Probability, Random Processes, and Statistical Analysis

Reader's Solution Manual (starred problems only)

(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 $n = 2, 3, 4, 5, \ldots$.

2.5* 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 Ω. 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

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

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)

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

2.3 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 ω does *not* belong to *both* A and B. This implies that either ω belongs to A^c or ω 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 ω belongs to either A^c or B^c. In the former case, ω does not belong to A, which implies that $\omega \notin A \cap B$. In the latter case, ω 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 $(A \cap B)^c = A^c \cup B^c$.

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

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

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:

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

Returning to (1), we have

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

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,

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

We are also given that

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

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

3 Solutions for Chapter 3: Discrete Random Variables

3.1 Random Variables

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

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

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

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

or

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

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

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

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)

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

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

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

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

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.

3.10* 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

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

which is equivalent to

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

where the equality holds iff

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

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

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

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

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

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

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

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).

3.18* 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)

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

3.20* 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

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

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

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

(b) Using the hint, we have

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

From this recursive relation, we have

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

(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

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

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

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

from which we can derive

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

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

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

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

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

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

$$\sum_{i=r}^{k} \binom{k}{i} p^i q^{k-i} \stackrel{?}{=} \binom{k-1}{k-r} p^r q^{k-r} + \sum_{i=r}^{k-1} \binom{k-1}{i} p^i q^{k-1-i}. \tag{7}$$

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

$$\text{LHS} = (p+q)^k - \sum_{i=0}^{r-1} \binom{k}{i} p^i q^{k-i} = 1 - \sum_{i=0}^{r-1} \binom{k}{i} p^i q^{k-i}. \tag{8}$$

and

$$\begin{aligned}
\text{RHS} &= \binom{k-1}{k-r} p^r q^{k-r} + (p+q)^{k-1} - \sum_{i=0}^{r-1} \binom{k-1}{i} p^i q^{k-1-i} \\
&= \binom{k-1}{r-1} + 1 - \binom{k-1}{r-1} p^{r-1} q^{k-r} - \sum_{i=0}^{r-2} \binom{k-1}{i} p^i q^{k-1-i} \\
&= 1 + \binom{k-1}{r-1} p^{r-1} (p-1) q^{k-r} - \sum_{i=0}^{r-2} \binom{k-1}{i} p^i q^{k-1-i} \\
&= 1 - \binom{k-1}{r-1} p^{r-1} q^{k+1-r} - \sum_{i=0}^{r-2} \binom{k-1}{i} p^i q^{k-1-i}.
\end{aligned} \tag{9}$$

Thus we need to examine

$$\binom{k-1}{r-1} p^{r-1} q^{k+1-r} \stackrel{?}{=} \sum_{i=0}^{r-1} \binom{k}{i} p^i q^{k-i} - \sum_{i=0}^{r-2} \binom{k-1}{i} p^i q^{k-1-i}. \tag{10}$$

Using the well-known formula

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

we rearrange the RHS of (10) 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 (10).

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.

4.4* 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

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

(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)

$$\mu_X = \int_0^{\infty} xdF_X(x) + \int_{-\infty}^{0} xdF_X(x)$$
$$= -\int_0^{\infty} xd(1 - F_X(x)) + \int_{-\infty}^{0} xdF_X(x)$$
$$= -[x(1 - F_X(x))]_0^{\infty} + \int_0^{\infty} (1 - F_X(x))\, dx + [xF_X(x)]_{-\infty}^{0} - \int_{-\infty}^{0} F_X(x)\, dx$$
$$= \int_0^{\infty} (1 - F_X(x))\, dx - \int_{-\infty}^{0} F_X(x)\, dx,$$

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

4.2 Important Continuous Random Variables and Their Distributions

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

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

which is the midpoint of the interval $[a, b]$.

The 2nd moment can be found in a similar fashion:

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

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)

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

4.13* 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

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

4.15* Mean and variance of the normal distribution.

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

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

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

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

We have

$$1 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \overset{(i)}{=} 2 \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$
$$\overset{(ii)}{=} \frac{1}{\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} e^{-x} dx \overset{(iii)}{=} \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right),$$

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,

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

4.3 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

$$u_1^2 - 2\rho u_1 u_2^2 = K_1, \tag{1}$$

where

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

Re-arranging the left-hand side of (1), 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 (1), 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

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

where

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}. \tag{3}$$

Since $\boldsymbol{x}_a$ is fixed, we need to analyse only the exponent of the numerator in the RHS of equation (2):

It is easy to verify that

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{I} & -\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_{aa}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1} & \boldsymbol{I} \end{bmatrix} \tag{4}$$

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

$$\begin{aligned} &(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}_a-\boldsymbol{\mu}_a)^\top \boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{x}_a-\boldsymbol{\mu}_a) \\ &+[-(\boldsymbol{x}_a-\boldsymbol{\mu}_a)^\top \boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab} + (\boldsymbol{x}_b-\boldsymbol{\mu}_b)^\top]\boldsymbol{S}^{-1}[-\boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{x}_a-\boldsymbol{\mu}_a) + (\boldsymbol{x}_b-\boldsymbol{\mu}_b)]. \end{aligned} \tag{5}$$

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

$$(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}_b-\boldsymbol{\mu}_b)^\top \boldsymbol{S}^{-1}(\boldsymbol{x}_b-\boldsymbol{\mu}_b) - 2(\boldsymbol{x}_b-\boldsymbol{\mu}_b)^\top \boldsymbol{b} + const. \tag{6}$$

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

$$(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = (\boldsymbol{x}_b-\boldsymbol{\mu}_b-\boldsymbol{S}\boldsymbol{b})^\top \boldsymbol{S}^{-1}(\boldsymbol{x}_b-\boldsymbol{\mu}_b-\boldsymbol{S}\boldsymbol{b}) + const \tag{7}$$

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,

$$f_{\boldsymbol{X}_b|\boldsymbol{X}_a}(\boldsymbol{x}_b|\boldsymbol{x}_a) \sim \exp\left\{-\frac{1}{2}(\boldsymbol{x}_b-\boldsymbol{\mu}_{b|a})^\top \boldsymbol{\Sigma}_{b|a}^{-1}(\boldsymbol{x}_b-\boldsymbol{\mu}_{b|a})\right\} \tag{8}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{b|a} &= \boldsymbol{\mu}_b + \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}(\boldsymbol{x}_a-\boldsymbol{\mu}_a) \\ \boldsymbol{\Sigma}_{b|a} &= \boldsymbol{S} = \boldsymbol{\Sigma}_{bb} - \boldsymbol{\Sigma}_{ba}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab} \end{aligned}. \tag{9}$$

4.4 Exponential Family of Distributions

4.26* Exponential families of distributions

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

4.5 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)

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

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 A Function of Two Random Variables

5.6* Leibniz's rule.

(a) The LHS of (5.95) equals

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

(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).

5.16* 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 $v \ge u$:

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

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 $a \le b$. (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 $u \le v$;

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

For $v \le u$

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

Thus, by combining the above results, we have

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

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.

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

$$2u_2 \le 2x, \text{ i.e., } u_2 \le x, \tag{1}$$

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.

5.20* Erlang variates. From (5.75) we see

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

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.

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

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

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

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

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 Θ 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 Θ 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

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

Then we can obtain (6.11)

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

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

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

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

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

6.11* 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

$$R_X(t) = E[R|X > t] = \int_0^\infty P[R > r|X > t]\,dr = \int_0^\infty S_X(r|t)\,dr$$
$$= \int_0^\infty \frac{S_X(t+r)}{S_X(t)}\,dr = \frac{\int_t^\infty S_X(u)\,du}{S_X(t)}, \tag{4}$$

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).

6.15* Covariance between two random variables. Since X is uniformly distributed between $-\pi$ and π, $E[X] = 0$. 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.

6.18* 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

$$E[s_{xy}] = \sigma_{XY}^2.$$

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 χ^2:

$$\begin{aligned} u_n &= -(u_1 + u_2 + \cdots + u_{n-1}), \\ u_n^2 &= u_1^2 + 2(u_1 u_2 + u_1 u_3 + \cdots + u_1 u_{n-1}) \\ &\quad + u_2^2 + 2(u_2 u_3 + \cdots + u_2 u_{n-1}) \\ &\qquad\qquad\qquad \vdots \\ &\qquad\qquad\qquad\quad + u_{n-1}^2. \end{aligned}$$

Then we can write χ^2 as

$$\begin{aligned}\frac{\chi^2}{2} &= u_1^2 + u_1u_2 + u_1u_3 + \cdots + u_1u_{n-1} \\ &\quad + u_2^2 + u_2u_3 + \cdots + u_2u_{n-1} \\ &\qquad \vdots \\ &\qquad + u_{n-1}^2 \\ &= \left[u_1 + \frac{1}{2}(u_2 + u_3 + \cdots + u_{n-1})\right]^2 \\ &\quad \frac{3}{4}u_2^2 + \frac{1}{2}(u_2u_3 + \cdots + u_2u_{n-1}) \\ &\qquad \frac{3}{4}u_3^2 + \frac{1}{2}(u_3u_4 + \cdots + u_3u_{n-1}) \\ &\qquad \vdots \\ &\qquad + \frac{3}{4}u_{n-1}^2.\end{aligned}$$

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 $E[U_r U_s] = 0$ 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.

7.3* Moments of gamma and χ^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)

$$f_{\chi_n^2}(\nu) d\nu = \frac{\nu^{\frac{n}{2}-1} e^{-\frac{\nu}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}, \;\; 0 \le \nu < \infty.$$

$$
\begin{aligned}
E[(\chi_n^2)^m] &= \int_0^\infty \nu^m f_{\chi_n^2}(\nu) d\nu \\
&= \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty \nu^{m+\frac{n}{2}-1} e^{-\nu/2}\, d\nu \\
&= \frac{2^{m+\frac{n}{2}}}{2^{n/2}\Gamma(n/2)} \int_0^\infty t^{m+\frac{n}{2}-1} e^{-t}\, dt \\
&= \frac{2^m \Gamma\left(\frac{n}{2}+m\right)}{\Gamma(n/2)}.
\end{aligned}
$$

7.2 Student's t-Distribution

7.7* Moments of the F-distribution.

From (7.39), the rth 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 $-n_1 < 2r < n_2$.

7.3 Lognormal Distribution

7.9* Median and mode of the lognormal distribution.

(a) The median of the distribution (7.46) is $y_{\text{med}} = \mu_Y$. The corresponding x_{med} is

$$x_{\text{med}} = e^{y_{\text{med}}} = e^{\mu_Y}.$$

Using (7.52), we find

$$x_{\text{med}} = e^{\mu_Y} = e^{\ln \mu_X - \frac{1}{2}\ln\left(1+\frac{\sigma_Y^2}{\mu_X^2}\right)}$$
$$= \mu_X \left(1 + \frac{\sigma_Y^2}{\mu_X^2}\right)^{-\frac{1}{2}} = \frac{\mu_X}{\sqrt{\left(1 + \frac{\sigma_Y^2}{\mu_X^2}\right)}}.$$

(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_{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

$$x_{\text{mode}} = \frac{\mu_X}{\left(1 + \frac{\sigma_Y^2}{\mu_X^2}\right)^{\frac{3}{2}}}.$$

7.4 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

$$M_{X^2}(t) = E[e^{t\sigma^2 U^2}] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{t\sigma^2 u^2} e^{-\frac{u^2}{2}}\, du$$
$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(u\sqrt{1-2\sigma^2 t})^2}{2}}\, du.$$

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

$$u = r\cos\theta, \;\; v = r\sin\theta, \;\; \text{hence } du\, dv = r dr\, d\theta,$$

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.

7.13* Nakagami distribution.

(a)

$$E[Z] = 2m\sigma^2 \triangleq \Omega, \tag{1}$$

Then by writing $X_i = \sigma U_i$, we readily find

$$Z = \sigma^2\chi^2_{2m}, \tag{2}$$

where χ^2_{2m} is the chi-squared variable with $2m$ degrees of freedom[1]. Thus, we find the PDF of Z as

$$f_Z(z) = \frac{1}{\sigma^2}f_{\chi^2_{2m}}\left(\frac{z}{\sigma^2}\right) = \frac{1}{\sigma^2}\frac{\left(\frac{z}{\sigma^2}\right)^{m-1}e^{-\frac{z}{2\sigma^2}}}{2^m\Gamma(m)}, \quad z \geq 0, \tag{3}$$

where $\Gamma(m)$ is the gamma function defined in (4.31), and $\Gamma(m) = (m-1)!$ when m is an integer. By substituting (1), we have

$$f_Z(z) = \frac{m^m}{\Omega^m\Gamma(m)}z^{m-1}e^{-\frac{mz}{\Omega}}, \quad z \geq 0. \tag{4}$$

(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

$$f_R(r) = \frac{2m^m}{\Omega^m\Gamma(m)}r^{2m-1}e^{-\frac{mr^2}{\Omega}}, \quad r \geq 0. \tag{5}$$

An alternative expression is given in terms of σ^2 as

$$f_R(r) = \frac{2\left(\frac{m}{2\sigma^2}\right)^m}{\Gamma(m)}r^{2m-1}e^{-\frac{mr^2}{2\sigma^2}}, \quad r \geq 0, \tag{6}$$

[1] Recall that $Y_{2m} \triangleq \chi^2_{2m}/2$ is an 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)).

(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 $E[R^2] - E^2[R]$.

7.5 Complex-valued normal variables

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:

$$\det \begin{bmatrix} \boldsymbol{I}_k & \boldsymbol{I}_k \\ \boldsymbol{I}_k & -\boldsymbol{I}_k \end{bmatrix} = (-2)^k. \tag{7}$$

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 (7) 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

$$m(0) = 0, \;\; m'(0) = M'(0) = E[X], \;\; m''(0) = M''(0) - M'(0)^2 = E[X^2] - E[X]^2 = \sigma^2.$$

8.3* Exponential distribution.

$$M(t) = \int_0^\infty e^{tx} \mu e^{-\mu x}\, dx = \mu \left[\frac{e^{(t-\mu)x}}{t-\mu}\right]_0^\infty = \begin{cases} \frac{\mu}{\mu - t} & t < \mu \\ \infty & t \geq \mu \end{cases}$$

8.7* Multivariate normal distribution. The derivation is essentially the same as that for the bivariate normal distribution. Let

$$Y \triangleq t_1 X_1 + \ldots + t_m X_m = \langle \boldsymbol{t}, \boldsymbol{X} \rangle.$$

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).

8.9* **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.

8.2 Characteristic Function (CF)

8.11* **CF of the binomial distribution.** By definition

$$\begin{aligned} \phi(u) &= \sum_{k=0}^n B(k; n, p) e^{iuk} = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{iuk} \\ &= \sum_{k=0}^n \binom{n}{k} (pe^{iu})^k (1-p)^{n-k} = (pe^{iu} + 1 - p)^n. \end{aligned}$$

8.15* **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:

$$\left| \int_{-auR}^{0} e^{-\frac{R+iy}{a}}\, i\, dy \right| \le e^{-\frac{R}{a}} \int_{-auR}^{0} dy = auRe^{-\frac{R}{a}} \overset{R\to\infty}{\longrightarrow} 0. \tag{4}$$

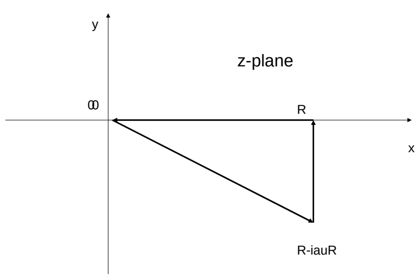
Therefore, we have

$$\lim_{R\to\infty} \int_0^{R-iauR} e^{-\frac{z}{a}}\, dz = \lim_{R\to\infty} \int_0^{R} e^{-\frac{x}{a}}\, dx = a. \tag{5}$$

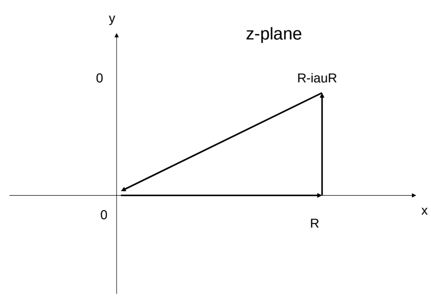
Thus, we find

$$\boxed{\phi_X(u) = \frac{1}{1-iau}, \quad -\infty < u < \infty} \tag{6}$$

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

$$\mu_X = (-i)\psi'_X(0) = a, \;\text{ and }\; \sigma_X^2 = (-i)^2\psi''_X(0) = a^2.$$

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 $|f_k| \leq M$ for all k and some constant M, then

$$|F(z)| \leq \sum_{k=0}^{\infty} |f_k z^k| \leq M \sum |z^k| = M \frac{1}{1-|z|}, \text{ for } |z| < 1.$$

(b) Similarly

$$|\tilde{F}(z)| \leq \sum_{k=0}^{\infty} |f_k z^{-k}| \leq M \sum |z^{-k}| = M \frac{1}{1-|z|^{-1}}, \text{ for } |z^{-1}| < 1, \text{ or } |z| > 1.$$

(c) Since $\sum_k p_k = 1$ by definition

$$|P(z)| \leq P(1) = \sum_k p_k = 1, \text{ for } |z| \leq 1.$$

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 $\left(\frac{pz}{1-qz}\right)^r$.

9.10* 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 kth 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

$$(1-qz)^{-r} = \sum_{j=0}^{\infty} \binom{-r}{j} (-qz)^j, \quad |z| < q^{-1}.$$

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):

$$\binom{-n}{i} = \frac{(-n)!}{i!(-n-i)!} = (-1)^i \frac{n(n+1)\cdots(n+i-1)}{i!} = (-1)^i \binom{n+i-1}{i}, \tag{2}$$

(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).

9.15* Derivation of the binomial distribution via a two-dimensional generating function $C(z, w)$.

(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

$$b(k;n,p) = \binom{n}{k} p^k q^{n-k}.$$

9.18* 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,

$$\Phi_g(s) = \int_0^{\infty} f_1(z)e^{-sz}\,dz \int_0^{\infty} f_2(y)e^{-sy}\,dy = \Phi_{f_1}(s)\Phi_{f_2}(s).$$

9.21* 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

$$\Phi_X(s) = \sum_k p_k e^{-sx_k} + \int g(x) e^{-sx} \, dx.$$

10 Solutions for Chapter 10: Inequalities, Bounds and Large Deviation Approximation

10.1 Inequalities frequently used in Probability Theory

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

$$P\left[\frac{S_n}{n} - p \geq \epsilon\right] \leq \exp\left(-\frac{n\epsilon^2}{3}\right), \quad \epsilon > 0.$$

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,

$$\begin{aligned} E[W_n|Y_{n-1}] &= e^{\lambda Y_{n-1}} E\left[e^{\lambda(Y_n - Y_{n-1})}|Y_{n-1}\right] \\ &\leq W_{n-1}\frac{b_n e^{-\lambda a_n} + a_n e^{\lambda b_n}}{a_n + b_n}, \end{aligned}$$

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

$$W_n = \max\{0, X_{n-1}, X_{n-1} + X_{n-2}, \dots, X_{n-1} + X_{n-2} + \dots + X_0\}.$$

For $\theta > 0$, $e^{\theta W_n}$ is a monotone increasing function of W_n. Hence,

$$e^{\theta W_n} = \max\left\{1, e^{\theta X_{n-1}}, e^{\theta(X_{n-1}+X_{n-2})}, \dots, e^{\theta(X_{n-1}+X_{n-2}+\dots+X_0)}\right\}$$
$$\max\{Y_0, Y_1, \dots, Y_n\}.$$

(b)

$$M_X(\theta) = E\left[e^{\theta X_n}\right].$$

The MGF is defined over an interval I_θ, in which the MGF is bounded. This domain I_θ 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_θ that satisfies

$$M_X(\theta) \geq 1. \tag{1}$$

Then

$$E[Y_n|Y_1, Y_2, \dots, Y_{n-1}] = E\left[e^{\theta(X_{n-1}+X_{n-2}+\cdots+X_0)} \middle| e^{\theta(X_{n-1}+X_{n-2})}, \dots, e^{\theta(X_{n-1}+X_{n-2}+\cdots+X_1)}\right]$$
$$= e\left[e^{\theta X_0}\right] e^{\theta(X_{n-1}+X_{n-2}+\cdots+X_1)} = M_X(\theta) Y_{n-1} \geq Y_{n-1},$$

which shows that Y_n is a submartingale.

(c) By applying the Doob-Kolmorogov' inequlaity, we obtain

$$F^c_{W_n}(t) = P[W_n > t] = P\left[e^{\theta W_n \geq e^{\theta t}}\right]$$
$$= P\left[\max\{Y_0, Y_1, \dots, Y_n\} \geq e^{\theta t}\right]$$
$$\leq \frac{P[Y_n]}{e^{\theta t}} = e^{-\theta t + n m_X(\theta)},$$

(c) The tightest bound is attained by finding

$$\min\left\{m_X(\theta) - \frac{t\theta}{n}\right\}$$

with the constraint (1), or equivalently

$$m_X(\theta) \geq 0. \tag{2}$$

10.2 Chernoff's bound

10.3 Large Deviation Theory

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

$$P[A^c] = P\left[\bigcup_{\epsilon>0} A^c(\epsilon)\right] \leq \sum_{\epsilon>0} P[A^c(\epsilon)]. \tag{9}$$

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 $m \geq N(\epsilon, \delta)$

$$P\left[\bigcap_{n=m}^{\infty} \{\omega : |X_n(\omega) - X(\omega)| < \epsilon\}\right] \geq 1 - \delta, \tag{15}$$

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

$$F_{Z_n}(z) = P[Z_n \leq z] = P[n(1 - Y_n) \leq z] = P\left[Y_n \geq 1 - \frac{z}{n}\right]. \tag{16}$$

Since $Y_n = \max\{X_1, X_2, \ldots, X_n\}$,

$$P\left[Y_n \leq 1 - \frac{z}{n}\right] = P\left[X_i \leq 1 - \frac{z}{n},\ 1 \leq i \leq n\right] = \left[F_X\left(1 - \frac{z}{n}\right)\right]^n. \tag{17}$$

Note that $0 \leq Y_n \leq 1$, almost surely, and hence $Z_n \geq 0$, a.s. When $n > z \geq 0$,

$$0 \leq 1 - \frac{z}{n} \leq 1,$$

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., $Z_n \stackrel{D}{\to} Z$.

11.3* 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,

$$\frac{1}{n}\sum_{k=1}^{n}N_k \stackrel{P}{\to} 0.$$

Therefore,

$$\lim_{n\to\infty}P[|\overline{X}_n-c|\geq\epsilon]=0,$$

for any $\epsilon>0$. Hence, $\overline{X}_n \stackrel{P}{\to} c$.

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 $\frac{1}{r}+\frac{1}{s}=1$,

$$\exp\left(\frac{u}{r}+\frac{v}{s}\right)\leq\frac{e^u}{r}+\frac{e^v}{s}. \tag{20}$$

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 (20) 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 (20) is

$$\text{RHS} = \frac{1}{r}\frac{|X|^r}{\|X\|_r^r} + \frac{1}{s}\frac{|Y|^s}{\|Y\|_s^s}.$$

Thus

$$\frac{|XY|}{\|X\|_r\|Y\|_s} \le \frac{1}{r}\frac{|X|^r}{\|X\|_r^r} + \frac{1}{s}\frac{|Y|^s}{\|Y\|_s^s}.$$

By taking the expectation, we find

$$\frac{\|XY\|_1}{\|X\|_r\|Y\|_s} \le \frac{1}{r} + \frac{1}{s} = 1.$$

(b) The proof given in part (a) can carry over to this case. From (20), we have

$$\sum_{i=1}^{n} \exp\left(\frac{u_i}{r} + \frac{v_i}{s}\right) \le \sum_{i=1}^{n}\left(\frac{e_i^u}{r} + \frac{e_i^v}{s}\right). \tag{21}$$

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 (21) 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,

$$\frac{x^r}{r} + \frac{x^{-s}}{s} \ge 1.$$

In order to derive

$$uv \le \frac{u^r}{r} + \frac{v^s}{s}$$

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:**

$$\begin{aligned}
E[|X+Y|^r] &= E[|X+Y||X+Y|^{r-1}] \\
&\leq E[(|X|+|Y|)|X+Y|^{r-1}] \\
&= E[|X||X+Y|^{r-1}] + E[|Y||X+Y|^{r-1}] && (22) \\
&\leq (E[|X|^r])^{1/r} \cdot \left(E[|X+Y|^{(r-1)s}]\right)^{1/s} \\
&\quad + (E[|Y|^r])^{1/r} \cdot \left(E[|X+Y|^{(r-1)s}]\right)^{1/s} && (23) \\
&= (||X||_r + ||Y||_r)\,(E[|X+Y|^r])^{\frac{r-1}{r}}\ , && (24)
\end{aligned}$$

where (23) is obtained by applying Hölder's inequality to each term in (22), with

$$\frac{1}{r} + \frac{1}{s} = 1 \Rightarrow s = \frac{r}{r-1}.$$

Multiplying both sides of the inequality (24) by the factor

$$\frac{||X+Y||_r}{E[|X+Y|^r]},$$

we obtained the desired result:

$$||X+Y||_r \leq ||X||_r + ||Y||_r.$$

11.3 Limit theorems

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)

$$R_X(\tau) = E[X(t+\tau)X(t)] = E\left[\left(\sum_{i=0}^{m}\{A_i\cos\omega_i(t+\tau) + B_i\sin\omega_i(t+\tau)\}\right) \cdot \left(\sum_{j=0}^{m}\{A_j\cos\omega_j t + B_j\sin\omega_j t\}\right)\right].$$

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].

12.4 Complex-Valued Gaussian Process

12.3* Condition for integration in mean-square.

We can expand the LHS of (12.39) as follows:

$$\begin{aligned} E[|Y-S_n|^2] &= E[YY^*] - E[YS_n^*] - E[Y^*S_n] + E[S_nS_n^*] \\ &= \int_a^b \int_a^b h(t)E[Z(t)Z^*(s)]h^*(s)\,dt\,ds \\ &- \int_a^b \sum_{i=1}^n h(t)E[Z(t)Z^*(t_i)]h^*(t_i)(t_{i+1}-t_i) \\ &- \int_a^b \sum_{i=1}^n h^*(t)E[Z^*(t)Z(t_i)]h(t_i)(t_{i+1}-t_i) \\ &+ \sum_{i=1}^n \sum_{j=1}^n h(t_i)E[Z(t_i)Z^*(t_j)]h^*(t_j)(t_{i+1}-t_i)(t_{j+1}-t_j). \end{aligned} \tag{1}$$

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

$$\begin{aligned} \lim_{n\to\infty} \int_a^b \sum_{i=1}^n h(t)E[Z(t)Z^*(t_i)]h^*(t_i)(t_{i+1}-t_i) &= \lim_{n\to\infty} \int_a^b \sum_{i=1}^n h(t)R(t,t_i)h^*(t_i)(t_{i+1}-t_i) \\ &= \int_a^b \int_a^b h(t)R(t,s)h^*(s)\,ds = Q \end{aligned}$$

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

$$\begin{aligned} Z(t)e^{i\theta} &= X(t)\cos\theta - Y(t)\sin\theta + i(X(t)\sin\theta + Y(t)\cos\theta) \\ &\triangleq U(t) + iV(t). \end{aligned}$$

If $Z(t) = X(t) + iY(t)$ and $Z(t)e^{i\theta} = U(t) + iV(t)$ have the same distribution, their 2×2-covariance function matrices must be the same. Let the four elements of the matrix be

$$\begin{aligned} E[X(s)X(t)] &\triangleq A(s,t), \quad E[X(s)Y(t)] \triangleq B(s,t) \\ E[Y(s)X(t)] &= B(s,t), \quad E[Y(s)Y(t)] \triangleq D(s,t). \end{aligned}$$

Then, the following relation must hold for any θ.

$$E[U(s)U(t)] = E\left[(X(s)\cos\theta - Y(s)\sin\theta)(X(t)\cos\theta - Y(t)\sin\theta)\top\right]$$

$$= \cos^2\theta A(s,t) - \sin\theta\cos\theta(B(s,t) + B(s,t)) + \sin^2\theta D(s,t) = A(s,t), \tag{2}$$

$$E[U(s)V(t)] = \sin\theta\cos\theta(A(s,t) - D(s,t)) - \sin^2\theta B(s,t) + \cos^2\theta B(s,t) = B(s,t). \tag{3}$$

$$E[V(s)U(t)] = \sin\theta\cos\theta(A(s,t) - D(s,t)) - \sin^2\theta B(s,t) + \cos^2\theta B(s,t) = B(s,t), \tag{4}$$

$$E[V(s)V(t)] = \sin^2\theta A(s,t) + \sin\theta\cos\theta(B(s,t) + B(s,t)) + \cos^2\theta D(s,t) = D(s,t). \tag{5}$$

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 θ, which implies that the distribution of $Z(t)e^{i\theta}$ is invariant under θ.

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

$$G^*(f) = \int_{-\infty}^{\infty} g^*(s) e^{2\pi f t}\, ds$$

we have

$$\begin{aligned}
\int_{-\infty}^{\infty} |G(f)|^2\, df = \int_{-\infty}^{\infty} G(f)G^*(f)\, df &= \int_{f=-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(t) e^{-i2\pi f t}\, dt \int_{-\infty}^{\infty} g^*(s) e^{2\pi f s}\, ds \right] df \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\int_{f=-\infty}^{\infty} e^{i2\pi(s-t)}\, df \right] g(t) g^*(s)\, dt\, ds \\
&= \int_{t=-\infty}^{\infty} \left[\int_{s=-\infty}^{\infty} \delta(s-t) g^*(s)\, ds \right] g(t)\, dt \\
&= \int_{t=-\infty}^{\infty} g^*(t) g(t)\, dt = \int_{-\infty}^{\infty} |g(t)|^2\, dt.
\end{aligned}$$

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.

13.9* 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 $\boldsymbol{u}^H \boldsymbol{u}_j = 0$.

13.12* 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 jth 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 α^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 mth eigenvalue is

$$\lambda_m = \sum_{k=0}^{n-1} c_k W^{km}, \tag{2}$$

and the mth 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 λ_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 α 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 ith element of the vector $\boldsymbol{u}$, i.e.,

$$\boldsymbol{u} = (u_0, u_1, \dots, u_{n-1})^\top. \tag{6}$$

Then from (4), we find

$$u_1 = \alpha u_0, u_2 = \alpha u_1 = \alpha^2 u_0, \dots, u_i = \alpha u_{i-1} = \alpha^i u_0, \dots, u_{n-1} = \alpha^{n-1} u_0. \tag{7}$$

From (4), we also find

$$\boldsymbol{V}^i \boldsymbol{u} = \alpha^i \boldsymbol{u}, \;\; i = 0, 1, 2, \dots. \tag{8}$$

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 α, which are given by

$$\alpha_m = \exp\left(\frac{i2\pi m}{n}\right) = W^m, \;\; m = 0, 1, \dots, n-1,$$

where $W = \exp\left(\frac{i2\pi}{n}\right)$, as defined in (1).
The mth eigenvector is found from equation (6) as

$$\boldsymbol{u}_m = (1, W^m, W^{2m}, \dots, W^{m(n-1)})^\top, \;\; m = 0, 1, \dots, n-1,$$

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

$$\lambda_m = \sum_{k=0}^{n-1} c_k W^{mk}, \;\; m = 0, 1, \dots, n-1,$$

which shows that the eigenvalues are the DFT of $(c_0, c_1, \dots, c_{n-1})$.

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

$$\begin{aligned}\mathrm{SNR} &= \frac{|\boldsymbol{h}^\top \boldsymbol{S}|^2}{\boldsymbol{h}^\top \boldsymbol{R}_N \boldsymbol{h}^*} = \frac{|(\boldsymbol{R}_N^{1/2}\boldsymbol{h})^\top (\boldsymbol{R}_N^{-1/2}\boldsymbol{S})|^2}{(\boldsymbol{R}^{1/2}\boldsymbol{h})^\top (\boldsymbol{R}_N^{1/2}\boldsymbol{h}^*)}\\ &\le \frac{\|\boldsymbol{R}_N^{1/2}\boldsymbol{h}\|^2 \|\boldsymbol{R}_N^{-1/2}\boldsymbol{S}\|^2}{\|\boldsymbol{R}^{1/2}\boldsymbol{h}\|^2} = \|\boldsymbol{R}_N^{-1/2}\boldsymbol{S}\|^2\\ &= \boldsymbol{S}^\top \boldsymbol{R}_N^{-1}\boldsymbol{S}^*,\end{aligned}$$

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

$$\begin{aligned}
\text{LHS} &= a^2 \left(\int_0^t s \left(\frac{e^{as}}{a} \right)' ds + t \left[\frac{e^{as}}{a} \right]_t^T \right) \\
&= a^2 \left(\left[\frac{se^{as}}{a} \right]_0^t - \int_0^t \frac{e^{as}}{a}\, ds + \frac{t(e^{aT} - e^{at})}{a} \right) \\
&= a^2 \left(\frac{te^{at}}{a} - \frac{e^{at} - 1}{a^2} + \frac{te^{aT} - te^{at}}{a} \right) \\
&= ate^{at} - e^{at} + 1 + ate^{aT} - ate^{at} = -e^{at} + ate^{aT} - 1.
\end{aligned}$$

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 λ and performing the integration, we have

$$\begin{aligned}\text{LHS} &= \omega^2 C_1\left(\frac{te^{i\omega t}}{i\omega}-\frac{e^{i\omega t}-1}{(i\omega)^2}+\frac{te^{i\omega T}+i\omega te^{i\omega t}}{i\omega}\right)\\ &+\omega^2 C_2\left(\frac{te^{-i\omega t}}{-i\omega}-\frac{e^{-i\omega t}-1}{(i\omega)^2}+\frac{te^{-i\omega T}-i\omega te^{-i\omega t}}{i\omega}\right)\\ &= C_1(-1-i\omega te^{i\omega T})+C_2(-1+i\omega te^{-\omega T})\\ &= -(C_1+C_2)-i\omega t(C_1e^{i\omega T}-C_2e^{-\omega T}).\end{aligned}$$

This equals the RHS ($C_1e^{i\omega t}+C_2e^{-i\omega t}$), 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,

$$\omega_n=\frac{(2n+2)\pi}{2T},\ \ n=0,\pm1,\pm2,\ldots.$$

Thus, the eigenvalues are

$$\lambda_n=\frac{\sigma^2}{\omega_n^2},\ \ n=0,\pm1,\pm2,\ldots.$$

The corresponding eigenfunctions are

$$v_n(t)=C_1e^{i\omega_n t}-C_1e^{-i\omega_n t}=i2C_1\sin\omega_n t.$$

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

$$W_n=\int_0^T \psi_n(t)W(t)\,dt=\sqrt{\frac{2}{T}}\int_0^T \sin\omega_n tW(t)\,dt.$$

From the theory of K-L expansion, we know the set of $\psi_n(t),\quad n=0,\pm1,\pm2,\ldots$ are orthogonal.

$$\begin{aligned}E[W_n^2] &= \frac{2}{T}\int_0^T\int_0^T \sin\omega_n t\omega_n sE[W(t)W(s)]\,dt\,ds\\ &= \frac{2\sigma^2}{T}\int_0^T\int_0^T \min(t,s)\sin\omega_n t\sin\omega_n s\,dt\,ds\end{aligned}$$

Now, we evaluate

$$\int_0^T \min(t,s)\sin\omega_n s\,ds = \int_0^t s\sin\omega_n s\,ds + t\int_t^T \sin\omega_n s\,ds$$
$$= \int_0^t s\left(-\frac{\cos\omega_n s}{\omega_n}\right)' ds + t\int_t^T \sin\omega_n s\,ds$$
$$= -\left[\frac{s\cos\omega_n t}{\omega_n}\right]_0^t + \int_0^t \frac{\cos\omega_n s}{\omega_n}\,ds - t\left[\frac{\cos\omega_n s}{\omega_n}\right]_t^T$$
$$= -\frac{t\cos\omega_n t}{\omega_n} + \frac{1}{\omega_2}\left[\sin\omega_n s\right]_0^t + t\frac{\cos\omega_n t}{\omega_n}$$
$$= \frac{\sin\omega_n t}{\omega^2}. \tag{10}$$

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:

$$E[W_n W_m] = \frac{2}{T}\int_0^T\int_0^T \sin\omega_n t\omega_m s E[W(t)W(s)]\,dt\,ds$$
$$= \frac{2\sigma^2}{T}\int_0^T \sin\omega_n t\left(\int_0^T \min(t,s)\sin\omega_m s\,ds\right)\,dt$$
$$= \frac{2\sigma^2}{T\omega^2}\int_0^T \sin\omega_n t\sin\omega_m t\,dt.$$

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

$$\sum_{n=-\infty}^{-1} W_n\sin\omega_n t = \sum_{m=1}^{\infty} W_{-m}\sin\omega_{-m} t = -\sum_{m=1}^{\infty} W_{-m}\sin\omega_{m-1} t$$
$$= -\sum_{n=0}^{\infty} W_{-n-1}\sin\omega_n t.$$

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

$$\begin{aligned} E[U_n^2] &= E[W_n^2] + E[W_{-n-1}^2] \\ &= 4\left(\frac{\sigma T}{(2n+1)\pi}\right)^2 + 4\left(\frac{\sigma T}{(-2n-1)\pi}\right)^2 \\ &= 8\left[\frac{\sigma T}{(2n+1)\pi}\right]^2. \end{aligned}$$

13.2 PCA and SVD

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 $\|\boldsymbol{\chi}_i\|^2 = \mu_i$.

13.26* 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).

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

$$F_{t_1}(t) = F_X(t) = 1 - e^{-\lambda t}, \quad t \geq 0. \tag{1}$$

(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

$$f_{t_2}(t) = \int_0^t \lambda e^{-\lambda(t-u)} \lambda e^{-\lambda u}\, du = \lambda^2 t e^{-\lambda t}, \quad t \geq 0. \tag{2}$$

By repeating the above step, we find

$$f_{t_n}(t) = \frac{(\lambda)^n t^{n-1}}{(n-1)!} e^{-\lambda t}, \quad t \geq 0, \; n = 1, 2, \ldots. \tag{3}$$

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 λt, if the lifetime distribution is the exponential distribution (14.99).

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:

$$(0, T] = (0, u_1] \cup I_1 \cup (u_1 + h_1, u_2] \cup \cdots \cup I_n \cup (u_n + h_n, T]. \tag{5}$$

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 (5). Therefore,

$$\begin{aligned} P[\mathcal{E}_n] &= (e^{-\lambda u_1})(\lambda h_1 e^{-\lambda h_1})(e^{-\lambda(u_2 - u_1 - h_1)}) \cdots (\lambda h_n e^{-\lambda h_n})(e^{-\lambda(T - u_n - h_n)}) \\ &= \lambda^n e^{-\lambda T} \prod_{j=1}^{n} h_j. \end{aligned} \tag{6}$$

We can also write

$$\begin{aligned} P[\mathcal{E}_n] &\overset{(a)}{=} \sum_{\sigma} P[U_1 \in I_{\sigma(1)}, U_2 \in I_{\sigma(2)}, \ldots, U_n \in I_{\sigma(n)}, N(T) = n] \\ &\overset{(b)}{=} n!\, P[U_1 \in I_1, U_2 \in I_2, \ldots, U_n \in I_n], \end{aligned} \tag{7}$$

where the summation on the right-hand side of (a) is over all permutations σ on the set $\{1, \ldots, n\}$. Step (b) follows because the n arrivals are unordered. From (6) and (7), we obtain

$$P[U_1 \in I_1, \ldots, U_n \in I_n] = \frac{\lambda^n}{n!} e^{-\lambda T} \prod_{j=1}^{n} h_j. \tag{8}$$

Hence,

$$\begin{aligned} &P[U_1 \in I_1, \ldots, U_n \in I_n | N(T) = n] \\ &\quad = \frac{P[U_1 \in I_1, \ldots, U_n \in I_n]}{P[N(T) = n]} = \frac{\frac{\lambda^n}{n!} e^{-\lambda T} \prod_{j=1}^{n} h_j}{\frac{(\lambda T)^n}{n!} e^{-\lambda T}} = \frac{\prod_{j=1}^{n} h_j}{T^n}. \end{aligned} \tag{9}$$

We remark that (9) holds for any ordering of the u_i's, although we assumed $u_1 < \cdots < u_n$ for convenience in obtaining (6). 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))

$$f_{U_1 \cdots U_n}(u_1, \ldots, u_n | N(t) = n) = \lim_{\substack{h_j \to 0 \\ j = 1, \ldots, n}} \frac{P[U_1 \in I_1, \ldots, U_n \in I_n | N(T) = n]}{\prod_{j=1}^{n} h_j}, \tag{10}$$

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 λT.

b. For $i = 1, \dots, n$, 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

14.13* 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:

$$P_n'(t) = -(\lambda + n\mu)P_n(t) + \lambda P_{n-1}(t) + (n+1)\mu P_{n+1}(t), \quad n = 1, 2, \cdots, \tag{11}$$

$$P_0'(t) = \lambda P_0(t) + \mu P_1(t). \tag{12}$$

Multiply both sides of (11) and (12) by z^n and sum from $n = 0$ to ∞ to obtain:

$$\begin{aligned} \sum_{n=0}^{\infty} P_n'(t) z^n &= -\lambda \sum_{n=0}^{\infty} P_n(t) z^n - \mu \sum_{n=1}^{\infty} n P_n(t) z^n \\ &\quad + \lambda \sum_{n=1}^{\infty} P_{n-1}(t) z^n + \mu \sum_{n=0}^{\infty} (n+1) P_{n+1}(t) z^n, \end{aligned} \tag{13}$$

which can be written as:

$$\frac{\partial}{\partial t} G(z,t) = -\lambda G(z,t) - \mu z \frac{\partial}{\partial z} G(z,t) + \lambda z G(z,t) + \mu \frac{\partial}{\partial z} G(z,t). \tag{14}$$

Re-arranging terms we have the following partial differential equation in $G(z,t)$:

$$\left[\frac{\partial}{\partial t} + \mu(z-1)\frac{\partial}{\partial z}\right] G(z,t) = \lambda(z-1)G(z,t). \tag{15}$$

It remains to verify that the solution

$$G(z,t) = \exp\left\{\frac{\lambda}{\mu}(1 - e^{-\mu t})(z-1)\right\} \tag{16}$$

satisfies (15). Alternatively, we may obtain the form of the solution (16) as follows. Based on (15), we suppose that $G(z,t)$ has the form: $G(z,t) = \exp(f(z,t))$. In this case, (16) reduces to the following partial differential equation:

$$\left[\frac{\partial}{\partial t} + \mu(z-1)\frac{\partial}{\partial z}\right] f(z,t) = \lambda(z-1). \tag{17}$$

Based on (17), we suppose that $f(z,t)$ is separable as follows: $f(z,t) = \lambda(z-1)f(t)$. Then (17) simplifies to:

$$f'(t) + \mu f(t) = 1. \tag{18}$$

This is a first-order differential equation that can be solved by multiplying both sides by the integrating factor $e^{\mu t}$, resulting in:

$$\frac{d}{dt}[e^{\mu t}f(t)] = e^{\mu t}. \tag{19}$$

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).

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} m_X, \tag{20}$$

as $n \to \infty$. Noting that

$$\frac{1}{N(t)}\sum_{i=1}^{N(t)} X_i = \frac{t_{N(t)}}{N(t)}, \tag{21}$$

we deduce from (20) that

$$\frac{N(t)}{t_{N(t)}} \xrightarrow{\text{a.s.}} \frac{1}{m_X}, \tag{22}$$

as $t \to \infty$. The left-hand side of (22) can be written as

$$\frac{N(t)}{t}\frac{t}{t_{N(t)}}. \tag{23}$$

Since $\frac{t}{t_{N(t)}} \xrightarrow{\text{a.s.}} 1$ as $t \to \infty$, we can establish from (22) and (23) that (14.72) holds.

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 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,

$$\boldsymbol{\pi}^{\top} = \left(\frac{2}{15}, \frac{1}{5}, \frac{2}{3} \right).$$

16 Solutions for Chapter 16: Semi-Markov Processes and Continuous-Time Markov Chains

16.1 Semi-Markov Process

16.2* Conditional independence of sojourn times.

Using a basic property of conditional probability (see (2.59) in Section 2.4.1), we have

$$\begin{aligned}
&P[\tau_1 \le u_1, \tau_2 \le u_2, \ldots, \tau_n \le u_n | X_0, X_1, \ldots] \\
&\quad = P[\tau_1 \le u_1 | \tau_2 \le u_2, \ldots, \tau_n \le u_n, X_0, X_1, \ldots] \\
&\quad\quad \cdot P[\tau_2 \le u_2 | \tau_3 \le u_2, \ldots, \tau_n \le u_n, X_0, X_1, \ldots] \\
&\quad\quad \cdot P[\tau_3 \le u_3 | \tau_4 \le u_4, \ldots, \tau_n \le u_n, X_0, X_1, \ldots] \\
&\quad\quad \cdots \\
&\quad\quad \cdot P[\tau_n \le u_n | X_0, X_1, \ldots]. \qquad (1)
\end{aligned}$$

Since τ_j depends only on X_{j-1} and X_j, we have for $1 \le j \le n-1$:

$$\begin{aligned}
&P[\tau_j \le u_j | \tau_{j+1} \le u_{j+1}, \ldots, \tau_n \le u_n, X_0, X_1, \ldots] \\
&\quad = P[\tau_j \le u_j | X_{j-1}, X_j] = F_{X_{j-1}X_j}(u_j). \qquad (2)
\end{aligned}$$

Applying (2) in (1), we obtain the desired result

$$\begin{aligned}
&P[\tau_1 \le u_1, \tau_2 \le u_2, \ldots, \tau_n \le u_n | X_0, X_1, \ldots] \\
&\quad\quad = F_{X_0X_1}(u_1) F_{X_1X_2}(u_2) \cdots F_{X_{n-1}X_n}(u_n).
\end{aligned}$$

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:

$$\begin{aligned} F_{ij}(t) &= P[t_{n+1} - t \le t | X_n = i, X_{n+1} = j] \\ &= \frac{P[X_{n+1}, t_{n+1} - t_n \le t | X_n = i]}{P[X_{n+1} = j | X_n = i]} \\ &= \frac{Q_{ij}(t)}{P_{ij}} \\ &= \frac{Q_{ij}(t)}{\lim_{t\to\infty} Q_{ij}(t)}. \end{aligned} \tag{3}$$

Conversely, given $\boldsymbol{P}$ and $\boldsymbol{F}(t)$, the semi-Markovian kernel can be obtained from (3) as follows:

$$Q_{ij}(t) = F_{ij}(t) P_{ij}. \tag{4}$$

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

$$F_{ij}(t) = 1 - e^{-\nu_i t}, \quad t \ge 0, \quad i, j \in \mathcal{S}. \tag{16.17}$$

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 $j \ne i$. For $s, t \ge 0$, we have

$$\begin{aligned} P[S_i > s + t \mid S_i > s] &= P[X(\tau) = i;\ 0 \le \tau \le s + t \mid X(\tau) = i;\ 0 \le \tau \le s] \\ &= P[X(\tau) = i;\ s \le \tau \le s + t \mid X(\tau) = i;\ 0 \le \tau \le s] \\ &= P[X(\tau) = i;\ s \le \tau \le s + t \mid X(s) = i] \qquad (5) \\ &= P[X(\tau) = i,\ 0 \le \tau \le t \mid X(0) = i] \qquad (6) \\ &= P[S_i > t], \qquad (7) \end{aligned}$$

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 ν_i. In other words, the sojourn time distributions of $X(t)$ are given by (16.17).

[1] Let $g(t) = P[S_i > t]$. Then (7) implies that $g(t + s) = g(t)g(s)$ for all $s, t \ge 0$. It is well-known that the unique solution to this functional equation has the form $g(t) = e^{\alpha t}$ for some constant α.

Let $t_0 = 0$ and let $t_1, t_2, \dots$ denote the jump times of $X(t)$. It remains to show that the process $\{X_n\}$ defined by $X(t_n)$, $n = 0, 1, 2, \dots$ is a DTMC. For $n \geq 1$, we have

$$\begin{aligned} P[X_n = j \mid X_0, X_1, \dots, X_{n-1}] & \\ = P[X(t_n) = j \mid X(t_0), X(t_n), \dots, X(t_{n-1})] & \quad (8) \\ = P[X(t_n) = j \mid X(t_{n-1})] & \\ = P[X_n = j \mid X_{n-1}] = P_{X_{n-1}, i}. & \quad (9) \end{aligned}$$

In the above derivation, (8) formally resembles the Markov property given in Definition 15.1. A key difference, however, is that the times $t_1, t_2, \dots$ 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 $t_1, t_2, \dots$ 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,

$$\lim_{h \to 0} \frac{P[S_i < h]}{h} = \lim_{h \to} \frac{1 - P_{ii}(h)}{h} = -Q_{ii}, \quad (10)$$

Since the left-hand side of (10) is given by ν_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

$$\begin{aligned} P_{ij} &= \lim_{h \to 0} P[X(S_i + h) = j \mid X(t) = i,\ 0 \leq t < S_i; X(S_i + h) \neq i] \\ &= \lim_{h \to 0} \frac{P[X(S_i + h) = j \mid X(S_i -) = i]}{P[X(S_i + h) \neq i \mid X(S_i -) = i]} \quad (11) \\ &= \lim_{h \to 0} \frac{P_{ij}(h)}{1 - P_{ii}(h)} = \lim_{h \to 0} \frac{\frac{P_{ij}(h)}{h}}{\frac{1 - P_{ii}(h)}{h}} = \frac{Q_{ij}}{-Q_{ii}}, \quad (12) \end{aligned}$$

[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.

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

$$F_{ij} = 1 - e^{Q_{ii}t}, \quad t \geq 0,\ i, j \in \mathcal{S}, \tag{13}$$

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 ($Q_{x_0x_0} \neq 0$), the dwell time τ_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 τ_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

$$Q_{ij}(t) = P[X_1 = j, \tau_1 \leq t | X_0 = i] = F_{ij}(t)P_{ij} = (1 - e^{-\nu_i t})P_{ij}, \tag{15}$$

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))

$$P_{ij}(t) = P[X(t) = j | X(0) = i] = P[X(t) = j | X_0 = i], \quad i, j \in \mathcal{S}, \ \ 0 \leq t < \infty. \tag{16}$$

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

$$\begin{aligned} &P[X(t) = j, T_1 \le t \mid X_0 = i] \\ &\quad = E[P[X(t) = j, T_1 \le t \mid X_0 = i, X_1, T_1] \mid X_0 = i] \\ &\quad = E[\mathbf{1}_{\{T_1 \le t\}} \cdot P[X(t) = j \mid X_1, T_1, X_0 = i] \mid X_0 = i] \\ &\quad = E[\mathbf{1}_{\{T_1 \le t\}} \cdot P[X(t - T_1) = j \mid X_1, X_0 = i] \mid X_0 = i] \\ &\quad = E[\mathbf{1}_{\{T_1 \le t\}} P_{X_1, j}(t - T_1) \mid X_0 = i] \\ &\quad = \sum_{k \in \mathcal{S}} \int_0^t P_{kj}(t - s)\, dQ_{ik}(s). \end{aligned} \tag{20}$$

Substituting (19) and (20) into (18), we obtain (17).
Applying (16.23),

$$\begin{aligned} Q_{ij} &= \left.\frac{dP_{ij}}{dt}\right|_{t=0} \\ &= -\delta_{ij} \sum_{k \in \mathcal{S}} Q'_{ik}(0) + \sum_{k \in \mathcal{S}} P_{kj}(0) Q'_{ik}(0) \\ &= -\delta_{ij} \sum_{k \in \mathcal{S}} \nu_i P_{ik} + \sum_{k \in \mathcal{S}} \delta_{kj} \nu_i P_{ik}. \end{aligned}$$

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

$$\nu_i = -Q_{ii}, \quad P_{ij} = \frac{Q_{ij}}{\nu_i} = \frac{Q_{ij}}{-Q_{ii}}. \tag{23}$$

16.10* 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).

16.3 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 $\{\tilde{X}_n\}$.

16.14* (a) The LHS of (16.85) can be written as

$$\begin{aligned}
\text{LHS} &= \frac{P[\tilde{X}(t_m) = x_0, \tilde{X}(t_{m-1}) = x_1, \tilde{X}(t_{m-2}) = x_2, \ldots, \tilde{X}(t_0) = x_m]}{P[\tilde{X}(t_{m-1}) = x_1, \tilde{X}(t_{m-2}) = x_2, \ldots, \tilde{X}(t_0) = x_m]} \\
&= \frac{P[X(-t_m) = x_0, X(-t_{m-1}) = x_1, X(-t_{m-2}) = x_2, \ldots, X(-t_0) = x_m]}{P[X(-t_{m-1}) = x_1, X(-t_{m-2}) = x_2, \ldots, X(-t_0) = x_m]} \\
&= \frac{\pi_{x_0} P_{x_0 x_1}(t_m - t_{m-1}) P_{x_1 x_2}(t_{m-1} - t_{m-2}) \cdots P_{x_{m-1} x_m}(t_1 - 1_0)}{\pi_{x_1} P_{x_1 x_2}(t_{m-1} - t_{m-2}) \cdots P_{x_{m-1} x_m}(t_1 - 1_0)} \\
&= \frac{\pi_{x_0} P_{x_0 x_1}(t_m - t_{m-1})}{\pi_{x_1}},
\end{aligned} \tag{24}$$

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:

$$\pi_i \tilde{Q}_{ij} = \pi_j Q_{ij}, \quad i, j \in \mathcal{S}. \tag{16.64}$$

16.16* 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

$$Q_{ij} = \tilde{Q}_{ij}, \quad i, j \in \mathcal{S}. \tag{25}$$

Applying (25) in the reversed balance equations (16.64), leads to the conclusion that $\{X(t)\}$ is reversible if and only if

$$\pi_i Q_{ij} = \pi_j Q_{ji}, \quad i, j \in \mathcal{S}. \tag{16.65}$$

16.4 An application: phylogenetic tree and its Markov chain representation

16.21* (a) It is easy to verify that the (i, j) element of the matrix $\boldsymbol{\Pi Q}$ is given by

$$[\boldsymbol{\Pi Q}]_{ij} = \pi_i Q_{ij}, \quad i, j \in \mathcal{S}, \tag{26}$$

and that the (i, j) element of the matrix $(\boldsymbol{\Pi Q})^\top$ is given by

$$[(\boldsymbol{\Pi Q})^\top]_{ij} = \pi_j Q_{ji}, \quad i, j \in \mathcal{S}. \tag{27}$$

It is then clear that the detailed balance equations (16.73) hold if and only if

$$[\boldsymbol{\Pi Q}]_{ij} = [(\boldsymbol{\Pi Q})^\top]_{ij},$$

i.e., if and only if $\boldsymbol{\Pi Q}$ is a symmetric matrix.

(b) Given a DTMC with TPM $\boldsymbol{P}$ and stationary probability vector $\boldsymbol{\pi}$, we again define the matrix $\boldsymbol{\Pi} = \text{diag}[\pi_i, \ i \in \mathcal{S}]$. Next, we verify that the (i, j) element of the matrix $\boldsymbol{\Pi P}$ is given by

$$[\boldsymbol{\Pi P}]_{ij} = \pi_i P_{ij}, \quad i, j \in \mathcal{S}, \tag{28}$$

and that the (i, j) element of the matrix $(\boldsymbol{\Pi P})^\top$ is given by

$$[(\boldsymbol{\Pi P})^\top]_{ij} = \pi_j P_{ji}, \quad i, j \in \mathcal{S}. \tag{29}$$

It is then clear that the detailed balance equations (16.63) hold if and only if

$$[\boldsymbol{\Pi P}]_{ij} = [(\boldsymbol{\Pi P})^\top]_{ij},$$

i.e., if and only if $\boldsymbol{\Pi P}$ 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 $\boldsymbol{\Pi}\boldsymbol{P}(\tau)$ is symmetric for any $\tau > 0$. We have that $\boldsymbol{P}(\tau) = e^{\boldsymbol{Q}\tau}$. Hence, it suffices to show that

$$\boldsymbol{\Pi}e^{\boldsymbol{Q}\tau} = [\boldsymbol{\Pi}e^{\boldsymbol{Q}\tau}]^\top = e^{\boldsymbol{Q}^\top\tau}\boldsymbol{\Pi}. \tag{30}$$

By part (a), the CTMC is reversible if and only if

$$\boldsymbol{\Pi}\boldsymbol{Q} = \boldsymbol{Q}^\top\boldsymbol{\Pi}. \tag{31}$$

Now suppose that

$$\boldsymbol{\Pi}\boldsymbol{Q}^n = (\boldsymbol{Q}^\top)^n\boldsymbol{\Pi} \tag{32}$$

for $n \geq 1$. Then

$$\begin{aligned}\boldsymbol{\Pi}\boldsymbol{Q}^{n+1} &= (\boldsymbol{\Pi}\boldsymbol{Q}^n)\boldsymbol{Q} = (\boldsymbol{Q}^\top)^n\boldsymbol{\Pi}\boldsymbol{Q}\\ &= (\boldsymbol{Q}^\top)^n\boldsymbol{Q}^\top\boldsymbol{\Pi} = (\boldsymbol{Q}^\top)^{n+1}\boldsymbol{\Pi}.\end{aligned}$$

By the induction principle, (32) holds for all $n \geq 1$. Now we have

$$\begin{aligned}\boldsymbol{\Pi}e^{\boldsymbol{Q}\tau} &= \boldsymbol{\Pi}\sum_{k=0}^{\infty}\frac{(\boldsymbol{Q}\tau)^k}{k!} = \sum_{k=0}^{\infty}\frac{\boldsymbol{\Pi}\boldsymbol{Q}^k\tau^k}{k!}\\ &= \sum_{k=0}^{\infty}\frac{(\boldsymbol{Q}^\top)^k\tau^k\boldsymbol{\Pi}}{k!} = e^{\boldsymbol{Q}^\top\tau}\boldsymbol{\Pi},\end{aligned}$$

which establishes (30).

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

$$\boldsymbol{\pi}^\top\boldsymbol{Q} = 0, \quad \boldsymbol{\pi}^\top\boldsymbol{1} = 1.$$

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 $\boldsymbol{\Pi} = \text{diag}\{0.1, 0.2, 0.2, 0.5\}$. Applying (16.75), the mean substitutions that occur on an edge $e \in \mathcal{E}$ is given by

$$\kappa(e) = \text{Tr}\{\boldsymbol{\Pi}\boldsymbol{Q}\} = 0.66.$$

(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

$$\begin{aligned} P[\boldsymbol{X} = \boldsymbol{x}] &= P[X_0 = x_0]P[X_1 = x_1 \mid X_0 = 0]P[X_2 = x_2 \mid X_1 = x_1] \\ &\quad P[X_3 = x_3 \mid X_1 = x_1]P[X_4 = x_4 \mid X_0 = x_0] \\ &= \pi_{x_0} P_{x_0x_1} P_{x_1x_2} P_{x_1x_3} P_{x_0x_4}, \end{aligned}$$

where P_{ij} is the (i, j) element of the TPM $\boldsymbol{P}(e)$ given in part (a). The likelihood of the character χ_3 is given by

$$\begin{aligned} P_{\chi_3} &= P[X_2 = \text{C}, X_3 = \text{T}, X_4 = \text{A}] \\ &= \sum_{x_0, x_1} P[\boldsymbol{X} = (x_0, x_1, \text{C}, \text{T}, \text{A})] \\ &= \sum_{x_0} \pi_{x_0} P_{x_0\text{A}} \sum_{x_1} P_{x_1\text{C}} P_{x_1\text{T}} P_{x_0x_1} \\ &= \sum_{x_0} \pi_{x_0} P_{x_0\text{A}} f_{x_0}, \end{aligned} \tag{33}$$

where we define

$$f_{x_0} = \sum_{x_1} P_{x_1\text{C}} P_{x_1\text{T}} P_{x_0x_1}.$$

We compute to four decimal places,

$$f_\text{A} = 0.0694, f_\text{C} = 0.1121, f_\text{G} = 0.0694, f_\text{T} = 0.0865. \tag{34}$$

Substituting into (33), we obtain $P_{\chi_3} = 0.0080$.

17 Solutions for Chapter 17: Random Walk, Brownian Motion and Diffusion Process

17.1 Random Walk

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_is, 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.

17.2 Brownian Motion or Wiener Process

17.10* Derivation of (17.104) and (17.106).

(a) Let X be a RV with mean μ and variance σ^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 σ^2 as $\sigma \to 0$. Thus, if σ^2 becomes very small, we can ignore the last term. Recall the following properties of Dirac's delta function

$$\int_{-\infty}^{\infty} \delta(x-a)g(x)\,dx = g(a), \tag{1}$$

$$\int_{-\infty}^{\infty} \delta^{(k)}(x-a)g(x)\,dx = (-1)^k g^{(k)}(a), \tag{2}$$

where (2) can be derived from (1) by applying integration by parts k times. Thus, for very small σ^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 ϵ, 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 σ^2. By identifying μ as $x' + \beta(x',t)h$ and σ^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:

$$f(x, t+h|x_0, t_0) = \int f(x, t+h|x', t) f(x', t|x_0, t_0)\, dx'. \tag{3}$$

Then

$$\mathrm{LHS} = f(x, t|x_0, t_0) + \frac{\partial f(x, t|x_0, t_0)}{\partial t} + o(h).$$

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):

$$f(y, t+h|x', t) = \delta(x - x' - \beta(x',t)h) + \delta^{(2)}(x - x' - \beta(x',t)h)\frac{\alpha(x',t)h}{2} + o(h). \tag{4}$$

Then the RHS of (3) is

$$\begin{aligned}\mathrm{RHS} &= \int \left[\delta(x - x' - \beta(x',t)h) + \delta^{(2)}(x - x' - \beta(x',t)h)\frac{\alpha(x',t)h}{2}\right] f(x', t|x_0, t_0)\, dx' \\ &\triangleq I_1 + I_2 + o(h),\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \int \delta(x - x' - \beta(x', t)h) f(x', t|x_0, t_0) dx' \\
&= \int [\delta(x - x') - h\delta^{(1)}(x - x')\beta(x', t)] f(x', t|x_0, t_0)\, dx' + o(h) \\
&= f(x, t|x_0, t_0) - h \int \delta^{(1)}(x - x')[\beta(x', t) f(x', t|x_0, t_0)]\, dx' + o(h) \\
&= f(x, t|x_0, t_0) - h \frac{\partial(\beta(x, t) f(x, t|x_0, t_0))}{\partial x} + o(h),
\end{aligned} \tag{5}$$

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 (5):

$$\begin{aligned}
x - x' - \beta(x', t)h &= x - x' - h\left[\beta(x, t) - (x - x')\frac{\partial \beta(x, t)}{\partial x}\right] + o(h) \\
&= (x - x')B(x, t, h) - h\beta(x, t) + o(h),
\end{aligned}$$

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 (5) obtained using the property of $\delta^{(1)}(x)$.

17.3 Stochastic Differential Equations and Itô Process

17.16* Conditional mean and variance of the geometric Brownian motion.
We can write

$$\begin{aligned}
E[Y(t)|Y(u), 0 \le u \le s] &= E\left[e^{X(t)}|X(u), 0 \le u \le s\right] \\
&= E\left[e^{X(s)+X(t)-X(s)}|X(u), 0 \le u \le s\right] \\
&= e^{X(s)}E\left[e^{X(t)-X(s)}|X(u), 0 \le u \le s\right] \quad \text{(independent increments)} \\
&= Y(s)E\left[e^{X(t-s)-X(0)}\right] \quad \text{(temporal homogeneity)} \\
&= Y(s)E\left[e^{X(t-s)}\right] \quad \text{(because } X(0)=0\text{)}
\end{aligned}$$

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

$$\begin{aligned}\mathrm{Var}[Y(t)|Y(u), 0 \le u \le s] &= E[Y(t)^2|Y(u), 0 \le u \le s] - (E[Y(t)|Y(u), 0 \le u \le s])^2\\ &= Y(s)^2e^{2(\beta+\alpha)(t-s)} - \left(Y(s)e^{\left(\beta+\frac{\alpha}{2}\right)(t-s)}\right)^2\\ &= Y(s)^2e^{(2\beta+\alpha)(t-s)}\left(e^{\alpha(t-s)}-1\right).\end{aligned}$$

17.19* **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

18.4* 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

$$\nabla_{\boldsymbol{\theta}} E[\boldsymbol{T}^\top(\boldsymbol{X},\boldsymbol{\theta})] = \int_x f(\boldsymbol{x},\boldsymbol{\theta})(\nabla_{\boldsymbol{\theta}}\boldsymbol{T}^\top(\boldsymbol{x},\boldsymbol{\theta}))d\boldsymbol{x} + \int_x (\nabla_{\boldsymbol{\theta}} f(\boldsymbol{x},\boldsymbol{\theta}))\boldsymbol{T}^\top(\boldsymbol{x},\boldsymbol{\theta})d\boldsymbol{x} \quad (1)$$

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

$$\nabla_{\boldsymbol{\theta}} E[\boldsymbol{T}^\top(\boldsymbol{X},\boldsymbol{\theta})] = E[\nabla_{\boldsymbol{\theta}}\boldsymbol{T}^\top(\boldsymbol{X},\boldsymbol{\theta})] + E[s(\boldsymbol{X},\boldsymbol{\theta})\boldsymbol{T}^\top(\boldsymbol{X},\boldsymbol{\theta})] \quad (2)$$

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 $E[\boldsymbol{s}(\boldsymbol{X};\boldsymbol{\theta})] = \boldsymbol{0}$ 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

$$E[\boldsymbol{s}(\boldsymbol{X},\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top] = \text{Cov}[\boldsymbol{s}(\boldsymbol{X},\boldsymbol{\theta}), \hat{\boldsymbol{\theta}}] = \boldsymbol{I}.$$

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

$$\int f_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}} \log f_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\theta}) \left(\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}\right) dx = \boldsymbol{0}.$$

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.

18.7* The CRLB and a sufficient statistic.
Apply the inverse operation of the operator $\nabla_{\boldsymbol{\theta}}$ to both sides in (18.43), leading to

$$\log f_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\theta}) = \int (\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta})^\top \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})\, d\boldsymbol{\theta} + C(\boldsymbol{x}), \tag{4}$$

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

$$\int \boldsymbol{a}^\top\, d\boldsymbol{\theta} \triangleq \sum_{i=1}^{M} \int a_i(\boldsymbol{\theta})\, d\theta_i. \tag{5}$$

Thus,

$$L_{\boldsymbol{x}}(\boldsymbol{x}; \boldsymbol{\theta}) = h(\boldsymbol{x}) \exp\left(\int (\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta})^\top \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}\right) = \exp\left(\boldsymbol{\eta}(\boldsymbol{\theta})^\top \hat{\boldsymbol{\theta}}(\boldsymbol{x}) - A(\boldsymbol{\theta})\right), \tag{6}$$

where $h(\boldsymbol{x}) = \exp C(\boldsymbol{x})$, and

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = \int \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \ \text{ and } \ A(\boldsymbol{\theta}) = \int \boldsymbol{\theta}^\top \boldsymbol{\mathcal{I}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}. \tag{7}$$

Hence, it is apparent from Theorem 18.1 that an *efficient* estimate $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ is a *sufficient statistic* for estimating $\boldsymbol{\theta}$.

18.2 Hypothesis Testing and Statistical Decision

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

$$E_f\left[\log \frac{g(\boldsymbol{X})}{f(\boldsymbol{X})}\right] \leq \log E_f\left[\frac{g(\boldsymbol{X})}{f(\boldsymbol{X})}\right] = \log \int f(\boldsymbol{x}) \frac{g(\boldsymbol{x})}{f(\boldsymbol{x})}\, d\boldsymbol{x} = 0, \quad (1)$$

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

$$\int f(\boldsymbol{x}) \log g(\boldsymbol{x})\, d\boldsymbol{x} \leq \int f(\boldsymbol{x}) \log f(\boldsymbol{x})\, d\boldsymbol{x}, \quad (2)$$

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

$$J(g, \lambda) = \int f(\boldsymbol{x}) \log g(\boldsymbol{x})\, d\boldsymbol{x} + \lambda \left(\int g(\boldsymbol{x})\, d\boldsymbol{x} - 1 \right). \quad (3)$$

Differentiate it with respect to g, and λ and set them all to zero:

$$\frac{\partial J(\boldsymbol{g}, \lambda)}{\partial g(\boldsymbol{x})} = \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} + \lambda = 0, \ \text{ for } \ -\infty < \boldsymbol{x} < \infty,$$
$$\frac{\partial J(\boldsymbol{g}, \lambda)}{\partial \lambda} = \int g(\boldsymbol{x})\, d\boldsymbol{x} - 1 = 0.$$

From the first equation we find

$$f(\boldsymbol{x}) = -\lambda g(\boldsymbol{x}), \ \text{ for } \ -\infty < \boldsymbol{x} < \infty, \quad (4)$$

and substituting to the last equation (i.e., the original constraint equation) and using $\int f(\boldsymbol{x})\, d\boldsymbol{x} = 1$, we find

$$\lambda = -1. \quad (5)$$

Thus, the condition for a stationary point is

$$g(\boldsymbol{x}) = f(\boldsymbol{x}), \ \text{ for } \ -\infty < \boldsymbol{x} < \infty. \quad (6)$$

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 Expectation-Maximization Algorithm

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.

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 $S_{t-1} = 1 \to S_t = j, \;\; (t \geq 1)$ 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*.

20.9* 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:

$$\alpha_t(j,\boldsymbol{y}_0^t) = \sum_{i\in\mathcal{S}} \alpha_{t-1}(i,\boldsymbol{y}_0^{t-1})c(i;j,y_t),\ \ j\in\mathcal{S},\ 1\le t\le T, \tag{8}$$

20.4 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.

20.18* 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)

$$\begin{aligned}
\breve{\alpha}_t(0) &= \max \left\{ \breve{\alpha}_{t-1}(0),\ \ \breve{\alpha}_{t-1}(1) - A y_t - \frac{A^2}{2} \right\}, \\
\breve{\alpha}_t(1) &= \max \left\{ \breve{\alpha}_{t-1}(0) + A y_t - \frac{A^2}{2},\ \ \breve{\alpha}_{t-1}(1) \right\}
\end{aligned} \tag{9}$$

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 $S_{t-1} = 1$.

20.25* 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

$$\boxed{Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) \triangleq E\left[\log p(\boldsymbol{S}, \boldsymbol{y}; \boldsymbol{\theta}) |\, \boldsymbol{y}; \boldsymbol{\theta}^{(p)} \right] = \sum_{\boldsymbol{s}} p(\boldsymbol{s}|\boldsymbol{y}; \boldsymbol{\theta}^{(p)}) \log p(\boldsymbol{s}, \boldsymbol{y}; \boldsymbol{\theta}),} \tag{10}$$

where $\boldsymbol{\theta}^{(p)}$ is the pth 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)

$$p(\boldsymbol{S}, \boldsymbol{y}; \boldsymbol{\theta}) = \alpha(S_0, y_0) \prod_{t=1}^{T} c(S_{t-1}; S_t, y_t). \tag{11}$$

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

$$p(\boldsymbol{S}, \boldsymbol{y}; \boldsymbol{\theta}) = \alpha_0(S_0, y_0) \prod_{i,j \in \mathcal{S}, k \in \mathcal{Y}} c(i; j, k)^{M(i;j,k)}, \tag{12}$$

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

$$\log p(\boldsymbol{S}, \boldsymbol{y}; \boldsymbol{\theta}) = \log \alpha_0(S_0, y_0) + \sum_{i,j \in \mathcal{S}, k \in \mathcal{Y}} M(i; j, k) \log c(i; j, k). \tag{13}$$

Thus, we can write $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(p)})$ as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = Q_0(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) + Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}), \tag{14}$$

where

$$Q_0(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = E\left[\log \alpha_0(S_0, y_0)|\boldsymbol{y}; \boldsymbol{\theta}^{(p)}\right], \tag{15}$$

$$Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = \sum_{i,j \in \mathcal{S}, k \in \mathcal{Y}} E\left[M(i; j, k)|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}\right] \log c(i; j, k). \tag{16}$$

(b) We denote the above conditional expectation of the random variable $M(i; j, k)$ as

$$E\left[M(i; j, k)|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}\right] \triangleq \overline{M}^{(p)}(i; j, k|\boldsymbol{y}). \tag{17}$$

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:

$$\overline{M}^{(p)}(i; j, k|\boldsymbol{y}) = \frac{\sum_{t=1}^{T} \alpha_{t-1}^{(p)}(i, \boldsymbol{y}_0^{t-1}) c^{(p)}(i; j, k) \beta_t^{(p)}(j; \boldsymbol{y}_{t+1}^T) \delta_{y_t, k}}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})}, \tag{18}$$

where $c^{(p)}(i; j, y_t)$ is the pth 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.

(18) 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 (18).
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 (21) to be given below.

(c) **Maximization step:**

Since the model parameters are the joint probability and conditional joint probability distributions, they must satisfy constraints

$$\sum_{j\in\mathcal{S},k\in\mathcal{Y}}\alpha_0(j,k)=1, \tag{19}$$

$$\sum_{j\in\mathcal{S},k\in\mathcal{Y}}c(i;j,k)=1,\ \ \text{for all } i\in\mathcal{S}. \tag{20}$$

We wish to find the value of $\boldsymbol{\theta}$ that maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ under the set of constraints (20). 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

$$\overline{M}_0^{(p)}(j,k|\boldsymbol{y})=\begin{cases}\frac{\alpha_0^{(p)}(j,y_0)\beta_0^{(p)}(j;\boldsymbol{y}_1^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})}, & k=y_0\\ 0, & k\neq y_0\end{cases} \tag{21}$$

Thus, $Q_0(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ of (15) can be written as

$$Q_0(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})=\sum_{j\in\mathcal{S}}\overline{M}_0^{(p)}(j,k|\boldsymbol{y})\log\alpha_0(j,y_0). \tag{22}$$

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

$$\alpha_0^{(p+1)}(j,y_0)=\frac{\overline{M}_0^{(p)}(j,k|\boldsymbol{y})}{\sum_{j\in\mathcal{S},k\in\mathcal{Y}}\overline{M}_0^{(p)}(j,k|\boldsymbol{y})}=\frac{\alpha_0^{(p)}(j,y_0)\beta_0^{(p)}(j;\boldsymbol{y}_1^T)}{\sum_{j\in\mathcal{S}}\alpha_0^{(p)}(j,y_0)\beta_0^{(p)}(j;\boldsymbol{y}_1^T)} \tag{23}$$

$$=\frac{\alpha_0^{(p)}(j,y_0)\beta_0^{(p)}(j;\boldsymbol{y}_1^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})}. \tag{24}$$

Maximization of $Q_1(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ is equivalent to maximizing the following expression for each $i\in\mathcal{S}$.

$$\sum_{j\in\mathcal{S},k\in\mathcal{Y}}\overline{M}^{(p)}(i;j,k|\boldsymbol{y})\log c(i;j,k),\quad i\in\mathcal{S}. \tag{25}$$

By using the log sum inequality (10.21) again, we find that the maximum of (25) can be achieved when

$$c(i;j,k)=\frac{\overline{M}^{(p)}(i;j,k|\boldsymbol{y})}{\sum_{j\in\mathcal{S},k\in\mathcal{Y}}\overline{M}^{(p)}(i;j,k|\boldsymbol{y})},\ \text{ for all }\ j\in\mathcal{S},k\in\mathcal{Y}. \tag{26}$$

By substituting (18) into (26), we obtain the following expression for the $(p+1)$st update of the model parameter $c(i;j,k)$:

$$c^{(p+1)}(i;j,k)=\frac{\sum_{t=1}^T\alpha_{t-1}^{(p)}(i,\boldsymbol{y}_0^{t-1})c^{(p)}(i;j,k)\beta_t^{(p)}(j;\boldsymbol{y}_{t+1}^T)\delta_{y_t,k}}{\sum_{j\in\mathcal{S}}\sum_{t=1}^T\alpha_{t-1}^{(p)}(i,\boldsymbol{y}_0^{t-1})c^{(p)}(i;j,y_t)\beta_t^{(p)}(j;\boldsymbol{y}_{t+1}^T)}$$

$$=\frac{\sum_{t=1}^T\xi_{t-1}^{(p)}(i,j|\boldsymbol{y})\delta_{y_t,k}}{\sum_{j\in\mathcal{S}}\sum_{t=1}^T\xi_{t-1}^{(p)}(i,j|\boldsymbol{y})}, \tag{27}$$

where we used the property (20.142) of the APP $\xi_t(i,j|\boldsymbol{y})$:

$$\boxed{\xi_t^{(p)}(i,j|\boldsymbol{y})=\frac{\alpha_t^{(p)}(i,\boldsymbol{y}_0^t)c^{(p)}(i;j,y_{t+1})\beta_{t+1}^{(p)}(j;\boldsymbol{y}_{t+2}^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})},\ \ i,j\in\mathcal{S},\ \ t\in\mathcal{T},} \tag{28}$$

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):

$$\gamma_t^{(p)}(i|\boldsymbol{y})=\sum_{j\in\mathcal{S}}\xi_t^{(p)}(i,j|\boldsymbol{y})=\frac{\alpha_t^{(p)}(i,\boldsymbol{y}_0^t)\beta_t^{(p)}(i;\boldsymbol{y}_{t+1}^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})} \tag{29}$$

with $\gamma_T^{(p)}(i|\boldsymbol{y})$ given, from (20.85) and (20.63), by

$$\gamma_T^{(p)}(i|\boldsymbol{y})=\frac{\alpha_T^{(p)}(i,\boldsymbol{y}_0^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})}=\frac{\alpha_T^{(p)}(i,\boldsymbol{y}_0^T)}{\sum_{i\in\mathcal{S}}\alpha_T^{(p)}(i,\boldsymbol{y}_0^T)},\ i\in\mathcal{S}. \tag{30}$$

Algorithm 20.1 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.

Thus, we can express $c^{(p+1)}(i; j, k)$ of (27), using (28) and (20.148), as

$$\boxed{c^{(p+1)}(i; j, k) = \frac{\sum_{t=1:\ y_t=k}^{T} \xi_{t-1}^{(p)}(i, j|\boldsymbol{y})}{\sum_{t=1}^{T} \gamma_{t-1}^{(p)}(i|\boldsymbol{y})},} \tag{31}$$

where $\sum_{t=1:\ y_t=k}^{T}$ in the numerator means summation with respect to t for which $y_t = k$. Algorithm 20.1 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

$$\Xi(i,j,k) = \sum_{t=1}^{T} \xi_t(i,j|\boldsymbol{y})\delta_{y_t=k} \tag{32}$$

$$\Gamma(i,k) = \sum_{t=1}^{T} \gamma_t(i|\boldsymbol{y})\delta_{y_t=k}. \tag{33}$$

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

$$\log p(\boldsymbol{S},\boldsymbol{y}|\boldsymbol{\theta}) = \log \pi_0(S_0) + \sum_{i,j\in\mathcal{S}} M(i,j)\log a(i;j) + \sum_{j\in\mathcal{S},k\in\mathcal{Y}} N(j,k)\log b(j;k). \tag{34}$$

Thus we can write $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)})$ as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(p)}) = E[\log \pi_0(S_0)|\boldsymbol{y},\boldsymbol{\theta}^{(p)}] + \sum_{i,j\in\mathcal{S}} E\left[M(i,j)|\boldsymbol{y},\boldsymbol{\theta}^{(p)}\right]\log a(i;j)$$
$$+ \sum_{j\in\mathcal{S},k\in\mathcal{Y}} E\left[N(j,k)|\boldsymbol{y},\boldsymbol{\theta}^{(p)}\right]\log b(j;k). \tag{35}$$

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:

$$\bar{M}^{(p)}(i;j|\boldsymbol{y}) = \frac{\sum_{t=1}^{T} \alpha_{t-1}^{(p)}(i,\boldsymbol{y}_0^{t-1})a^{(p)}(i;j)\beta_t^{(p)}(j;\boldsymbol{y}_{t+1}^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})} \tag{36}$$

$$\bar{N}^{(p)}(j,k) = \frac{\sum_{t=0}^{T} \alpha_t^{(p)}(j,\boldsymbol{y}_0^t)\beta_t^{(p)}(j;\boldsymbol{y}_{t+1}^T)\delta(y_t,k)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})} \tag{37}$$

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 (35) 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 (35) can be maximized when we set the model parameters to the following values in the $(p+1)$st iteration:

$$
\begin{aligned}
\pi_0^{(p+1)}(j) &= P[S_0 = j|\boldsymbol{y}, \boldsymbol{\theta}^{(p)}] = \gamma_0^{(p)}(j|\boldsymbol{y}) = \frac{\alpha_0^{(p)}(j, y_0)\beta_0^{(p)}(j; \boldsymbol{y}_1^T)}{L_{\boldsymbol{y}}(\boldsymbol{\theta}^{(p)})} \\
a^{(p+1)}(i; j) &= \frac{\bar{M}^{(p)}(i, j|\boldsymbol{y})}{\sum_{j \in \mathcal{S}} \bar{M}^{(p)}(i, j|\boldsymbol{y})} \quad i, j \in \mathcal{S} \\
b^{(p+1)}(j; k) &= \frac{\bar{N}^{(p)}(j, k|\boldsymbol{y})}{\sum_{k \in \mathcal{Y}} \bar{N}^{(p)}(j, k|\boldsymbol{y})}, \quad j \in \mathcal{S}.
\end{aligned}
$$

The mean values $\bar{M}^{(p)}(i; j|\boldsymbol{y}$ of (36) and $\bar{N}^{(p)}(j, k)$ of (37) 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:

$$
\bar{M}^{(p)}(i; j|\boldsymbol{y}) = \sum_{t=0}^{T} \xi_{t-1}^{(p)}(i, j|\boldsymbol{y}), \text{ and } \bar{N}^{(p)}(j, k) = \sum_{t=0}^{T} \gamma_t^{(p)}(j|\boldsymbol{y}) b^{(p)}(j; k)\delta(y_t, k), \qquad (38)
$$

Thus, we have the expressions (20.105) as the M-step solution:

20.5 Application Example: Parameter Estimation in Mixture Distributions

21 Solutions for Chapter 21: Elements of Machine Learning

21.4* Sum-product algorithm for a phylogenetic tree.

The character χ 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 χ 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

$$\begin{aligned} P[\boldsymbol{X} = \boldsymbol{x}] &= P[X_0 = x_0] \prod_{v \in \tilde{\mathcal{V}} \setminus \{0\}} P[X_v = x_v \big| X_u = x_u, \ u \leq v] \\ &\overset{(a)}{=} \pi_{x_0}(0) \prod_{e=(u,v) \in \mathcal{E}} P_{x_u x_v}(e), \end{aligned} \tag{2}$$

where the Markov property (16.68) is applied in step (a). Combining (1) and (2) we obtain

$$P_\chi = \sum_{\boldsymbol{x}:\tilde{\boldsymbol{x}}=\boldsymbol{w}} \pi_{x_0}(0) \prod_{e=(u,v) \in \mathcal{E}} P_{x_u x_v}(e). \tag{3}$$

We now develop a sum-product algorithm to compute P_χ 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

$$P_\chi = \sum_{i \in \mathcal{S}} P[\tilde{\boldsymbol{X}} = \boldsymbol{w} | X_0 = i] P[X_0 = i] = \sum_{i \in \mathcal{S}} P[\tilde{\boldsymbol{X}} = \boldsymbol{w} | X_0 = i] \pi_i(0). \tag{4}$$

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

$$\begin{aligned} P[\tilde{\boldsymbol{X}}^u = \boldsymbol{w}^u | X_u = i] &= P[\tilde{\boldsymbol{X}}^v = \boldsymbol{w}^v,\ v \in \mathrm{ch}(u) | X_v = i] \\ &= \prod_{v \in \mathrm{ch}(u)} \sum_{j \in \mathcal{S}} P[\tilde{\boldsymbol{X}}^v = \boldsymbol{w}^v, X_v = j | X_v = i] \\ &\overset{(a)}{=} \prod_{v \in \mathrm{ch}(u)} \sum_{j \in \mathcal{S}} P[\tilde{\boldsymbol{X}}^v = \boldsymbol{w}^v | X_v = i] P_{ij}((u,v)), \end{aligned} \tag{5}$$

where the Markov property (16.68) was used in step (a). Conditioning on X_0 and applying (5) in (1), we have

$$\begin{aligned} P_\chi = P[\tilde{\boldsymbol{X}} = \boldsymbol{w}] &= \sum_{i \in \mathcal{S}} P[\tilde{\boldsymbol{X}} = \boldsymbol{w} | X_0 = i] \pi_d(0) \\ &= \sum_{i \in \mathcal{S}} \pi_d(0) \prod_{v \in \mathrm{ch}(0)} \sum_{j \in \mathcal{S}} P[\tilde{\boldsymbol{X}}^v = \boldsymbol{w}^v | X_v = c] P_{ij}((0,v)). \end{aligned} \tag{6}$$

By applying (5) recursively to (6), we obtain an efficient sum-product algorithm to compute P_χ: 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

$$0 < x < \begin{cases} \frac{\pi_2}{\pi_1} & \text{if} \ \ \pi_2 \leq 0.5 \\ 1 & \text{otherwise} \end{cases}$$

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.

21.7* Stationary distribution in the block MH algorithm.

$$\begin{aligned}
&\int_{\boldsymbol{x}_1}\int_{\boldsymbol{x}_2} f_1(\boldsymbol{x}_1;\boldsymbol{y}_1|\boldsymbol{x}_2) f_2(\boldsymbol{x}_2;\boldsymbol{y}_2|\boldsymbol{y}_1)\pi(\boldsymbol{x}_1,\boldsymbol{x}_2)\,d\boldsymbol{x}_1\,d\boldsymbol{x}_2 \\
&= \int_{\boldsymbol{x}_2} f_2(\boldsymbol{x}_2;\boldsymbol{y}_2|\boldsymbol{y}_1)\pi_2(\boldsymbol{x}_2)\,d\boldsymbol{x}_2 \left(\int_{\boldsymbol{x}_1} f_1(\boldsymbol{x}_1;\boldsymbol{y}_1|\boldsymbol{x}_2)\pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\,d\boldsymbol{x}_1\right) && (9)\\
&= \int_{\boldsymbol{x}_2} f_2(\boldsymbol{x}_2;\boldsymbol{y}_2|\boldsymbol{y}_1)\pi_2(\boldsymbol{x}_2)\pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\,d\boldsymbol{x}_2 && (10)\\
&= \int_{\boldsymbol{x}_2} f_2(\boldsymbol{x}_2;\boldsymbol{y}_2|\boldsymbol{y}_1)\pi(\boldsymbol{y}_1,\boldsymbol{x}_2)\,d\boldsymbol{x}_2 && (11)\\
&= \pi_1(\boldsymbol{y}_1)\int_{\boldsymbol{x}_2} f_2(\boldsymbol{x}_2;\boldsymbol{y}_2|\boldsymbol{y}_1)\pi_2(\boldsymbol{x}_2|\boldsymbol{y}_1)\,d\boldsymbol{x}_2 && (12)\\
&= \pi_1(\boldsymbol{y}_1)\pi_2(\boldsymbol{x}_2|\boldsymbol{y}_1) && (13)\\
&= \pi(\boldsymbol{y}_1,\boldsymbol{y}_2). && (14)
\end{aligned}$$

where equations (9), (11), (12) and (14) follow from Bayes' rule, while (10) and (13) follow from (21.50) and (21.51), respectively.

21.8* Stationary distribution in the Gibbs sampler.

$$f(\boldsymbol{x};\boldsymbol{y}) = \pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\pi_2(\boldsymbol{y}_2|\boldsymbol{y}_1), \ \ \text{where } \boldsymbol{x} = (\boldsymbol{x}_1,\boldsymbol{x}_2), \ \ \boldsymbol{y} = (\boldsymbol{y}_1, by_2)$$

$$\begin{aligned}
\int_{\boldsymbol{x}} \pi(\boldsymbol{x}) f(\boldsymbol{x};\boldsymbol{y}) d\boldsymbol{x} &= \int_{\boldsymbol{x}_1}\int_{\boldsymbol{x}_2} \pi(\boldsymbol{x}_1, bx_2)\pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\pi_2(\boldsymbol{y}_2|\boldsymbol{y}_1) d\boldsymbol{x}_1 dx_2 \\
&= \int_{\boldsymbol{x}_1}\int_{\boldsymbol{x}_2} \pi_1(\boldsymbol{x}_1|bx_2)\pi_2(\boldsymbol{x}_2)\pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\pi_2(\boldsymbol{y}_2|\boldsymbol{y}_1) d\boldsymbol{x}_1 dx_2 \\
&= \int_{\boldsymbol{x}_2} \pi_2(\boldsymbol{x}_2)\pi_1(\boldsymbol{y}_1|\boldsymbol{x}_2)\pi(\boldsymbol{y}_2|\boldsymbol{y}_1) d\boldsymbol{x}_2 \\
&= \int_{\boldsymbol{x}_2} \pi(\boldsymbol{y}_1, \boldsymbol{x}_2)\pi(\boldsymbol{y}_2|\boldsymbol{y}_1) d\boldsymbol{x}_2 \\
&= \pi_1(\boldsymbol{y}_1)\pi_2(\boldsymbol{y}_2|\boldsymbol{y}_1) = \pi(\boldsymbol{y}_1, \boldsymbol{y}_2) \\
&= \pi(\boldsymbol{y}).
\end{aligned}$$

22 Solutions for Chapter 22: Filtering and Prediction of Random Processes

22.1 Conditional Expectation and MMSE Estimation

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

$$\begin{aligned}\langle S - E[S|\boldsymbol{X}], h(\boldsymbol{X})\rangle &\triangleq E\left[(S - E[S|\boldsymbol{X}])h^*(\boldsymbol{X})\right] \\ &= E_x\left[\left(E_{s|x}(S - E[S|\boldsymbol{X}])\right]\right) h^*(\boldsymbol{X})\right] = 0,\end{aligned} \tag{2}$$

because the term in the parenthesis is zero for all $\boldsymbol{X}$:

$$E_{s|x}(S - E[S|\boldsymbol{X}]) = E_{s|x}(S - E[S|X]) = E_{s|x}[S|X] - E_{s|x}[S|X] = 0. \tag{3}$$

22.13* Regression coefficient estimates.

Note: There is a typo in (22.57) and (22.61). (22.57) should be

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top\right]^{-1} \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}})y_j. \tag{4}$$

and the second equation of (22.61) should be

$$\mathrm{Var}[\hat{\beta}_0] = \frac{\sigma_\epsilon^2}{n}\left(1 + n\overline{\boldsymbol{x}}^\top \left[\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top\right]^{-1} \overline{\boldsymbol{x}}\right). \tag{5}$$

We first derive (4), 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 $\mathbf{0}$:

$$\frac{\partial Q}{\partial \boldsymbol{\beta}^\top} = -2\sum_{j=1}^{n}(y_j - \beta_0 - \boldsymbol{\beta}^\top \boldsymbol{x}_j)\boldsymbol{x}_j = \mathbf{0}, \tag{6}$$

Similarly, by differentiating Q with respect to β_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 (6), we obtain

$$\sum_{j=1}^{n}(y_j - \overline{y} + \hat{\boldsymbol{\beta}}^\top \overline{\boldsymbol{x}} - \hat{\boldsymbol{\beta}}^\top \boldsymbol{x}_j)\boldsymbol{x}_j = \sum_{j=1}^{n}[(y_j - \overline{y}) - \hat{\boldsymbol{\beta}}^\top(\boldsymbol{x}_j - \overline{\boldsymbol{x}})]\boldsymbol{x}_j = 0. \tag{7}$$

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 (7) as

$$\sum_{j=1}^{n}[(y_j - \overline{y}) - \hat{\boldsymbol{\beta}}^\top(\boldsymbol{x}_j - \overline{\boldsymbol{x}})](\boldsymbol{x}_j - \overline{\boldsymbol{x}}) = \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}})[(y_j - \overline{y}) - (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \hat{\boldsymbol{\beta}}] = 0 \tag{8}$$

or

$$\boldsymbol{\Sigma}_{\boldsymbol{x}} \hat{\boldsymbol{\beta}} = \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}})(y_j - \overline{y}), \tag{9}$$

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,

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}})(y_j - \overline{y}) = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n}(\boldsymbol{x}_j - \overline{\boldsymbol{x}}) y_j \tag{10}$$

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 ϵ_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,

$$E[y_j - \overline{y}] = \boldsymbol{\beta}^\top (\boldsymbol{x}_j - \overline{\boldsymbol{x}}) = (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \boldsymbol{\beta}. \tag{11}$$

Taking the expectation of (10) and using (11), we obtain

$$\begin{aligned} E[\hat{\boldsymbol{\beta}}] &= \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}}) E[y_j - \overline{y}] \\ &= \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \left(\sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \right) \boldsymbol{\beta} = \boldsymbol{\beta}. \end{aligned} \tag{12}$$

In order to derive the first equation of (22.62), note that

$$y_j = E[y_j] + \epsilon_j.$$

Similar to the derivation of (12), we obtain

$$\hat{\boldsymbol{\beta}} = E[\hat{\boldsymbol{\beta}}] + \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}}) \epsilon_j = \boldsymbol{\beta} + \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \overline{\boldsymbol{x}}) \epsilon_j, \tag{13}$$

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 ϵ_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):

$$E[\hat{\beta}_0] = E[\overline{y}] - E[\hat{\boldsymbol{\beta}}^\top \overline{\boldsymbol{x}}] = \beta_0 + \boldsymbol{\beta}^\top \overline{\boldsymbol{x}} - \boldsymbol{\beta}^\top \overline{\boldsymbol{x}} = \beta_0. \tag{14}$$

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 (13) for $\hat{\boldsymbol{\beta}}$. Thus,

$$\hat{\beta}_0 = \beta_0 + \sum_{j=1}^{n} \left(\frac{1}{n} - (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \right) \epsilon_j \tag{15}$$

Therefore,

$$\begin{aligned}
\mathrm{Var}[\hat{\boldsymbol{\beta}}_0] &= E\left[\epsilon_j \epsilon_k \sum_{j=1}^{n} \left(\frac{1}{n} - (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \right) \sum_{k=1}^{n} \left(\frac{1}{n} - (\boldsymbol{x}_k - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \right) \right] \\
&= \frac{1}{n^2} \sum_j \sum_k \sigma_\epsilon^2 \delta_{jk} + \sum_j \sum_k \sigma_\epsilon^2 \delta_{jk} \overline{\boldsymbol{x}}^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_k - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \\
&- \frac{1}{n} \sum_j \sum_k \sigma_\epsilon^2 \delta_{jk} (\boldsymbol{x}_j - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} - \frac{1}{n} \sum_j \sum_k \sigma_\epsilon^2 \delta_{jk} (\boldsymbol{x}_k - \overline{\boldsymbol{x}})^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} \\
&= \frac{\sigma_\epsilon^2}{n} + \sigma_\epsilon^2 \overline{\boldsymbol{x}}^\top \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \overline{\boldsymbol{x}} - 0 - 0 \\
&= \frac{\sigma_\epsilon^2}{n} \left(1 + n \overline{\boldsymbol{x}}^\top \left[\sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^\top \right]^{-1} \overline{\boldsymbol{x}} \right).
\end{aligned} \tag{16}$$

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

$$R_{yy}[d] = \sum_{k=0}^{n} \sum_{j=0}^{n} h^*[k] h[j] R_{xx}[d + j - k], \tag{17}$$

which in matrix form becomes

$$R_{yy}[d] = \boldsymbol{h}^H \boldsymbol{R}_{xx}[d] \boldsymbol{h},$$

where $\boldsymbol{h}^H = (h^*[0], h^*[1], \ldots, h^*[n])$.

22.3 Kalman Filter

23 Solutions for Chapter 23 Queuing and Loss Models

23.1 Introduction

23.2 Little's Formula

23.3 Queueing Models

23.8* Derivation of the waiting time distribution (23.37).
From (23.36) we have

$$
\begin{aligned}
F_W(x) &= 1-\rho+(1-\rho)\sum_{n=1}^{\infty}\rho^n-(1-\rho)e^{-\mu x}\sum_{n=1}^{\infty}\sum_{j=0}^{n-1}\rho^n\frac{(\mu x)^j}{j!}\\
&= 1-(1-\rho)e^{-\mu x}\sum_{j=0}^{\infty}\sum_{n=j+1}^{\infty}\rho^n\frac{(\mu x)^j}{j!}\\
&= 1-(1-\rho)e^{-\mu x}\sum_{j=0}^{\infty}\frac{\rho^{j+1}}{1-\rho}\frac{(\mu x)^j}{j!}=1-\rho e^{-\mu x}\sum_{j=0}^{\infty}\frac{(\rho\mu x)^j}{j!}\\
&= 1-\rho e^{-\mu(1-\rho)x}.
\end{aligned}
$$

23.10* 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

$$p_n(t)=\frac{(\frac{\lambda}{\mu}(1-e^{-\mu t}))^j}{j!}\exp\left\{-\frac{\lambda}{\mu}(1-e^{-\mu t})\right\},\ 0\leq n<\infty. \tag{1}$$

23.15* Waiting time distribution in the M/M/m queue.

(a)

$$F_W^c(x)\sum_{n=0}^{\infty} a_{m+n}F_W^c(x|m+n) = \sum_{n=0}^{\infty} \pi_{m+n}F_W^c(x|m+n)$$
$$= F_W^c(0)(1-\rho)\sum_{n=0}^{\infty} \rho^n F_W^c(x|m+n), \tag{2}$$

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 (2) 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 ith completion, $i = 2, 3, \ldots, n+1$, as illustrated in Figure 23.1.

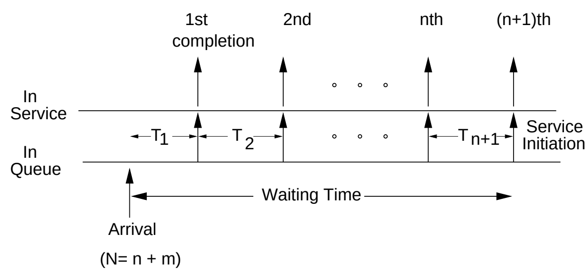


Figure 23.1 Relationship between the waiting time and service completion intervals T_i's in M/M/m.

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 jth server. The X_j's are independent and identically distributed with complementary distribution function

$$F_{X_j}^c(t) = P\{X_j \ge t\}e^{-\mu t}.$$

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

$$F_T^c(t) = P\{T \ge t\} = P\{X_j \ge t : \forall j,\ 1 \le j \le m\}$$
$$= \prod_{j=1}^{m} P\{X_j \ge t\} = \prod_{j=1}^{m} e^{-\mu t} = e^{-m\mu t}.$$

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 (2) gives

$$F_W^c(x) = F_W^c(0)e^{-m\mu x}(1-\rho)\sum_{n=0}^{\infty}\rho^n\sum_{j=0}^{n}\frac{(m\mu x)^j}{j!}. \tag{3}$$

The double summation over (n, j) can be rewritten as a double summation over (k, j), where $k = n - j$, as follows:

$$\begin{aligned}\sum_{n=0}^{\infty}\sum_{j=0}^{n}\rho^n\frac{(m\mu x)^j}{j!} &= \sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\rho^{k+j}\frac{(m\mu x)^j}{j!} \\ &= \sum_{k=0}^{\infty}\rho^k e^{\rho m\mu x} = \frac{1}{1-\rho}e^{\rho m\mu x} \end{aligned} \tag{4}$$

Using (4)in (3) yields where we interchanged the order of summation and using the formula for a geometric series, we obtain

$$F_W^c(x) = F_W^c(0)e^{-m\mu(1-\rho)x} \tag{5}$$

or (23.149).

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.,

$$F_W^c(x|n) = P[W > x|N = n] = 0, \quad 0 \le n \le m-1, \;\; x \ge 0. \tag{6}$$

On the other hand, when $n \ge m$, the waiting time is given by:

$$W = R_1 + S_2 + \cdots + S_{n-m+1}, \tag{7}$$

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 nth 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 μ. 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.,

$$F_W^c(x|n) = e^{-\mu x}\sum_{j=0}^{n-m}\frac{(\mu x)^j}{j!}, \quad m \le n \le K. \tag{8}$$

Alternatively, we can write

$$F_W^c(x|n+m) = e^{-\mu x}\sum_{j=0}^{n}\frac{(\mu x)^j}{j!} = Q(n; m\mu x), \; 0 \le n \le K-m, \tag{9}$$

which is (23.71). Therefore,

$$F_W^c(x) = \sum_{n=0}^{K-m} a_n(K) F_W^c(x|n+m). \tag{10}$$

Since $a_n(K) = \pi_n(K-1)$ for the $M(K)/M/m$ system, we have:

$$F_W^c(x) = \sum_{n=0}^{K-m} \pi_{n+m}(K-1) F_W^c(x|n+m), \tag{11}$$

which is (23.70), where $F_W^c(x|n+m)$ is given by (9). 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.,

$$F_W^c(x) = \sum_{n=0}^{K-m-1} \pi_{n+m}(K-1) F_W^c(x|n+m). \tag{23.70'}$$

23.31* 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 $f_W^*(s)$.

23.4 Loss Models

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 λ_n and μ_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).

23.43* 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).

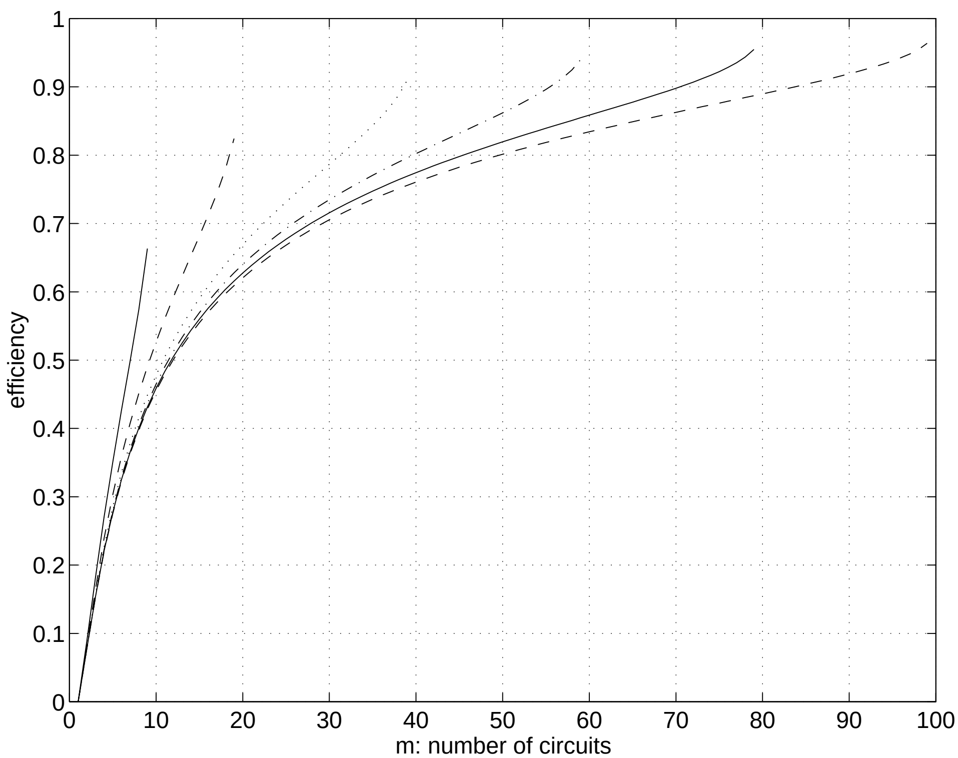


Figure Exercise 3.2-8: The efficiency η vs. the number of circuits (output lines) m. The number of input lines (sources) is $K = 10, 20, 40, 60, 80, 100$, and the specified QoS is $L = 0.01$.

(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

$$\pi_{\boldsymbol{N}}(\boldsymbol{n}|m, K_1) = \frac{1}{G(m, K_1)} \binom{K_1}{n_1} r_1^{n_1} \frac{a_2^{n_2}}{n_2!}, \quad \boldsymbol{n} \in \mathcal{F}_{\boldsymbol{N}}(m, K_1), \tag{12}$$

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 (12) 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.