rho_2)z + 1 + \rho_1 + \rho_2 - \rho}, \tag{57}$$

where

$$\rho = p_1\rho_1 + p_2\rho_2$$

is the traffic intensity.

The probability distribution $\{\pi_n\}$ is obtained, e.g., through partial fraction expansion of $P(z)$. This requires us to solve for the roots of the characteristic equation obtained by setting the denominator of (57) to zero. If $z_1$ and $z_2$ are the characteristic roots, $P(z)$ has the following representation:

$$P(z) = \frac{C_1 z_1}{z_1 - z} + \frac{C_2 z_2}{z_2 - z}, \tag{58}$$

where $C_1$ and $C_2$ are determined by setting $z = 0$ and $z = 1$ in $P(z)$: $P(1) = 1$ and $P(0) = \pi_0 = 1 - \rho$. Thus, we find

$$C_1 = \frac{(z_1 - 1)(1 - \rho z_2)}{z_1 - z_2} \;\text{ and }\; C_2 = \frac{(z_2 - 1)(1 - \rho z_1)}{z_2 - z_1}. \tag{59}$$

Finally we obtain the equilibrium-state distribution of $N(t)$:

$$\pi_n = C_1 z_1^{-n} + C_2 z_2^{-n}.$$

Alternatively, we can derive a recurrence formula to compute the sequence $\{\pi_n\}$ starting with the initial condition $\pi_0 = 1 - \rho$.

**23.27 M/G/1 with Erlangian service time distribution [203].**

The LT of the pdf is given by

$$f_S^*(s) = \left(1 + \frac{s}{k\mu}\right)^{-k}. \tag{60}$$

Thus, we readily obtain

$$A(z) = f_S^*(\lambda - \lambda z) = \left(1 + \frac{\rho(1-z)}{k}\right)^{-k}$$

and

$$P(z) = \frac{(1-\rho)(z-1)}{z\left(1 + \frac{\rho(1-\rho)}{k}\right)^k - 1}. \tag{61}$$

For a small value of $k$, the PGF $P(z)$ can be inverted by the partial fraction method. The recursion method can be used to find the solution $\{\pi_n\}$ even for large $k$.
We will derive below, instead, a closed form expression for $\{\pi_n\}$. By defining the dimensionless parameters

$$R = 1 + \frac{\rho}{k} \text{ and } r = \frac{\rho}{k+\rho} = 1 - R^{-1},$$

we can write $P(z)$ as

$$P(z) = \frac{(1-\rho)(1-z)}{1 - z[R(1-rz)]^k} = (1-\rho)(1-z)\sum_{j=0}^{\infty} z^j R^{kj}(1-rz)^{kj}. \tag{62}$$

By equating the coefficients of the terms $z^n$ on both sides of the equation, we find

$$\pi_n = (1-\rho)\sum_{j=0}^{n}(-1)^{n-j} r^{n-j-1} R^{kj}\left[\binom{kj}{n-j} r + \binom{kj}{n-j-1}\right]. \tag{63}$$

In Figure 23.7, we plot the distributions $\{\pi_n\}$ for the cases $k = 1, 2,$ and $4$ with traffic intensity $\rho = 0.75$. By letting $k$ approach infinity, we can obtain the solution for the M/D/1 system discussed in Example 23.1. The solution (63) should converge to that of M/D/1, as $k$ tends to $\infty$.

**23.28 Derivation of Pollaczek–Khintchine formula.**Take the natural logarithm of (23.84) and differentiate with respect to $z$ to obtain

$$\begin{aligned}\frac{P'(z)}{P(z)} &= \frac{1}{z-1} - \frac{\lambda f_S^{*\prime}(\lambda - \lambda z)}{f_S^*(\lambda - \lambda z)} - \frac{1 + \lambda f_S^{*\prime}(\lambda - \lambda z)}{z - f_S^*(\lambda - \lambda z)} \\ &= -\frac{\lambda f_S^{*\prime}(\lambda - \lambda z)}{f_S^*(\lambda - \lambda z)} + \frac{1 - f_S^*(\lambda - \lambda z) - (z-1)\lambda f_S^{*\prime}(\lambda - \lambda z)}{(z-1)[z - f_S(\lambda - \lambda s)]}.\end{aligned}$$

Letting $z \to 1$ and using l'Hospital's rule in the second term of the right-hand side, we get

$$\begin{aligned}
P'(1) &= -\lambda f_S^{*\prime}(0) + \lim_{z\to 1} \frac{\lambda^2(z-1) f_S^{*\prime\prime}(\lambda - \lambda z)}{z - f_S^*(\lambda-\lambda z) + (z-1)[1+\lambda f_S^{*\prime}(\lambda-\lambda z)]} \\
&= \lambda E[S] + \lim_{z\to 1} \frac{\lambda^2 f_S^{*\prime\prime}(\lambda-\lambda z)}{\frac{z - f_S^*(\lambda-\lambda z)}{z-1} + 1 + \lambda f_S^{*\prime}(\lambda - \lambda z)} \\
&= \lambda E[S] + \frac{\lambda^2 f_S^{*\prime\prime}(0)}{z[1+\lambda f_S^{*\prime}(0)]} = \lambda E[S] + \frac{\lambda^2 E[S^2]}{2(1-\lambda E[S])}.
\end{aligned}$$

**23.29 Alternative derivation of Pollaczek–Khintchine formula [203].**

(a) Taking expectations of (23.76), we get

$$E[N_k] = E[N_{k+1}] - E[U(N_{k-1})] + E[A_k]. \tag{64}$$

Letting $k \to \infty$, we have

$$\begin{aligned}
\lim_{k\to\infty} E[U(N_{k-1})] &= \lim_{k\to\infty} P[N_{k-1} > 0] = P[\text{server is busy}], \\
\lim_{k\to\infty} E[A_k] &= A'(1) = \lambda E[S],
\end{aligned}$$

where the last equation can be obtained from (23.82). Thus, in the limit as $k \to \infty$, (64) yields

$$P[\text{server is busy}] = \lambda E[S] = \rho.$$

(b) Squaring both sides of (64), and then taking expectations yields

$$\begin{aligned}
E[N_k^2] &= E[N_{k-1}^2] + E[U^2(N_{k-1})] + E[A_k^2] - 2E[N_{k-1}U(N_{k-1})] \\
&\quad - 2E[A_k]E[U(N_{k-1})] + 2E[N_{k-1}]E[A_k] \qquad (65) \\
&= E[N_{k-1}^2] + E[U(N_{k-1})] + E[A_k^2] - 2E[N_{k-1}] \\
&\quad - 2E[A_k]E[U(N_{k-1})] + 2E[N_{k-1}]E[A_k]. \qquad (66)
\end{aligned}$$

Also note that

$$\lim_{k\to\infty} E[A_k^2] = A''(1) = \rho + \lambda^2 E[S^2].$$

Taking the limit as $k \to \infty$ in (66), and using the result of part (a) we have

$$0 = \rho + \rho + \lambda^2 E[S^2] - 2E[N] - 2\rho^2 + 2E[N]\rho.$$

Hence,

$$E[N] = \rho + \frac{\lambda^2 E[S^2]}{2(1-\rho)}.$$

**23.30 Derivation of (23.98).** As discussed in the section

$$A_k = \text{Number of arrivals during the service time } S_k.$$

From (23.82), we know that the PGF $A(z)$ of $\{A_k\}$ is related to the LT $f_S^*(s)$ of the service time according to

$$A(z) = f_S^*(\lambda(1-z)).$$

Since we have a parallel relationship between $N_k$ and $T_k$, i.e.,

$$N_k = \text{Number of arrivals during the system time } T_k,$$

we can immediately infer that the PGF $P(z)$ of $\{N_k\}$, is related to the LT $f_T^*(s)$ of the system time according to

$$P(z) = f_T^*(\lambda(1-z)),$$

which is (23.98).

**23.31* Derivation of waiting time distribution (23.103).**

$$f_W^*(s) = \frac{1-\rho}{1-\lambda\frac{1-f_S^*(s)}{s}}.$$

By substituting

$$\frac{1-f_S^*(s)}{s} = E[S]f_R^*(s),$$

we find the desired expression for $f_W^*(s)$.

**23.32 System time distribution of an M/H$_2$/1 queue.**

From (58) and (59) of the solution to Problem 23.26, we have for the M/H$_2$/1 queue:

$$P(z) = \frac{C_1 z_1}{z_1 - z} + \frac{C_2 z_2}{z_2 - z}, \tag{67}$$

where

$$C_1 = \frac{(z_1-1)(1-\rho z_2)}{z_1 - z_2} \text{ and } C_2 = \frac{(z_2-1)(1-\rho z_1)}{z_2 - z_1}. \tag{68}$$

By setting $\lambda - \lambda z = s$ in (4.5-53), we have

$$f_T^*(s) = P\left(1 - \frac{s}{\lambda}\right) = \frac{C_1 z_1}{z_1 - 1 + \frac{s}{\lambda}} + \frac{C_2 z_2}{z_2 - 1 + \frac{s}{\lambda}} = \frac{p_\alpha \alpha}{s+\alpha} + \frac{p_\beta \beta}{s+\beta},$$

where

$$\alpha = \lambda(z_1 - 1), \ \ \beta = \lambda(z_2 - 1),$$
$$p_\alpha = \frac{C_1 z_1}{z_1 - 1}, \ \ p_\beta = \frac{C_2 z_2}{z_2 - 1}.$$

Using (68), we can show that $p_\alpha + p_\beta = 1$. By inverting $f_T^*(s)$, we obtain

$$f_T(t) = p_\alpha \alpha e^{-\alpha t} + p_\beta \beta e^{-\beta t}, \ \ t \geq 0,$$

and

$$F_T(t) = 1 - p_\alpha e^{-\alpha t} - p_\beta e^{-\beta t}, \ \ t \geq 0.$$

**23.33 Conditional mean system time under FCFS.**

From (23.97) the expected waiting time under FCFS is given by (23.96). Then for a customer with service time $Sx$, its expected system time is

$$E[T_{\text{FCFS}}|S] = S + E[W|S] = S + E[W_{\text{FCFS}}],$$

because $S$ is independent of $W$ under FCFS. Thus, (23.151) follows.

**23.34 Comparison of $\overline{T}_{\mathrm{PS}}$ and $\overline{T}_{\mathrm{FCFS}}$.**

We have from (23.151)

$$E[T_{\mathrm{FCFS}}(S)|S] = S + \frac{\lambda E[S^2]}{2(1-\rho)}.$$

and

$$E[T_{\mathrm{PS}}(S)|S] = \frac{S}{(1-\rho)}.$$

Thus, Writing $E[S] = \overline{S}$, $E[S^2] = \overline{S}^2 + \sigma_S^2$, and $\overline{S} = \rho/\lambda$, we have

$$\overline{T}_{\mathrm{FCFS}} - \overline{T}_{\mathrm{PS}} = \frac{\rho}{\lambda}\left[1 - \frac{1}{1-\rho} + \frac{\rho(1+c_S^2)}{2(1-\rho)}\right],$$

where $c_S = \frac{\sigma_S}{\overline{S}}$ If $\sigma_S \geq 1$, then

$$\overline{T}_{\mathrm{FCFS}} - \overline{T}_{\mathrm{PS}} \geq \overline{S}\left[1 - \frac{1}{1-\rho} + \frac{\rho}{1-\rho}\right] = 0,$$

When $c_M < 1$, the opposite inequality holds.

## 23.4 Loss Models

**23.35 PASTA in the Erlang loss model.**

(a) $\pi_n T$ is the period when there are $n$ customers in the system. During this interval there are, on the average, $\lambda\pi_n T$ arrivals and they find exactly $n$ customers in the system. Hence,

$$a_n = \frac{\lambda\pi_n T}{\sum_{n=0}^{m}\lambda\pi_n T} = \pi_n,$$

since $\sum_{n=0}^{m}\pi_n = 1$.

(b) If you observe $N(t)$ over a sufficiently long interval $T$, then the number of "up jumps" from level $n$ to $n+1$ and the number of "down jumps" from level $n+1$ to $n$. Such an "up jump" occurs when a customer arrives to find $n$ customers in the system. Similarly such a "down jump" occurs when a customer departs the system, leaving $n$ customers behind him/her. Thus,

$$a_n = \frac{\text{number of arrivals who see } n \text{ customers}}{\text{Total number of arrivals in } (0,T)}$$

and

$$d_n = \frac{\text{number of departures who leave } n \text{ customers behind}}{\text{Total number of departures in } (0,T)}$$

are equal as $T \to \infty$, since the ratio of the total numbers of arrivals and departures in $(0,T)$ converges to unity as $T \to \infty$. Hence $\{a_n\} = \{d_n\}$. For the Poisson arrival, we know $\{a_n\} = \{\pi_n\}$ (i.e., PASTA). Hence $\{d_n\} = \{\pi_n\}$.

(c) This question should be postponed until the next section on Engset model, as an example of non-Poisson arrivals. It turns out rather difficult to give a specific counter example other than the state dependent arrival case as in the Engset model. Syski [515] discusses a loss system D/M/$m(0)$, but it is rather complicated to discuss here.

**23.36 Erlang loss model and the machine repairman model.**

(a) The diagram asked to draw is basically similar to the multiple access model of Figure 23.5, except that $K$ parallel sources (or $K$ machines that are up and running) is now replaced by $m$-parallel servers or by an infinite-server (IS) station, and the time spent in IS is generally distributed, not restricted to an exponential distribution. Since there are only $m$ customers or machines are circulating around in this closed system, it makes no difference whether the number of parallel servers is $m$, $m+1$ or infinite.

(b) The single server (i.e., the repairman) is idle only when all the $m$ machines are up and running (i.e., in the IS station). Hence, $m$ servers out of infinite servers are busy.

(c) Consider a loss model M/G/$m$(0) with Poisson arrivals of rate $\mu$ and general service times of mean $\nu^{-1}$. Those arriving calls that find all the $m$ servers (or lines) busy will be lost. Those calls that find $m-1$ or less lines busy will enter service immediately, and these accepted calls form an IPP. This IPP stream is equivalent to the completed from the exponential server in the G($m$)/M/1.

(d) The probability that $n$ out of $m$ machines are in service at the exponential server or its queue of G($m$)/M/1 is given from (23.66)

$$\pi_n(m) = \frac{P(m-n, r^{-1})}{Q(m, r^{-1})}.$$

This is the probability that $m - n \triangle n'$ machines are up and running (i.e., in the IS station side). In the corresponding Erlang loss model, the probability that $n'$ calls are in service is given from (23.116) by

$$\pi_{n'} = \frac{P(n', a)}{Q(n', a)},$$

where $n' = m - n$ and $a = \frac{\mu}{\nu} = r^{-1}$. Clearly these two probabilities are equivalent.

**23.37 Recursive computation of Erlang's loss formula.** We proceed as follows:

$$\frac{1}{B(m,a)} = \frac{\sum_{j=0}^{m} \frac{a^j}{j!}}{\frac{a^m}{m!}} = \frac{\frac{a^m}{m!} + \sum_{j=0}^{m-1} \frac{a^j}{j!}}{\frac{a}{m} \cdot \frac{a^{m-1}}{(m-1)!}} = 1 + \frac{m}{a} \frac{1}{B(m-1,a)}.$$

Taking the reciprocal of both sides, we have:

$$B(m,a) = \frac{aB(m-1,a)}{m + aB(m-1,a)}$$

**23.38 Bounds for Erlang's loss formula [155]**

(a) Clearly

$$\sum_{i=0}^{m} \frac{a^i}{i!} < \sum_{i=0}^{\infty} \frac{a^i}{i!} = e^a.$$

Hence the lower bound of $B(m,a)$ is immediate. The inequality of RHS is perhaps the most difficult part of this exercise, but can be proved as follows. Let

$$S = \sum_{i=m}^{\infty} \frac{a^i}{i!} e^{-a}.$$

Then clearly

$$1 > S > \frac{a^m}{m!} e^{-a},$$

which implies

$$S(1-S) > (1-S)\frac{a^m}{m!} e^{-a},$$

which in turn implies

$$\frac{\frac{a^m}{m!} e^{-a}}{1 - S + \frac{a^m}{m!} e^{-a}} < S.$$

It is easy to see the LHS is $B(m,a)$. Hence the inequality of RHS in the question has been proved.

(b) We write

$$i! = i(i-1)\cdots(m+1)m! > m^{i-m} m!,$$

which leads to the inequality to be proved.

(c) By letting $j \triangleq i - m$ in the inequality (b), and sum from $i = m$ (hence $j = 0$) to $\infty$.

$$\sum_{i=m}^{\infty} \frac{a^i}{i} \le \sum_{j=0}^{\infty} \left(\frac{a}{m}\right)^j \frac{a^m}{m!} e^{-a},$$

which results in the RHS of (23.153).

**23.39 TDMA-based digital cellular network [203].**

(a) The set of feasible states is

$$\mathcal{F}_{\boldsymbol{N}}(m, \boldsymbol{K}) = \{\boldsymbol{n} \ge \boldsymbol{0} : n_1 + 2n_2 \le m, n_1 \le K_1, n_2 \le K_2\},$$

where $\boldsymbol{K} = (K_1, K_2)$.

(b) We can formulate this problem as a GLS with $A_1 = 1$ and $A_2 = 2$. The stationary distribution of $\boldsymbol{N}(t)$ is given by

$$\pi(\boldsymbol{n}) = \lim_{t\to\infty} P[\boldsymbol{N}(t) = \boldsymbol{n}] = \frac{1}{G(m, \boldsymbol{K})} \binom{K_1}{n_1} r_1^{n_1} \binom{K_2}{n_2} r_2^{n_2}, \;\; \boldsymbol{n} \in \mathcal{F}_{\boldsymbol{N}}(m, \boldsymbol{K}),$$

where $r_r = \nu_r/\mu_r, r = 1, 2$, and the normalization constant is

$$G(m, \boldsymbol{K}) = \sum_{\boldsymbol{n}\in\mathcal{F}_{\boldsymbol{N}}(m,\boldsymbol{K})} \binom{K_1}{n_1} r_1^{n_1} \binom{K_2}{n_2} r_2^{n_2}.$$

(c)

$$\mathcal{F}(4, \boldsymbol{K}) = \{(0,0), (0,1), (0,2), (1,0), (1,1), (2,0), (2,1), (3,0)\}.$$

and

$$G(4, \boldsymbol{K}) = 1 + 2 \cdot \frac{2}{3} + \left(\frac{2}{3}\right)^2 + 3 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} \cdot 2 \cdot \frac{2}{3} + 3 \cdot \left(\frac{1}{2}\right)^2 + 3 \cdot \left(\frac{1}{2}\right)^2 \cdot 2 \cdot \frac{2}{3} + \left(\frac{1}{2}\right)^3 = \frac{587}{72}.$$

$\mathcal{F}(5, \boldsymbol{K}) = \mathcal{F}(4, \boldsymbol{K}) \cup \{(1,2), (3,1)\}$, and $G(5, \boldsymbol{K}) = G(4, \boldsymbol{K}) + 3 \cdot \frac{1}{2} \cdot \left(\frac{2}{3}\right)^2 + \left(\frac{1}{2}\right)^3 \cdot 2 \cdot \frac{2}{3} = G(4, \boldsymbol{K}) + \frac{5}{6} = \frac{647}{72}$. Similarly, $\mathcal{F}(6, \boldsymbol{K}) = \mathcal{F}(5, \boldsymbol{K}) \cup \{(2,2)\}$ and $G(6, \boldsymbol{K}) = G(5, \boldsymbol{K}) + \frac{1}{3} = 671/72$.

**23.40 TDMA-based digital cellular network—continued. [203].**

(a) Now that we have found the normalization constants $G(m, \boldsymbol{K})$ for $m = 4, 5$, and 6, we can use the formula (23.142) to find the blocking probabilities:

$$B_1(6, \boldsymbol{K}) = 1 - \frac{G(5, \boldsymbol{K})}{G(6, \boldsymbol{K})} = \frac{24}{671} = 0.0358,$$
$$B_2(6, \boldsymbol{K}) = 1 - \frac{G(4, \boldsymbol{K})}{G(6, \boldsymbol{K})} = \frac{84}{671} = 0.1252.$$

(b) We need to find the normalization constant with $\boldsymbol{K} - \boldsymbol{e}_1 = (2,2)$ and $\boldsymbol{K} - \boldsymbol{e}_2 = (3,1)$.

$$\mathcal{F}(5, \boldsymbol{K} - \boldsymbol{e}_1) = \{(0,0), (0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,0)\},$$
$$G(5, \boldsymbol{K} - \boldsymbol{e}_1) = 1 + 2 \cdot \frac{2}{3} + \left(\frac{2}{3}\right)^2 + \cdots + \left(\frac{1}{2}\right)^2 \cdot 2 \cdot \frac{2}{3} = \frac{221}{36},$$
$$\mathcal{F}(6, \boldsymbol{K} - \boldsymbol{e}_1) = \mathcal{F}(5, \boldsymbol{K} - \boldsymbol{e}_1) \cup \{(2,2)\},$$
$$G(6, \boldsymbol{K} - \boldsymbol{e}_1) = G(5, \boldsymbol{K} - \boldsymbol{e}_1) + \left(\frac{1}{2}\right)^2 \left(\frac{2}{3}\right)^2 = \frac{221}{36} + \frac{2}{9} = \frac{229}{36}.$$

Hence, we find the call congestion for class-1 calls:

$$L_1(6, \boldsymbol{K}) = B_1(6, \boldsymbol{K} - \boldsymbol{e}_1) = 1 - \frac{G(5, \boldsymbol{K} - \boldsymbol{e}_1)}{G(6, \boldsymbol{K} - \boldsymbol{e}_1)} = \frac{2/9}{229/36} = \frac{8}{229} \approx 0.0349,$$

which is slightly smaller than the blocking probability $B_1(6, \boldsymbol{K}) = 0.0358$.
To compute the call congestion for class-2 calls, we find the feasible states and normalization constant with $\boldsymbol{K} - \boldsymbol{e}_2 = (3,1)$:

$$\mathcal{F}(4, \boldsymbol{K} - \boldsymbol{e}_2) = \{(0,0), (0,1), (1,0), (1,1), (2,0), (2,1), (3,0)\},$$
$$G(4, \boldsymbol{K} - \boldsymbol{e}_2) = 1 + \frac{2}{3} + 3 \cdot \frac{1}{2} + \cdots + \left(\frac{1}{2}\right)^3 = \frac{133}{24},$$
$$\mathcal{F}(6, \boldsymbol{K} - \boldsymbol{e}_2) = \mathcal{F}(5, \boldsymbol{K} - \boldsymbol{e}_2) = \mathcal{F}(4, \boldsymbol{K} - \boldsymbol{e}_2) \cup \{(3,1)\},$$
$$G(6, \boldsymbol{K} - \boldsymbol{e}_2) = G(4, \boldsymbol{K} - \boldsymbol{e}_2) + \left(\frac{1}{2}\right)^3 \cdot \frac{2}{3} = \frac{133}{24} + \frac{1}{12} = \frac{135}{24}.$$

Thus, we finally obtain

$$L_2(6, \boldsymbol{K}) = B_2(6, \boldsymbol{K} - \boldsymbol{e}_2) = 1 - \frac{G(4, (3,1))}{G(6, (3,1))} = \frac{1/12}{135/24} = \frac{2}{135} \approx 0.0148, \qquad (69)$$

which is almost one order of magnitude smaller than the blocking probability $B_2(6, \boldsymbol{K}) = 0.125$. This disparity between $L_2$ and $B_2$ can be explained as follows: the time congestion corresponds to the probability that the system state $\boldsymbol{N}(t) = (N_1(t), N_2(t))$ is in $(2,2)$, $(1,2)$, or $(3,1)$. Thus,

$$B_2(6,\boldsymbol{K}) = \pi(2,2)+\pi(1,2)+\pi(3,1) = \frac{1}{G(6,\boldsymbol{K})}\left(\frac{1}{3}+\frac{2}{3}+\frac{1}{6}\right) = \frac{7/6}{671/72} = \frac{84}{671}.$$

When the system is in state $(2,2)$ or $(1,2)$, however, there is no new class-2 call coming, since all $K_2 = 2$ users are already in service. When the system is in state (3, 1), there is one class-2 user who will generate a call at rate $\nu_2 = 1/3$. Thus, the arrival rate of class-2 calls to be lost is given as

$$\pi(3,1)\nu_2 = \frac{1/6\cdot\nu_2}{G(6,\boldsymbol{K})}.$$

The total arrival rate of class-2 calls is given by

$$\begin{aligned}
&2\nu_2 P[n_2=0] + \nu_2 P[n_2=1] \\
&= 2\nu_2[\pi(0,0)+\pi(1,0)+\pi(2,0)+\pi(3,0)] + \nu_2[\pi(0,1)+\pi(1,1)+\pi(2,1)+\pi(3,1)] \\
&= \frac{\nu_2}{G(6,\boldsymbol{K})}\left[\left(1+\frac{3}{2}+\frac{3}{4}+\frac{1}{8}\right)2+\left(\frac{4}{3}+2+1+\frac{1}{6}\right)\right] = \frac{45}{4}\frac{\nu_2}{G(6,\boldsymbol{K})}.
\end{aligned}$$

By taking the ratio of the last two equations, we find the call congestion of class-2 call:

$$L_2(6,\boldsymbol{K}) = \frac{1/6}{45/4} = \frac{2}{135},$$

which agrees with (69).

**23.41* Differential-difference equation for the Engset model [203].**

(a) Let $N(t)$ be the number of calls in progress at time $t$: $0 \le N(t) \le m$. This process is a BD process with $\lambda_n$ and $\mu_n$ given by (23.60) and (23.111). Then the differential-difference equations for $p_n(t;K)$ is the same as those for $p_n(t)$ given by (14.45), where $n = 1,2,3,\ldots,m$ and $p_n = 0$ for $n \ge m+1$.

(b) The balance equations in the steady state are given by (14.51), i.e.,

$$n\mu\pi_n(K) = (K-n)\nu\pi_{n-1}, \text{ for all } n = 1,2,\ldots,m.$$

Thus, from the above (or from (14.53)), we have

$$\pi_n(K) = \pi_0(K)\prod_{i=0}^{n-1}\frac{(K-i)\nu}{(i+1)\mu} = \pi_0(K)\binom{K}{n}.$$

Thus, the normalization constant $G$ is given by (23.124) and $\pi_n(K) = G^{-1}\binom{K}{n}$, hence we obtain (23.125).

(c) We have

$$P[A|B_i] = (K-i)\nu\delta t, \;\; P[B_i] = \pi_i(K).$$

Hence,

$$P[A] = \sum_{i=0}^{m} P[A,B_i] = \nu\delta t\sum_{i=0}^{m}(K-i)\pi_i(K).$$

Hence,

$$a_n(K) = P[B_n|A] = \frac{P[A|B_n]P[B_n]}{P[A]} = \frac{(K-n)\pi_n(K)}{\sum_{i=0}^{m}(K-i)\pi_i(K)}.$$

Substitution of the result of (b) (or (23.125)) leads to (23.130).

(d) If we keep $K\nu = \text{constant} \triangleq \lambda$, then in the limit $K \to \infty$, the Engset distribution of (23.125) converges to the Erlang distribution (23.115).

**23.42 Engset model without restriction $K > m$ [203].**

(a) The birth and death rates are given by

$$\lambda_n = (K-n)\nu, \;\; n = 0, 1, \ldots, \min\{m, K\},$$

and

$$\mu_n = n\mu, \;\; n = 0, 1, \ldots, \min\{k, K\}.$$

Then the normalization constant is given by

$$G = \sum_{i=0}^{\min\{m,K\}} \binom{K}{i} r^i,$$

Thus, we have

$$\pi_n(K) = \frac{\binom{K}{n} r^n}{\sum\limits_{i=0}^{\min\{m,K\}} \binom{K}{i} r^i}, \;\; 0 \le n \le \min\{m, K\} \tag{70}$$

(b) When a customer arrives, he or she finds at most $\min\{m, K-1\}$ customers in service at the $m$ parallel servers. Thus,

$$a_n(K) = \frac{\binom{K-1}{n} r^n}{\sum\limits_{i=0}^{\min\{m,K-1\}} \binom{K-1}{i} r^i} = \pi_n(K-1), \text{ for } 0 \le n \le \min\{K-1, m\}.$$

**23.43* Link efficiency [203].**

(a) The formula (23.122) can apply to the Engset model, i.e.,

$$L(K) = 1 - \frac{a_c}{a}, \;\; \text{or } a_c = a(1 - L(K)).$$

Substituting this into (23.132), we obtain (23.154).

(b)

**23.44* Example of MLN [203].**

(a)

$$r_1 = \frac{\nu_1}{\mu_1} = 0.5 \text{ [erl] per class-1 subscriber, } \; a_2 = \frac{\lambda_2}{\mu_2} = 0.5 \text{ [erl] for the entire class 2.}$$

(b) Equation (23.137) reduces in this case to

$$\pi_{\boldsymbol{N}}(\boldsymbol{n}|m, K_1) = \frac{1}{G(m, K_1)} \binom{K_1}{n_1} r_1^{n_1} \frac{a_2^{n_2}}{n_2!}, \;\; \boldsymbol{n} \in \mathcal{F}_{\boldsymbol{N}}(m, K_1), \tag{71}$$

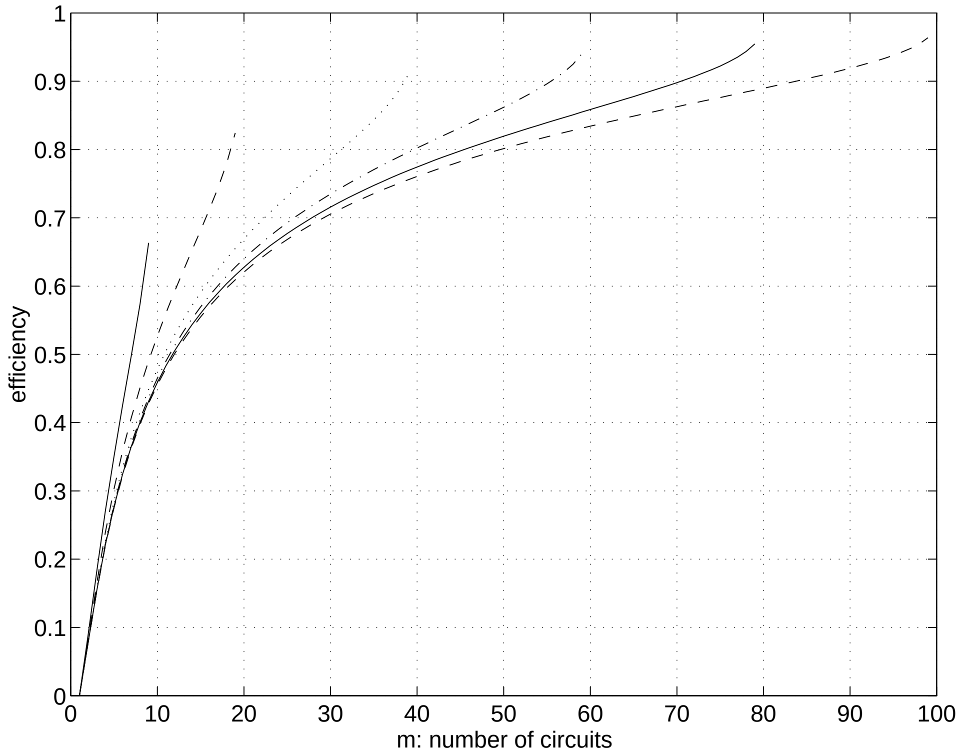



Figure Exercise 3.2-8: The efficiency $\eta$ vs. the number of circuits (output lines) $m$. The number of input lines (sources) is $K = 10, 20, 40, 60, 80, 100$, and the specified QoS is $L = 0.01$.

where

$$\mathcal{F}_{\boldsymbol{N}}(m, K_1) = \{\boldsymbol{n} = (n_1, n_2) \geq (0, 0) : n_1 + 2n_2 \leq m, n_1 \leq K_1\}$$

and

$$G(m, K_1) = \sum_{\boldsymbol{n} \in \mathcal{F}_{\boldsymbol{N}}} (m, K_1) \binom{K_1}{n_1} \frac{a_2^{n_2}}{n_2!}.$$

(c) Start with $m = 0$. Obviously, $\mathcal{F}_{\boldsymbol{N}}(0, 3) = \{(0, 0)\}$, and $G(0, 3) = 1$. For $m = 1$, we find $\mathcal{F}_{\boldsymbol{N}}(1, 3) = \{(0, 0), (1, 0)\}$ and $G(1, 3) = 1 + \binom{3}{1}\frac{1}{2} = \frac{5}{2}$. By proceeding in a similar manner, we find the feasible set for $m = 4, K_1 = 3$:

$$\mathcal{F}_{\boldsymbol{N}}(4, 3) = \{(0, 0), (1, 0), (2, 0), (0, 1), (3, 0), (1, 1), (2, 1), (0, 2)\},$$

and the corresponding normalization constant: $G(4, 3) = \frac{41}{8} = 5.125$.

For $m = 5$,

$$\mathcal{F}_{\boldsymbol{N}}(5,3) = \mathcal{F}_{\boldsymbol{N}}(4,3) \cup \{(3,1),(1,2)\},$$

$$G(5,3) = G(4,3) + \binom{3}{3}\left(\frac{1}{2}\right)^3 \frac{1}{2} + \binom{3}{1}\frac{1}{2}\frac{(1/2)^2}{2!} = \frac{41}{8} + \frac{1}{4} = \frac{43}{8} = 5.375.$$

For $m = 6$,

$$\mathcal{F}_{\boldsymbol{N}}(6,3) = \mathcal{F}_{\boldsymbol{N}}(5,3) \cup \{(2,2),(0,3)\},$$

$$G(6,3) = G(5,3) + \binom{3}{2}\left(\frac{1}{2}\right)^2 \frac{1}{2}\frac{(1/2)^2}{2!} + \frac{(1/2)^3}{3!} = \frac{43}{8} + \frac{11}{96} = \frac{527}{96} = 5.486,$$

to three decimal places.

(d) Using (23.141), we find

$$B_2(6,3) = 1 - \frac{G(4,3)}{G(6,3)} = \frac{4/16 + 11/96}{527/96} = \frac{35}{527} = 0.0664,$$

$$L_2(6,3) = B_2(6,3) = 0.0664.$$

(e) We need to find $\mathcal{F}_{\boldsymbol{N}}(m, K_1)$ and $G(m, K_1)$ for $K_1 = 2$. Since $\mathcal{F}_{\boldsymbol{N}}(m,2) \subset \mathcal{F}_{\boldsymbol{N}}(m,3)$, this does not really require an additional effort: it is easy to find

$$\mathcal{F}_{\boldsymbol{N}}(5,2) = \{(0,0),(1,0),(2,0),(0,1),(1,1),(2,1),(0,2),(1,2)\},$$

$$G(5,2) = 1 + 1 + \frac{1}{4} + \frac{1}{2} + \frac{1}{2} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{29}{8} = 3.625,$$

and

$$\mathcal{F}_{\boldsymbol{N}}(6,2) = \mathcal{F}_{\boldsymbol{N}}(5,2) \cup \{(2,2),(0,3)\},$$

$$G(6,2) = G(5,2) + \binom{2}{2}\left(\frac{1}{2}\right)^2 \frac{(1/2)^2}{2!} + \frac{(1/2)^3}{3!} = \frac{29}{8} + \frac{5}{96} = \frac{353}{96} = 3.677.$$

Then, from the formula (23.142) we find

$$L_1(6,3) = B_1(6,2) = 1 - \frac{G(5,2)}{G(6,2)} = \frac{5/96}{353/96} = \frac{5}{353} = 0.0142.$$

**Remarks:** It will be instructive to make the following observations: the time congestion for class-1 customers (in the closed chain) occurs when the GLS is in states $(n_1, n_2) = (0,3)$ or $(2,2)$. The stationary probabilities of these states are found from (71) as

$$\pi_{\boldsymbol{N}}((0,3)|6,3) = \frac{1/48}{G(6,3)} \text{ and } \pi_{\boldsymbol{N}}((2,2)|6,3) = \frac{3/32}{G(6,3)}.$$

By adding these probabilities, we find $B_1(3,6) = \frac{1/48+3/32}{527/96} = \frac{11}{527} = 0.0209$, as was obtained above.

Similarly, time congestion for class-2 customers (in the open route) occurs when the GLS is in one of the following four states: $(0,3),(2,2),(1,2),(3,1)$. By adding the stationary probabilities of these states, we have

$$B_2(3,6) = \frac{1/48 + 3/32 + 3/16 + 1/16}{G(6,3)} = \frac{35}{527} = 0.0664,$$

as expected.