8

Random Vectors

In this chapter, we expand on the concepts presented in Chapter 5. While Chapter 5 introduced the CDF and PDF of n random variables $X_1, \ldots, X_n$, this chapter focuses on the random vector $\mathbf{X} = \begin{bmatrix} X_1 & \cdots & X_n \end{bmatrix}'$. A random vector treats a collection of n random variables as a single entity. Thus, vector notation provides a concise representation of relationships that would otherwise be extremely difficult to represent.

The first section of this chapter presents vector notation for a set of random variables and the associated probability functions. The subsequent sections define marginal probability functions of subsets of n random variables, n independent random variables, independent random vectors, and expected values of functions of n random variables. We then introduce the covariance matrix and correlation matrix, two collections of expected values that play an important role in stochastic processes and in estimation of random variables. The final two sections cover Gaussian random vectors and the application of MATLAB, which is especially useful in working with multiple random variables.

8.1 Vector Notation

> A random vector with n dimensions is a concise representation of a set of n random variables. There is a corresponding notation for the probability model (CDF, PMF, or PDF) of a random vector.

When an experiment produces two or more random variables, vector and matrix notation provide a concise representation of probability models and their properties. This section presents a set of definitions that establish the mathematical notation of random vectors. We use boldface notation $\mathbf{x}$ for a column vector. Row vectors are transposed column vectors; $\mathbf{x}'$ is a row vector. The components of a column vector are, by definition, written in a column. However; to save space, we will often use the transpose of a row vector to display a column vector: $\mathbf{y} = [y_1 \ \cdots \ y_n]'$ is a column vector.

Definition 8.1 — Random Vector

A ***random vector*** *is a column vector* $\mathbf{X} = [X_1 \ \cdots \ X_n]'$*. Each* X_i *is a random variable.*

A random variable is a random vector with $n = 1$. The sample values of the components of a random vector constitute a column vector.

Definition 8.2 — Vector Sample Value

A sample value of a random vector *is a column vector* $\mathbf{x} = [x_1 \ \cdots \ x_n]'$*. The ith component,* x_i*, of the vector* $\mathbf{x}$ *is a sample value of a random variable,* X_i*.*

Following our convention for random variables, the uppercase $\mathbf{X}$ is the random vector and the lowercase $\mathbf{x}$ is a sample value of $\mathbf{X}$. However, we also use boldface capitals such as $\mathbf{A}$ and $\mathbf{B}$ to denote matrices with components that are not random variables. It will be clear from the context whether $\mathbf{A}$ is a matrix of numbers, a matrix of random variables, or a random vector.

The CDF, PMF, or PDF of a random vector is the joint CDF, joint PMF, or joint PDF of the components.

Definition 8.3 — Random Vector Probability Functions

(a) The ***CDF of a random vector*** $\mathbf{X}$ *is*

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1,\ldots,X_n}(x_1,\ldots,x_n).$$

(b) The ***PMF of a discrete random vector*** $\mathbf{X}$ *is*

$$P_{\mathbf{X}}(\mathbf{x}) = P_{X_1,\ldots,X_n}(x_1,\ldots,x_n).$$

(c) The ***PDF of a continuous random vector*** $\mathbf{X}$ *is*

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,\ldots,X_n}(x_1,\ldots,x_n).$$

We use similar notation for a function $g(\mathbf{X}) = g(X_1,\ldots,X_n)$ of n random variables and a function $g(\mathbf{x}) = g(x_1,\ldots,x_n)$ of n numbers. Just as we described the relationship of two random variables in Chapter 5, we can explore a pair of random vectors by defining a joint probability model for vectors as a joint CDF, a joint PMF, or a joint PDF.

Definition 8.4 — Probability Functions of a Pair of Random Vectors

For random vectors $\mathbf{X}$ *with* n *components and* $\mathbf{Y}$ *with* m *components:*

(a) *The joint CDF of* $\mathbf{X}$ *and* $\mathbf{Y}$ *is*

$$F_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = F_{X_1,\ldots,X_n,Y_1,\ldots,Y_m}(x_1,\ldots,x_n,y_1,\ldots,y_m);$$

(b) *The joint PMF of discrete random vectors* $\mathbf{X}$ *and* $\mathbf{Y}$ *is*

$$P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = P_{X_1,\ldots,X_n,Y_1,\ldots,Y_m}(x_1,\ldots,x_n,y_1,\ldots,y_m);$$

(c) *The joint PDF of continuous random vectors* $\mathbf{X}$ *and* $\mathbf{Y}$ *is*

$$f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = f_{X_1,\ldots,X_n,Y_1,\ldots,Y_m}(x_1,\ldots,x_n,y_1,\ldots,y_m).$$

The logic of Definition 8.4 is that the pair of random vectors $\mathbf{X}$ and $\mathbf{Y}$ is the same as $\mathbf{W} = \begin{bmatrix}\mathbf{X}' & \mathbf{Y}'\end{bmatrix}' = \begin{bmatrix}X_1 & \cdots & X_n & Y_1 & \cdots & Y_m\end{bmatrix}'$, a concatenation of $\mathbf{X}$ and $\mathbf{Y}$. Thus a probability function of the pair $\mathbf{X}$ and $\mathbf{Y}$ corresponds to the same probability function of $\mathbf{W}$; for example, $F_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$ is the same CDF as $F_{\mathbf{W}}(\mathbf{w})$.

If we are interested only in $\mathbf{X} = X_1,\ldots,X_n$, we can use the methods introduced in Section 5.10 to derive a marginal probability model of $X_1,\ldots,X_n$ from the complete probability model for $X_1,\ldots,X_n,Y_1,\ldots,Y_m$. That is, if an experiment produces continuous random vectors $\mathbf{X}$ and $\mathbf{Y}$, then the joint vector PDF $f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y})$ is a complete probability model, while $f_{\mathbf{X}}(\mathbf{x})$ and $f_{\mathbf{Y}}(\mathbf{y})$ are marginal probability models for $\mathbf{X}$ and $\mathbf{Y}$.

Example 8.1

Random vector $\mathbf{X}$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 6e^{-\mathbf{a}'\mathbf{x}} & \mathbf{x} \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{8.1}$$

where $\mathbf{a} = \begin{bmatrix}1 & 2 & 3\end{bmatrix}'$. What is the CDF of $\mathbf{X}$?

Because $\mathbf{a}$ has three components, we infer that $\mathbf{X}$ is a three-dimensional random vector. Expanding $\mathbf{a}'\mathbf{x}$, we write the PDF as a function of the vector components,

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 6e^{-x_1-2x_2-3x_3} & x_i \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{8.2}$$

Applying Definition 8.4, we integrate the PDF with respect to the three variables to obtain

$$F_{\mathbf{X}}(\mathbf{x}) = \begin{cases} (1-e^{-x_1})(1-e^{-2x_2})(1-e^{-3x_3}) & x_i \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{8.3}$$

Quiz 8.1

Discrete random vectors $\mathbf{X} = [x_1 \; x_2 \; x_3]'$ and $\mathbf{Y} = [y_1 \; y_2 \; y_3]'$ are related by $\mathbf{Y} = \mathbf{A}\mathbf{X}$. Find the joint PMF $P_{\mathbf{Y}}(\mathbf{y})$ if $\mathbf{X}$ has joint PMF

$$P_{\mathbf{X}}(\mathbf{x}) = \begin{cases} (1-p)p^{x_3} & x_1 < x_2 < x_3; \\ & x_1, x_2, x_3 \in \{1, 2, \ldots\}, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

8.2 Independent Random Variables and Random Vectors

The probability model of the pair of independent random vectors $\mathbf{X}$ and $\mathbf{Y}$ is the product of the probability model of $\mathbf{X}$ and the probability model of $\mathbf{Y}$.

In considering the relationship of a pair of random vectors, we have the following definition of independence:

Definition 8.5 Independent Random Vectors

Random vectors $\mathbf{X}$ *and* $\mathbf{Y}$ *are independent if*

$$\textit{Discrete:} \quad P_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = P_{\mathbf{X}}(\mathbf{x})\, P_{\mathbf{Y}}(\mathbf{y});$$

$$\textit{Continuous:} \quad f_{\mathbf{X},\mathbf{Y}}(\mathbf{x},\mathbf{y}) = f_{\mathbf{X}}(\mathbf{x})\, f_{\mathbf{Y}}(\mathbf{y}).$$

Example 8.2

As in Example 5.23, random variables $Y_1, \ldots, Y_4$ have the joint PDF

$$f_{Y_1,\ldots,Y_4}(y_1, \ldots, y_4) = \begin{cases} 4 & 0 \le y_1 \le y_2 \le 1, 0 \le y_3 \le y_4 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8.4}$$

Let $\mathbf{V} = [Y_1 \;\; Y_4]'$ and $\mathbf{W} = [Y_2 \;\; Y_3]'$. Are $\mathbf{V}$ and $\mathbf{W}$ independent random vectors?

We first note that the components of $\mathbf{V}$ are $V_1 = Y_1$, and $V_2 = Y_4$. Also, $W_1 = Y_2$, and $W_2 = Y_3$. Therefore,

$$f_{\mathbf{V},\mathbf{W}}(\mathbf{v}, \mathbf{w}) = f_{Y_1,\ldots,Y_4}(v_1, w_1, w_2, v_2) = \begin{cases} 4 & 0 \le v_1 \le w_1 \le 1; \\ & 0 \le w_2 \le v_2 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8.5}$$

Since $\mathbf{V} = \begin{bmatrix} Y_1 & Y_4 \end{bmatrix}'$ and $\mathbf{W} = \begin{bmatrix} Y_2 & Y_3 \end{bmatrix}'$,

$$f_{\mathbf{V}}(\mathbf{v}) = f_{Y_1,Y_4}(v_1, v_2), \qquad f_{\mathbf{W}}(\mathbf{w}) = f_{Y_2,Y_3}(w_1, w_2). \tag{8.6}$$

In Example 5.23, we found $f_{Y_1,Y_4}(y_1, y_4)$ and $f_{Y_2,Y_3}(y_2, y_3)$ in Equations (5.78) and (5.80). From these marginal PDFs, we have

$$f_{\mathbf{V}}(\mathbf{v}) = \begin{cases} 4(1-v_1)v_2 & 0 \le v_1, v_2 \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{8.7}$$

$$f_{\mathbf{W}}(\mathbf{w}) = \begin{cases} 4w_1(1-w_2) & 0 \le w_1, w_2 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8.8}$$

Therefore,

$$f_{\mathbf{V}}(\mathbf{v})\, f_{\mathbf{W}}(\mathbf{w}) = \begin{cases} 16(1-v_1)v_2w_1(1-w_2) & 0 \le v_1, v_2, w_1, w_2 \le 1, \\ 0 & \text{otherwise,} \end{cases} \tag{8.9}$$

which is not equal to $f_{\mathbf{V},\mathbf{W}}(\mathbf{v}, \mathbf{w})$. Therefore $\mathbf{V}$ and $\mathbf{W}$ are not independent.

Quiz 8.2

Use the components of $\mathbf{Y} = \begin{bmatrix} Y_1, \ldots, Y_4 \end{bmatrix}'$ in Example 8.2 to construct two independent random vectors $\mathbf{V}$ and $\mathbf{W}$. Prove that $\mathbf{V}$ and $\mathbf{W}$ are independent.

8.3 Functions of Random Vectors

$P_W(w)$, the PMF of $W = g(\mathbf{X})$, a function of discrete random vector $\mathbf{X}$, is the sum of the probabilities of all sample vectors $\mathbf{x}$ for which $g(\mathbf{x}) = w$. To obtain the PDF of W, a function of a continuous random vector, we derive the CDF of W and then differentiate. The expected value of a function of a discrete random vector is the sum over the range of the random vector of the product of the function and the PMF. The expected value of a function of a continuous random vector is the integral over the range of the random vector of the product of the function and the PDF.

Just as we did for one random variable and two random variables, we can derive a random variable $W = g(\mathbf{X})$ that is a function of an arbitrary number of random variables. If W is discrete, the probability model can be calculated as $P_W(w)$, the probability of the event $A = \{W = w\}$ in Theorem 5.24. If W is continuous, the probability model can be expressed as $F_W(w) = \mathrm{P}[W \le w]$.

Theorem 8.1

For random variable $W = g(\mathbf{X})$,

$$\textit{Discrete:} \quad P_W(w) = \mathrm{P}[W = w] = \sum_{\mathbf{x}:g(\mathbf{x})=w} P_{\mathbf{X}}(\mathbf{x});$$

$$\textit{Continuous:} \quad F_W(w) = \mathrm{P}[W \le w] = \int_{g(\mathbf{x})\le w} \cdots \int f_{\mathbf{X}}(\mathbf{x})\, dx_1 \cdots dx_n.$$

Example 8.3

Consider an experiment that consists of spinning the pointer on the wheel of circumference 1 meter in Example 4.1 n times and observing Y_n meters, the maximum position of the pointer in the n spins. Find the CDF and PDF of Y_n.

If X_i is the position of the pointer on spin i, then $Y_n = \max\{X_1, X_2, \ldots, X_n\}$. As a result, $Y_n \le y$ if and only if each $X_i \le y$. This implies

$$F_{Y_n}(y) = \mathrm{P}[Y_n \le y] = \mathrm{P}[X_1 \le y, X_2 \le y, \ldots, X_n \le y]. \tag{8.10}$$

If we assume the spins to be independent, the events $\{X_1 \le y\}$, $\{X_2 \le y\}$, ..., $\{X_n \le y\}$ are independent events. Thus

$$F_{Y_n}(y) = \mathrm{P}[X_1 \le y] \cdots \mathrm{P}[X_n \le y] = (\mathrm{P}[X \le y])^n = (F_X(y))^n. \tag{8.11}$$

Example 4.2 derives Equation (4.8):

$$F_X(x) = \begin{cases} 0 & x < 0, \\ x & 0 \le x < 1, \\ 1 & x \ge 1. \end{cases} \tag{8.12}$$

Equations (8.11) and (8.12) imply that the CDF and corresponding PDF are

$$F_{Y_n}(y) = \begin{cases} 0 & y < 0, \\ y^n & 0 \le y \le 1, \\ 1 & y > 1, \end{cases} \qquad f_{Y_n}(y) = \begin{cases} ny^{n-1} & 0 \le y \le 1, \\ 0 & \text{otherwise}. \end{cases} \tag{8.13}$$

The following theorem is a generalization of Example 8.3. It expresses the PDF of the maximum and minimum values of a sequence of independent and identically distributed (iid) continuous random variables in terms of the CDF and PDF of the individual random variables.

Scanned with CamScanner

Theorem 8.2

Let $\mathbf{X}$ be a vector of n iid continuous random variables, each with CDF $F_X(x)$ and PDF $f_X(x)$.

(a) The CDF and the PDF of $Y = \max\{X_1, \ldots, X_n\}$ are

$$F_Y(y) = (F_X(y))^n, \qquad f_Y(y) = n(F_X(y))^{n-1} f_X(y).$$

(b) The CDF and the PDF of $W = \min\{X_1, \ldots, X_n\}$ are

$$F_W(w) = 1 - (1 - F_X(w))^n, \qquad f_W(w) = n(1 - F_X(w))^{n-1} f_X(w).$$

Proof By definition, $F_Y(y) = P[Y \le y]$. Because Y is the maximum value of $\{X_1, \ldots, X_n\}$, the event $\{Y \le y\} = \{X_1 \le y, X_2 \le y, \ldots, X_n \le y\}$. Because all the random variables X_i are iid, $\{Y \le y\}$ is the intersection of n independent events. Each of the events $\{X_i \le y\}$ has probability $F_X(y)$. The probability of the intersection is the product of the individual probabilities, which implies the first part of the theorem: $F_Y(y) = (F_X(y))^n$. The second part is the result of differentiating $F_Y(y)$ with respect to y. The derivations of $F_W(w)$ and $f_W(w)$ are similar. They begin with the observations that $F_W(w) = 1 - P[W > w]$ and that the event $\{W > w\} = \{X_1 > w, X_2 > w, \ldots X_n > w\}$, which is the intersection of n independent events, each with probability $1 - F_X(w)$.

In some applications of probability theory, we are interested only in the expected value of a function, not the complete probability model. Although we can always find $E[W]$ by first deriving $P_W(w)$ or $f_W(w)$, it is easier to find $E[W]$ by applying the following theorem.

Theorem 8.3

For a random vector $\mathbf{X}$, the random variable $g(\mathbf{X})$ has expected value

$$\textit{Discrete:} \quad E[g(\mathbf{X})] = \sum_{x_1 \in S_{X_1}} \cdots \sum_{x_n \in S_{X_n}} g(\mathbf{x}) P_{\mathbf{X}}(\mathbf{x});$$

$$\textit{Continuous:} \quad E[g(\mathbf{X})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})\, dx_1 \cdots dx_n.$$

If $W = g(\mathbf{X})$ is the product of n univariate functions and the components of $\mathbf{X}$ are mutually independent, $E[W]$ is a product of n expected values.

Theorem 8.4

When the components of $\mathbf{X}$ are independent random variables,

$$E[g_1(X_1)g_2(X_2)\cdots g_n(X_n)] = E[g_1(X_1)]E[g_2(X_2)]\cdots E[g_n(X_n)].$$

Proof When $\mathbf{X}$ is discrete, independence implies $P_{\mathbf{X}}(\mathbf{x}) = P_{X_1}(x_1)\cdots P_{X_n}(x_n)$. This implies

$$\mathrm{E}[g_1(X_1)\cdots g_n(X_n)] = \sum_{x_1\in S_{X_1}} \cdots \sum_{x_n\in S_{X_n}} g_1(x_1)\cdots g_n(x_n) P_{\mathbf{X}}(\mathbf{x}) \tag{8.14}$$

$$= \left(\sum_{x_1\in S_{X_1}} g_1(x_1)P_{X_1}(x_1)\right)\cdots\left(\sum_{x_n\in S_{X_n}} g_n(x_n)P_{X_n}(x_n)\right) \tag{8.15}$$

$$= \mathrm{E}[g_1(X_1)]\,\mathrm{E}[g_2(X_2)]\cdots\mathrm{E}[g_n(X_n)]. \tag{8.16}$$

The derivation is similar for independent continuous random variables.

We have considered the case of a single random variable $W = g(\mathbf{X})$ derived from a random vector $\mathbf{X}$. Some experiments may yield a new random vector $\mathbf{Y}$ with components $Y_1, \ldots, Y_n$ that are functions of the components of $\mathbf{X}$: $Y_k = g_k(\mathbf{X})$. We can derive the PDF of $\mathbf{Y}$ by first finding the CDF $F_{\mathbf{Y}}(\mathbf{y})$ and then applying Definition 5.11. The following theorem demonstrates this technique.

Theorem 8.5

Given the continuous random vector $\mathbf{X}$, define the derived random vector $\mathbf{Y}$ such that $Y_k = aX_k + b$ for constants $a > 0$ and b. The CDF and PDF of $\mathbf{Y}$ are

$$F_{\mathbf{Y}}(\mathbf{y}) = F_{\mathbf{X}}\left(\frac{y_1 - b}{a}, \ldots, \frac{y_n - b}{a}\right), \quad f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{a^n} f_{\mathbf{X}}\left(\frac{y_1 - b}{a}, \ldots, \frac{y_n - b}{a}\right).$$

Proof We observe $\mathbf{Y}$ has CDF $F_{\mathbf{Y}}(\mathbf{y}) = \mathrm{P}[aX_1 + b \le y_1, \ldots, aX_n + b \le y_n]$. Since $a > 0$,

$$F_{\mathbf{Y}}(\mathbf{y}) = \mathrm{P}\left[X_1 \le \frac{y_1 - b}{a}, \ldots, X_n \le \frac{y_n - b}{a}\right] = F_{\mathbf{X}}\left(\frac{y_1 - b}{a}, \ldots, \frac{y_n - b}{a}\right). \tag{8.17}$$

Definition 5.13 defines the joint PDF of $\mathbf{Y}$,

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{\partial^n F_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)}{\partial y_1 \cdots \partial y_n} = \frac{1}{a^n} f_{\mathbf{X}}\left(\frac{y_1 - b}{a}, \ldots, \frac{y_n - b}{a}\right). \tag{8.18}$$

Theorem 8.5 is a special case of a transformation of the form $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$. The following theorem is a consequence of the change-of-variable theorem (Appendix B, Math Fact B.13) in multivariable calculus.

Theorem 8.6

If $\mathbf{X}$ is a continuous random vector and $\mathbf{A}$ is an invertible matrix, then $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ has PDF

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|det(\mathbf{A})|} f_{\mathbf{X}}\left(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})\right)$$

Proof Let $B = \{\mathbf{y}|\mathbf{y} \le \hat{\mathbf{y}}\}$ so that $F_{\mathbf{Y}}(\hat{\mathbf{y}}) = \int_B f_{\mathbf{Y}}(\mathbf{y})\,d\mathbf{y}$. Define the vector transformation $\mathbf{x} = T(\mathbf{y}) = \mathbf{A}^{-1}(\mathbf{y}-\mathbf{b})$. It follows that $\mathbf{Y} \in B$ if and only if $\mathbf{X} \in T(B)$, where $T(B) = \{\mathbf{x}|\mathbf{A}\mathbf{x}+\mathbf{b} \le \hat{\mathbf{y}}\}$ is the image of B under transformation T. This implies

$$F_{\mathbf{Y}}(\hat{\mathbf{y}}) = P[\mathbf{X} \in T(B)] = \int_{T(B)} f_{\mathbf{X}}(\mathbf{x})\,d\mathbf{x} \tag{8.19}$$

By the change-of-variable theorem (Math Fact B.13),

$$F_{\mathbf{Y}}(\hat{\mathbf{y}}) = \int_B f_{\mathbf{X}}\left(\mathbf{A}^{-1}(\mathbf{y}-\mathbf{b})\right) \left|\det\left(\mathbf{A}^{-1}\right)\right| d\mathbf{y} \tag{8.20}$$

where $|\det(\mathbf{A}^{-1})|$ is the absolute value of the determinant of $\mathbf{A}^{-1}$. Definition 8.3 for the CDF and PDF of a random vector combined with Theorem 5.23(b) imply that $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{A}^{-1}(\mathbf{y}-\mathbf{b}))|\det(\mathbf{A}^{-1})|$. The theorem follows, since $|\det(\mathbf{A}^{-1})| = 1/|\det(\mathbf{A})|$.

Quiz 8.3

(A) A test of light bulbs produced by a machine has three possible outcomes: L, long life; A, average life; and R, reject. The results of different tests are independent. All tests have the following probability model: $P[L] = 0.3$, $P[A] = 0.6$, and $P[R] = 0.1$. Let X_1, X_2, and X_3 be the number of light bulbs that are L, A, and R respectively in five tests. Find the PMF $P_{\mathbf{X}}(\mathbf{x})$; the marginal PMFs $P_{X_1}(x_1)$, $P_{X_2}(x_2)$, and $P_{X_3}(x_3)$; and the PMF of $W = \max(X_1, X_2, X_3)$.

(B) The random vector $\mathbf{X}$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} e^{-x_3} & 0 \le x_1 \le x_2 \le x_3, \\ 0 & \text{otherwise.} \end{cases} \tag{8.21}$$

Find the PDF of $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$. where $\mathbf{A} = \text{diag}[2, 2, 2]$ and $\mathbf{b} = \begin{bmatrix} 4 & 4 & 4 \end{bmatrix}'$.

8.4 Expected Value Vector and Correlation Matrix

> The expected value of a random vector is a vector containing the expected values of the components of the vector. The covariance of a random vector is a symmetric matrix containing the variances of the components of the random vector and the covariances of all pairs of random variables in the random vector.

Corresponding to the expected value of a single random variable, the expected value of a random vector is a column vector in which the components are the expected values of the components of the random vector. There is a corresponding definition of the variance and standard deviation of a random vector.

Definition 8.6—Expected Value Vector

The expected value of a random vector $\mathbf{X}$ *is a column vector*

$$\mathrm{E}[\mathbf{X}] = \mu_{\mathbf{X}} = \begin{bmatrix} \mathrm{E}[X_1] & \mathrm{E}[X_2] & \cdots & \mathrm{E}[X_n] \end{bmatrix}'.$$

The correlation and covariance (Definition 5.7 and Definition 5.5) are numbers that contain important information about a pair of random variables. Corresponding information about random vectors is reflected in the set of correlations and the set of covariances of all pairs of components. These sets are referred to as *second-order statistics*. They have a concise matrix notation. To establish the notation, we first observe that for random vectors $\mathbf{X}$ with n components and $\mathbf{Y}$ with m components, the set of all products, X_iY_j, is contained in the $n \times m$ *random matrix* $\mathbf{XY}'$. If $\mathbf{Y} = \mathbf{X}$, the random matrix $\mathbf{XX}'$ contains all products, X_iX_j, of components of $\mathbf{X}$.

Example 8.4

If $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix}'$, what are the components of $\mathbf{XX}'$?

$$\mathbf{XX}' = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix} = \begin{bmatrix} X_1^2 & X_1X_2 & X_1X_3 \\ X_2X_1 & X_2^2 & X_2X_3 \\ X_3X_1 & X_3X_2 & X_3^2 \end{bmatrix}. \qquad (8.22)$$

In Definition 8.6, we defined the expected value of a random vector as the vector of expected values. This definition can be extended to random matrices.

Definition 8.7—Expected Value of a Random Matrix

For a random matrix $\mathbf{A}$ *with the random variable* A_{ij} *as its* i,j*th element,* $\mathrm{E}[\mathbf{A}]$ *is a matrix with* i,j*th element* $\mathrm{E}[A_{ij}]$.

Applying this definition to the random matrix $\mathbf{XX}'$, we have a concise way to define the correlation matrix of random vector $\mathbf{X}$.

Definition 8.8—Vector Correlation

The correlation of a random vector $\mathbf{X}$ *is an* $n \times n$ *matrix* $\mathbf{R_X}$ *with* i,j*th element* $R_{\mathbf{X}}(i,j) = \mathrm{E}[X_iX_j]$. *In vector notation,*

$$\mathbf{R_X} = \mathrm{E}[\mathbf{XX}'].$$

Example 8.5

If $\mathbf{X} = [X_1 \quad X_2 \quad X_3]'$, the correlation matrix of $\mathbf{X}$ is

$$\mathbf{R_X} = \begin{bmatrix} E[X_1^2] & E[X_1X_2] & E[X_1X_3] \\ E[X_2X_1] & E[X_2^2] & E[X_2X_3] \\ E[X_3X_1] & E[X_3X_2] & E[X_3^2] \end{bmatrix} = \begin{bmatrix} E[X_1^2] & r_{X_1,X_2} & r_{X_1,X_3} \\ r_{X_2,X_1} & E[X_2^2] & r_{X_2,X_3} \\ r_{X_3,X_1} & r_{X_3,X_2} & E[X_3^2] \end{bmatrix}.$$

The i,jth element of the correlation matrix is the expected value of the random variable X_iX_j. The *covariance matrix* of $\mathbf{X}$ is a similar generalization of the covariance of two random variables.

Definition 8.9 Vector Covariance

The covariance of a random vector $\mathbf{X}$ *is an* $n \times n$ *matrix* $\mathbf{C_X}$ *with components* $C_X(i,j) = \mathrm{Cov}[X_i, X_j]$. *In vector notation,*

$$\mathbf{C_X} = E[(\mathbf{X} - \mu_\mathbf{X})(\mathbf{X} - \mu_\mathbf{X})']$$

Example 8.6

If $\mathbf{X} = [X_1 \quad X_2 \quad X_3]'$, the covariance matrix of $\mathbf{X}$ is

$$\mathbf{C_X} = \begin{bmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1, X_2] & \mathrm{Cov}[X_1, X_3] \\ \mathrm{Cov}[X_2, X_1] & \mathrm{Var}[X_2] & \mathrm{Cov}[X_2, X_3] \\ \mathrm{Cov}[X_3, X_1] & \mathrm{Cov}[X_3, X_2] & \mathrm{Var}[X_3] \end{bmatrix} \tag{8.23}$$

Theorem 5.16(a), which connects the correlation and covariance of a pair of random variables, can be extended to random vectors.

Theorem 8.7

For a random vector $\mathbf{X}$ *with correlation matrix* $\mathbf{R_X}$, *covariance matrix* $\mathbf{C_X}$, *and vector expected value* $\mu_\mathbf{X}$,

$$\mathbf{C_X} = \mathbf{R_X} - \mu_\mathbf{X}\mu_\mathbf{X}'.$$

Proof The proof is essentially the same as the proof of Theorem 5.16(a), with vectors replacing scalars. Cross multiplying inside the expectation of Definition 8.9 yields

$$\begin{aligned} \mathbf{C_X} &= E[\mathbf{XX}' - \mathbf{X}\mu_\mathbf{X}' - \mu_\mathbf{X}\mathbf{X}' + \mu_\mathbf{X}\mu_\mathbf{X}'] \\ &= E[\mathbf{XX}'] - E[\mathbf{X}\mu_\mathbf{X}'] - E[\mu_\mathbf{X}\mathbf{X}'] + E[\mu_\mathbf{X}\mu_\mathbf{X}']. \end{aligned} \tag{8.24}$$

Since $E[\mathbf{X}] = \mu_\mathbf{X}$ is a constant vector,

$$\mathbf{C_X} = \mathbf{R_X} - E[\mathbf{X}]\mu'_\mathbf{X} - \mu_\mathbf{X} E[\mathbf{X}'] + \mu_\mathbf{X}\mu'_\mathbf{X} = \mathbf{R_X} - \mu_\mathbf{X}\mu'_\mathbf{X}. \quad (8.25)$$

Example 8.7

Find the expected value $E[\mathbf{X}]$, the correlation matrix $\mathbf{R_X}$, and the covariance matrix $\mathbf{C_X}$ of the two-dimensional random vector $\mathbf{X}$ with PDF

$$f_\mathbf{X}(\mathbf{x}) = \begin{cases} 2 & 0 \le x_1 \le x_2 \le 1, \\ 0 & \text{otherwise.} \end{cases} \quad (8.26)$$

The elements of the expected value vector are

$$E[X_i] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i f_\mathbf{X}(\mathbf{x})\, dx_1 dx_2 = \int_0^1\int_0^{x_2} 2x_i\, dx_1 dx_2, \quad i = 1, 2. \quad (8.27)$$

The integrals are $E[X_1] = 1/3$ and $E[X_2] = 2/3$, so that $\mu_\mathbf{X} = E[\mathbf{X}] = [1/3 \quad 2/3]'$. The elements of the correlation matrix are

$$E[X_1^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1^2 f_\mathbf{X}(\mathbf{x})\, dx_1 dx_2 = \int_0^1\int_0^{x_2} 2x_1^2\, dx_1 dx_2, \quad (8.28)$$

$$E[X_2^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2^2 f_\mathbf{X}(\mathbf{x})\, dx_1 dx_2 = \int_0^1\int_0^{x_2} 2x_2^2\, dx_1 dx_2, \quad (8.29)$$

$$E[X_1X_2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2 f_\mathbf{X}(\mathbf{x})\, dx_1 dx_2 = \int_0^1\int_0^{x_2} 2x_1x_2\, dx_1 dx_2. \quad (8.30)$$

These integrals are $E[X_1{}^2] = 1/6$, $E[X_2{}^2] = 1/2$, and $E[X_1X_2] = 1/4$. Therefore,

$$\mathbf{R_X} = \begin{bmatrix} 1/6 & 1/4 \\ 1/4 & 1/2 \end{bmatrix}. \quad (8.31)$$

We use Theorem 8.7 to find the elements of the covariance matrix.

$$\mathbf{C_X} = \mathbf{R_X} - \mu_\mathbf{X}\mu'_\mathbf{X} = \begin{bmatrix} 1/6 & 1/4 \\ 1/4 & 1/2 \end{bmatrix} - \begin{bmatrix} 1/9 & 2/9 \\ 2/9 & 4/9 \end{bmatrix} = \begin{bmatrix} 1/18 & 1/36 \\ 1/36 & 1/18 \end{bmatrix}. \quad (8.32)$$

In addition to the correlations and covariances of the elements of one random vector, it is useful to refer to the correlations and covariances of elements of two random vectors.

Scanned with CamScanner

Definition 8.10 — Vector Cross-Correlation

The cross-correlation of random vectors, $\mathbf{X}$ with n components and $\mathbf{Y}$ with m components, is an $n \times m$ matrix $\mathbf{R_{XY}}$ with i,jth element $R_{XY}(i,j) = E[X_iY_j]$, or, in vector notation,

$$\mathbf{R_{XY}} = E[\mathbf{XY'}].$$

Definition 8.11 — Vector Cross-Covariance

The cross-covariance of a pair of random vectors $\mathbf{X}$ with n components and $\mathbf{Y}$ with m components is an $n \times m$ matrix $\mathbf{C_{XY}}$ with i,jth element $C_{XY}(i,j) = \text{Cov}[X_i, Y_j]$, or, in vector notation,

$$\mathbf{C_{XY}} = E[(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})(\mathbf{Y} - \boldsymbol{\mu}_\mathbf{Y})'].$$

To distinguish the correlation or covariance of a random vector from the correlation or covariance of a pair of random vectors, we sometimes use the terminology *autocorrelation* and *autocovariance* when there is one random vector and *cross-correlation* and *cross-covariance* when there is a pair of random vectors. Note that when $\mathbf{X} = \mathbf{Y}$ the autocorrelation and cross-correlation are identical (as are the covariances). Recognizing this identity, some texts use the notation $\mathbf{R_{XX}}$ and $\mathbf{C_{XX}}$ for the correlation and covariance of a random vector.

When $\mathbf{Y}$ is a linear transformation of $\mathbf{X}$, the following theorem states the relationship of the second-order statistics of $\mathbf{Y}$ to the corresponding statistics of $\mathbf{X}$.

Theorem 8.8

$\mathbf{X}$ is an n-dimensional random vector with expected value $\boldsymbol{\mu}_\mathbf{X}$, correlation $\mathbf{R_X}$, and covariance $\mathbf{C_X}$. The m-dimensional random vector $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{b}$ is an m-dimensional vector, has expected value $\boldsymbol{\mu}_\mathbf{Y}$, correlation matrix $\mathbf{R_Y}$, and covariance matrix $\mathbf{C_Y}$ given by

$$\begin{aligned}
\boldsymbol{\mu}_\mathbf{Y} &= \mathbf{A}\boldsymbol{\mu}_\mathbf{X} + \mathbf{b}, \\
\mathbf{R_Y} &= \mathbf{AR_XA'} + (\mathbf{A}\boldsymbol{\mu}_\mathbf{X})\mathbf{b'} + \mathbf{b}(\mathbf{A}\boldsymbol{\mu}_\mathbf{X})' + \mathbf{bb'}, \\
\mathbf{C_Y} &= \mathbf{AC_XA'}.
\end{aligned}$$

Proof We derive the formulas for the expected value and covariance of $\mathbf{Y}$. The derivation for the correlation is similar. First, the expected value of $\mathbf{Y}$ is

$$\boldsymbol{\mu}_\mathbf{Y} = E[\mathbf{AX} + \mathbf{b}] = \mathbf{A}E[\mathbf{X}] + E[\mathbf{b}] = \mathbf{A}\boldsymbol{\mu}_\mathbf{X} + \mathbf{b}. \tag{8.33}$$

It follows that $\mathbf{Y} - \boldsymbol{\mu}_\mathbf{Y} = \mathbf{A}(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})$. This implies

$$\begin{aligned}
\mathbf{C_Y} &= E[(\mathbf{A}(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X}))(\mathbf{A}(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X}))'] \\
&= E[\mathbf{A}(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X}))(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})'\mathbf{A}'] = \mathbf{A}E[(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})(\mathbf{X} - \boldsymbol{\mu}_\mathbf{X})']\mathbf{A}' = \mathbf{AC_XA'}.
\end{aligned} \tag{8.34}$$

Example 8.8

Given the expected value μ_X, the correlation R_X, and the covariance C_X of random vector X in Example 8.7, and $Y = AX + b$, where

$$A = \begin{bmatrix} 1 & 0 \\ 6 & 3 \\ 3 & 6 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 0 \\ -2 \\ -2 \end{bmatrix}, \tag{8.35}$$

find the expected value μ_Y, the correlation R_Y, and the covariance C_Y.

From the matrix operations of Theorem 8.8, we obtain $\mu_Y = \begin{bmatrix} 1/3 & 2 & 3 \end{bmatrix}'$ and

$$R_Y = \begin{bmatrix} 1/6 & 13/12 & 4/3 \\ 13/12 & 7.5 & 9.25 \\ 4/3 & 9.25 & 12.5 \end{bmatrix}; \qquad C_Y = \begin{bmatrix} 1/18 & 5/12 & 1/3 \\ 5/12 & 3.5 & 3.25 \\ 1/3 & 3.25 & 3.5 \end{bmatrix}. \tag{8.36}$$

The cross-correlation and cross-covariance of two random vectors can be derived using algebra similar to the proof of Theorem 8.8.

Theorem 8.9

The vectors X and $Y = AX + b$ have cross-correlation R_{XY} and cross-covariance C_{XY} given by

$$R_{XY} = R_X A' + \mu_X b'; \qquad C_{XY} = C_X A'.$$

In the next example, we see that covariance and cross-covariance matrices allow us to quickly calculate the correlation coefficient between any pair of component random variables.

Example 8.9

Continuing Example 8.8 for random vectors X and $Y = AX + b$, calculate

(a) The cross-correlation matrix R_{XY} and the cross-covariance matrix C_{XY}.

(b) The correlation coefficients ρ_{Y_1,Y_3} and ρ_{X_2,Y_1}.

(a) Direct matrix calculation using Theorem 8.9 yields

$$R_{XY} = \begin{bmatrix} 1/6 & 13/12 & 4/3 \\ 1/4 & 5/3 & 29/12 \end{bmatrix}; \qquad C_{XY} = \begin{bmatrix} 1/18 & 5/12 & 1/3 \\ 1/36 & 1/3 & 5/12 \end{bmatrix}. \tag{8.37}$$

(b) Referring to Definition 5.6 and recognizing that $\text{Var}[Y_i] = C_Y(i,i)$, we have

$$\rho_{Y_1,Y_3} = \frac{\text{Cov}[Y_1, Y_3]}{\sqrt{\text{Var}[Y_1]\,\text{Var}[Y_3]}} = \frac{C_Y(1,3)}{\sqrt{C_Y(1,1)C_Y(3,3)}} = 0.756 \tag{8.38}$$

Similarly,

$$\rho_{X_2,Y_1} = \frac{\text{Cov}[X_2, Y_1]}{\sqrt{\text{Var}[X_2]\,\text{Var}[Y_1]}} = \frac{C_{XY}(2,1)}{\sqrt{C_X(2,2)C_Y(1,1)}} = 1/2. \tag{8.39}$$

Quiz 8.4

The three-dimensional random vector $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix}'$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 6 & 0 \le x_1 \le x_2 \le x_3 \le 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8.40}$$

Find $E[\mathbf{X}]$ and the correlation and covariance matrices $\mathbf{R_X}$ and $\mathbf{C_X}$.

8.5 Gaussian Random Vectors

The multivariate Gaussian PDF is a probability model for a vector in which all the components are Gaussian random variables. The parameters of the model are the expected value vector and the covariance matrix of the components. A linear function of a Gaussian random vector is also a Gaussian random vector. The components of the standard normal random vector are mutually independent standard normal random variables.

Multiple Gaussian random variables appear in many practical applications of probability theory. The *multivariate Gaussian distribution* is a probability model for n random variables with the property that the marginal PDFs are all Gaussian. A set of random variables described by the multivariate Gaussian PDF is said to be *jointly Gaussian*. A vector whose components are jointly Gaussian random variables is said to be a *Gaussian random vector*. The PDF of a Gaussian random vector has a particularly concise notation.

Definition 8.12 — Gaussian Random Vector

$\mathbf{X}$ is the Gaussian $(\mu_{\mathbf{X}}, \mathbf{C_X})$ random vector with expected value $\mu_{\mathbf{X}}$ and covariance $\mathbf{C_X}$ if and only if

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}[\det(\mathbf{C_X})]^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_{\mathbf{X}})'\mathbf{C_X}^{-1}(\mathbf{x} - \mu_{\mathbf{X}})\right)$$

where $\det(\mathbf{C_X})$, the determinant of $\mathbf{C_X}$, satisfies $\det(\mathbf{C_X}) > 0$.

Definition 8.12 is a generalization of Definition 4.8 and Definition 5.10. When $n = 1$, $\mathbf{C_X}$ and $\mathbf{x} - \boldsymbol{\mu}_\mathbf{X}$ are σ_X^2 and $x - \mu_X$, and the PDF in Definition 8.12 reduces to the ordinary Gaussian PDF of Definition 4.8. That is, a 1-dimensional Gaussian (μ, σ^2) random vector is a Gaussian (μ, σ) random variable.[1] In Problem 8.5.8, we ask you to show that for $n = 2$, Definition 8.12 reduces to the bivariate Gaussian PDF in Definition 5.10. The condition that $\det(\mathbf{C_X}) > 0$ is a generalization of the requirement for the bivariate Gaussian PDF that $|\rho_{X,Y}| < 1$. Basically, $\det(\mathbf{C_X}) > 0$ reflects the requirement that no random variable X_i is a linear combination of the other random variables in $\mathbf{X}$.

For a Gaussian random vector $\mathbf{X}$, an important special case is $\text{Cov}[X_i, X_j] = 0$ for all $i \neq j$. In the covariance matrix $\mathbf{C_X}$, the off-diagonal elements are all zero and the ith diagonal element is simply $\text{Var}[X_i] = \sigma_i^2$. In this case, we write $\mathbf{C_X} = \text{diag}[\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2]$. When the covariance matrix is diagonal, X_i and X_j are uncorrelated for $i \neq j$. In Theorem 5.20, we showed that uncorrelated bivariate Gaussian random variables are independent. The following theorem generalizes this result.

Theorem 8.10

A Gaussian random vector $\mathbf{X}$ *has independent components if and only if* $\mathbf{C_X}$ *is a diagonal matrix.*

Proof First, if the components of $\mathbf{X}$ are independent, then for $i \neq j$, X_i and X_j are independent. By Theorem 5.17(c), $\text{Cov}[X_i, X_j] = 0$. Hence the off-diagonal terms of $\mathbf{C_X}$ are all zero. If $\mathbf{C_X}$ is diagonal, then

$$\mathbf{C_X} = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix} \quad \text{and} \quad \mathbf{C_X}^{-1} = \begin{bmatrix} 1/\sigma_1^2 & & \\ & \ddots & \\ & & 1/\sigma_n^2 \end{bmatrix}. \tag{8.41}$$

It follows that $\mathbf{C_X}$ has determinant $\det(\mathbf{C_X}) = \prod_{i=1}^n \sigma_i^2$ and that

$$(\mathbf{x} - \boldsymbol{\mu}_\mathbf{X})' \mathbf{C_X}^{-1} (\mathbf{x} - \boldsymbol{\mu}_\mathbf{X}) = \sum_{i=1}^n \frac{(X_i - \mu_i)^2}{\sigma_i^2}. \tag{8.42}$$

From Definition 8.12, we see that

$$f_\mathbf{X}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \prod_{i=1}^n \sigma_i^2} \exp\left(-\sum_{i=1}^n (x_i - \mu_i)/2\sigma_i^2 \right) \tag{8.43}$$

$$= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-(x_i - \mu_i)^2 / 2\sigma_i^2 \right). \tag{8.44}$$

Thus $f_\mathbf{X}(\mathbf{x}) = \prod_{i=1}^n f_{X_i}(x_i)$, implying $X_1, \ldots, X_n$ are independent.

[1] For the Gaussian random variable, we specify parameters μ and σ because they have the same units. However, the PDF of the Gaussian random vector displays $\boldsymbol{\mu}_\mathbf{X}$ and $\mathbf{C_X}$ as parameters, and for one dimension $\mathbf{C_X} = \sigma_X^2$

Example 8.10

Consider the outdoor temperature at a certain weather station. On May 5, the temperature measurements in units of degrees Fahrenheit taken at 6 AM, 12 noon, and 6 PM are all Gaussian random variables, X_1, X_2, X_3, with variance 16 degrees^2. The expected values are 50 degrees, 62 degrees, and 58 degrees respectively. The covariance matrix of the three measurements is

$$\mathbf{C_X} = \begin{bmatrix} 16.0 & 12.8 & 11.2 \\ 12.8 & 16.0 & 12.8 \\ 11.2 & 12.8 & 16.0 \end{bmatrix}. \tag{8.45}$$

(a) Write the joint PDF of X_1, X_2 using the algebraic notation of Definition 5.10.
(b) Write the joint PDF of X_1, X_2 using vector notation.
(c) Write the joint PDF of $\mathbf{X} = \begin{bmatrix} X_1 & X_2 & X_3 \end{bmatrix}'$ using vector notation.

(a) First we note that X_1 and X_2 have expected values $\mu_1 = 50$ and $\mu_2 = 62$, variances $\sigma_1^2 = \sigma_2^2 = 16$, and covariance $\text{Cov}[X_1, X_2] = 12.8$. It follows from Definition 5.6 that the correlation coefficient is

$$\rho_{X_1,X_2} = \frac{\text{Cov}[X_1, X_2]}{\sigma_1\sigma_2} = \frac{12.8}{16} = 0.8. \tag{8.46}$$

From Definition 5.10, the joint PDF is

$$f_{X_1,X_2}(x_1, x_2) = \frac{\exp\left[-\frac{(x_1-50)^2 - 1.6(x_1-50)(x_2-62) + (x_2-62)^2}{19.2}\right]}{60.3}.$$

(b) Let $\mathbf{W} = \begin{bmatrix} X_1 & X_2 \end{bmatrix}'$ denote a vector representation for random variables X_1 and X_2. From the covariance matrix $\mathbf{C_X}$, we observe that the 2×2 submatrix in the upper left corner is the covariance matrix of the random vector $\mathbf{W}$. Thus

$$\mu_{\mathbf{W}} = \begin{bmatrix} 50 \\ 62 \end{bmatrix}, \qquad \mathbf{C_W} = \begin{bmatrix} 16.0 & 12.8 \\ 12.8 & 16.0 \end{bmatrix}. \tag{8.47}$$

We observe that $\det(\mathbf{C_W}) = 92.16$ and $\det(\mathbf{C_W})^{1/2} = 9.6$. From Definition 8.12, the joint PDF of $\mathbf{W}$ is

$$f_{\mathbf{W}}(\mathbf{w}) = \frac{1}{60.3} \exp\left(-\frac{1}{2}(\mathbf{w} - \mu_{\mathbf{W}})^T \mathbf{C_W^{-1}} (\mathbf{w} - \mu_{\mathbf{W}})\right). \tag{8.48}$$

(c) Since $\mu_{\mathbf{X}} = \begin{bmatrix} 50 & 62 & 58 \end{bmatrix}'$ and $\det(\mathbf{C_X})^{1/2} = 22.717$, $\mathbf{X}$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{357.8} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_{\mathbf{X}})^T \mathbf{C_X^{-1}} (\mathbf{x} - \mu_{\mathbf{X}})\right). \tag{8.49}$$

The following theorem is a generalization of Theorem 4.13. It states that a linear transformation of a Gaussian random vector results in another Gaussian random vector.

Theorem 8.11

Given an n-dimensional Gaussian random vector $\mathbf{X}$ *with expected value* $\boldsymbol{\mu}_\mathbf{X}$ *and covariance* $\mathbf{C}_\mathbf{X}$, *and an* $m \times n$ *matrix* $\mathbf{A}$ *with* $\operatorname{rank}(\mathbf{A}) = m$,

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

is an m-dimensional Gaussian random vector with expected value $\boldsymbol{\mu}_\mathbf{Y} = \mathbf{A}\boldsymbol{\mu}_\mathbf{X} + \mathbf{b}$ *and covariance* $\mathbf{C}_\mathbf{Y} = \mathbf{A}\mathbf{C}_\mathbf{X}\mathbf{A}'$.

Proof The proof of Theorem 8.8 contains the derivations of $\boldsymbol{\mu}_\mathbf{Y}$ and $\mathbf{C}_\mathbf{Y}$. Our proof that $\mathbf{Y}$ has a Gaussian PDF is confined to the special case when $m = n$ and $\mathbf{A}$ is an invertible matrix. The case of $m < n$ is addressed in Problem 8.5.14. When $m = n$, we use Theorem 8.6 to write

$$f_\mathbf{Y}(\mathbf{y}) = \frac{1}{|\det(\mathbf{A})|} f_\mathbf{X}(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})) \tag{8.50}$$

$$= \frac{\exp\left(-\frac{1}{2}[\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu}_\mathbf{X}]'\mathbf{C}_\mathbf{X}^{-1}[\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu}_\mathbf{X}]\right)}{(2\pi)^{n/2}|\det(\mathbf{A})|\,|\det(\mathbf{C}_\mathbf{X})|^{1/2}}. \tag{8.51}$$

In the exponent of $f_\mathbf{Y}(\mathbf{y})$, we observe that

$$\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - \boldsymbol{\mu}_\mathbf{X} = \mathbf{A}^{-1}[\mathbf{y} - (\mathbf{A}\boldsymbol{\mu}_\mathbf{X} + \mathbf{b})] = \mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu}_\mathbf{Y}), \tag{8.52}$$

since $\boldsymbol{\mu}_\mathbf{Y} = \mathbf{A}\boldsymbol{\mu}_\mathbf{X} + \mathbf{b}$. Applying (8.52) to (8.51) yields

$$f_\mathbf{Y}(\mathbf{y}) = \frac{\exp\left(-\frac{1}{2}[\mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu}_\mathbf{Y})]'\mathbf{C}_\mathbf{X}^{-1}[\mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu}_\mathbf{Y})]\right)}{(2\pi)^{n/2}|\det(\mathbf{A})|\,|\det(\mathbf{C}_\mathbf{X})|^{1/2}}. \tag{8.53}$$

Using the identities $|\det(\mathbf{A})||\det(\mathbf{C}_\mathbf{X})|^{1/2} = |\det(\mathbf{A}\mathbf{C}_\mathbf{X}\mathbf{A}')|^{1/2}$ and $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$, we can write

$$f_\mathbf{Y}(\mathbf{y}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_\mathbf{Y})'(\mathbf{A}')^{-1}\mathbf{C}_\mathbf{X}^{-1}\mathbf{A}^{-1}(\mathbf{y} - \boldsymbol{\mu}_\mathbf{Y})\right)}{(2\pi)^{n/2}|\det(\mathbf{A}\mathbf{C}_\mathbf{X}\mathbf{A}')|^{1/2}}. \tag{8.54}$$

Since $(\mathbf{A}')^{-1}\mathbf{C}_\mathbf{X}^{-1}\mathbf{A}^{-1} = (\mathbf{A}\mathbf{C}_\mathbf{X}\mathbf{A}')^{-1}$, we see from Equation (8.54) that $\mathbf{Y}$ is a Gaussian vector with expected value $\boldsymbol{\mu}_\mathbf{Y}$ and covariance matrix $\mathbf{C}_\mathbf{Y} = \mathbf{A}\mathbf{C}_\mathbf{X}\mathbf{A}'$.

Example 8.11

Continuing Example 8.10, use the formula $Y_i = (5/9)(X_i - 32)$ to convert the three temperature measurements to degrees Celsius.

(a) What is $\boldsymbol{\mu}_\mathbf{Y}$, the expected value of random vector $\mathbf{Y}$?

(b) What is $\mathbf{C}_\mathbf{Y}$, the covariance of random vector $\mathbf{Y}$?

(c) Write the joint PDF of $\mathbf{Y} = [Y_1 \quad Y_2 \quad Y_3]'$ using vector notation.

(a) In terms of matrices, we observe that $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ where

$$\mathbf{A} = \begin{bmatrix} 5/9 & 0 & 0 \\ 0 & 5/9 & 0 \\ 0 & 0 & 5/9 \end{bmatrix}, \qquad \mathbf{b} = -\frac{160}{9}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \tag{8.55}$$

(b) Since $\mu_{\mathbf{X}} = [50 \quad 62 \quad 58]'$, from Theorem 8.11,

$$\mu_{\mathbf{Y}} = \mathbf{A}\mu_{\mathbf{X}} + \mathbf{b} = \begin{bmatrix} 10 \\ 50/3 \\ 130/9 \end{bmatrix}. \tag{8.56}$$

(c) The covariance of $\mathbf{Y}$ is $\mathbf{C_Y} = \mathbf{A}\mathbf{C_X}\mathbf{A}'$. We note that $\mathbf{A} = \mathbf{A}' = (5/9)\mathbf{I}$ where $\mathbf{I}$ is the 3×3 identity matrix. Thus $\mathbf{C_Y} = (5/9)^2\mathbf{C_X}$ and $\mathbf{C_Y^{-1}} = (9/5)^2\mathbf{C_X^{-1}}$. The PDF of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{24.47}\exp\left(-\frac{81}{50}(\mathbf{y} - \mu_{\mathbf{Y}})^T\mathbf{C_X^{-1}}(\mathbf{y} - \mu_{\mathbf{Y}})\right). \tag{8.57}$$

A standard normal random vector is a generalization of the standard normal random variable in Definition 4.9.

Definition 8.13 Standard Normal Random Vector

The n-dimensional standard normal random vector $\mathbf{Z}$ *is the n-dimensional Gaussian random vector with* $E[\mathbf{Z}] = \mathbf{0}$ *and* $\mathbf{C_Z} = \mathbf{I}$.

From Definition 8.13, each component Z_i of $\mathbf{Z}$ has expected value $E[Z_i] = 0$ and variance $\text{Var}[Z_i] = 1$. Thus Z_i is the Gaussian (0,1) random variable. In addition, $E[Z_iZ_j] = 0$ for all $i \neq j$. Since $\mathbf{C_Z}$ is a diagonal matrix, $Z_1, \ldots, Z_n$ are independent.

In many situations, it is useful to transform the Gaussian (μ_X, σ_X) random variable X to the standard normal random variable $Z = (X - \mu_X)/\sigma_X$. For Gaussian vectors, we have a vector transformation to transform $\mathbf{X}$ into a standard normal random vector.

Theorem 8.12

For a Gaussian $(\mu_{\mathbf{X}}, \mathbf{C_X})$ *random vector, let* $\mathbf{A}$ *be an* $n \times n$ *matrix with the property* $\mathbf{A}\mathbf{A}' = \mathbf{C_X}$. *The random vector*

$$\mathbf{Z} = \mathbf{A}^{-1}(\mathbf{X} - \mu_{\mathbf{X}})$$

is a standard normal random vector.

Proof Applying Theorem 8.11 with $\mathbf{A}$ replaced by $\mathbf{A}^{-1}$, and $\mathbf{b} = \mathbf{A}^{-1}\mu_{\mathbf{X}}$, we have that $\mathbf{Z}$ is a Gaussian random vector with expected value

$$E[\mathbf{Z}] = E\left[\mathbf{A}^{-1}(\mathbf{X} - \mu_{\mathbf{X}})\right] = \mathbf{A}^{-1}E[\mathbf{X} - \mu_{\mathbf{X}}] = \mathbf{0} \tag{8.58}$$

and covariance

$$\mathbf{C_Z} = \mathbf{A}^{-1}\mathbf{C_X}(\mathbf{A}^{-1})' = \mathbf{A}^{-1}\mathbf{A}\mathbf{A}'(\mathbf{A}')^{-1} = \mathbf{I}. \tag{8.59}$$

The transformation in this theorem is considerably less straightforward than the scalar transformation $Z = (X - \mu_X)/\sigma_X$, because it is necessary to find for a given $\mathbf{C_X}$ a matrix $\mathbf{A}$ with the property $\mathbf{A}\mathbf{A}' = \mathbf{C_X}$. The calculation of $\mathbf{A}$ from $\mathbf{C_X}$ can be achieved by applying the linear algebra procedure *singular value decomposition*. Section 8.6 describes this procedure in more detail and applies it to generating sample values of Gaussian random vectors.

The inverse transform of Theorem 8.12 is particularly useful in computer simulations.

Theorem 8.13

Given the n-dimensional standard normal random vector $\mathbf{Z}$, an invertible $n \times n$ matrix $\mathbf{A}$, and an n-dimensional vector $\mathbf{b}$,

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{b}$$

is an n-dimensional Gaussian random vector with expected value $\mu_{\mathbf{X}} = \mathbf{b}$ and covariance matrix $\mathbf{C_X} = \mathbf{A}\mathbf{A}'$.

Proof By Theorem 8.11, $\mathbf{X}$ is a Gaussian random vector with expected value

$$\mu_{\mathbf{X}} = E[\mathbf{X}] = E[\mathbf{A}\mathbf{Z} + \mu_{\mathbf{X}}] = \mathbf{A}\,E[\mathbf{Z}] + \mathbf{b} = \mathbf{b}. \tag{8.60}$$

The covariance of $\mathbf{X}$ is

$$\mathbf{C_X} = \mathbf{A}\mathbf{C_Z}\mathbf{A}' = \mathbf{A}\mathbf{I}\mathbf{A}' = \mathbf{A}\mathbf{A}'. \tag{8.61}$$

Theorem 8.13 says that we can transform the standard normal vector $\mathbf{Z}$ into a Gaussian random vector $\mathbf{X}$ whose covariance matrix is of the form $\mathbf{C_X} = \mathbf{A}\mathbf{A}'$. The usefulness of Theorems 8.12 and 8.13 depends on whether we can always find a matrix $\mathbf{A}$ such that $\mathbf{C_X} = \mathbf{A}\mathbf{A}'$. In fact, as we verify below, this is possible for every Gaussian vector $\mathbf{X}$.

Theorem 8.14

For a Gaussian vector $\mathbf{X}$ with covariance $\mathbf{C_X}$, there always exists a matrix $\mathbf{A}$ such that $\mathbf{C_X} = \mathbf{A}\mathbf{A}'$.

Proof To verify this fact, we connect some simple facts:

- In Problem 8.4.12, we ask you to show that every random vector $\mathbf{X}$ has a positive semidefinite covariance matrix $\mathbf{C_X}$. By Math Fact B.17, every eigenvalue of $\mathbf{C_X}$ is nonnegative.
- The definition of the Gaussian vector PDF requires the existence of $\mathbf{C_X^{-1}}$. Hence, for a Gaussian vector $\mathbf{X}$, all eigenvalues of $\mathbf{C_X}$ are nonzero. From the previous step, we observe that all eigenvalues of $\mathbf{C_X}$ must be positive.
- Since $\mathbf{C_X}$ is a real symmetric matrix, Math Fact B.15 says it has a singular value decomposition (SVD) $\mathbf{C_X} = \mathbf{UDU'}$ where $\mathbf{D} = \text{diag}[d_1, \ldots, d_n]$ is the diagonal matrix of eigenvalues of $\mathbf{C_X}$. Since each d_i is positive, we can define $\mathbf{D}^{1/2} = \text{diag}[\sqrt{d_1}, \ldots, \sqrt{d_n}]$, and we can write

$$\mathbf{C_X} = \mathbf{UD}^{1/2}\mathbf{D}^{1/2}\mathbf{U'} = \left(\mathbf{UD}^{1/2}\right)\left(\mathbf{UD}^{1/2}\right)'. \tag{8.62}$$

We see that $\mathbf{A} = \mathbf{UD}^{1/2}$.

From Theorems 8.12, 8.13, and 8.14, it follows that any Gaussian $(\boldsymbol{\mu}_\mathbf{X}, \mathbf{C_X})$ random vector $\mathbf{X}$ can be written as a linear transformation of uncorrelated Gaussian (0, 1) random variables. In terms of the SVD $\mathbf{C_X} = \mathbf{UDU'}$ and the standard normal vector $\mathbf{Z}$, the transformation is

$$\mathbf{X} = \mathbf{UD}^{1/2}\mathbf{Z} + \boldsymbol{\mu}_\mathbf{X}. \tag{8.63}$$

We recall that $\mathbf{U}$ has orthonormal columns $\mathbf{u}_1, \ldots, \mathbf{u}_n$. When $\boldsymbol{\mu}_\mathbf{X} = \mathbf{0}$, Equation (8.63) can be written as

$$\mathbf{X} = \sum_{i=1}^{n} \sqrt{d_i}\mathbf{u}_i Z_i. \tag{8.64}$$

The interpretation of Equation (8.64) is that a Gaussian random vector $\mathbf{X}$ is a combination of orthogonal vectors $\sqrt{d_i}\mathbf{u}_i$, each scaled by an independent Gaussian random variable Z_i. In a wide variety of problems involving Gaussian random vectors, the transformation from the Gaussian vector $\mathbf{X}$ to the standard normal random vector $\mathbf{Z}$ is the key to an efficient solution. Also, we will see in the next section that Theorem 8.13 is essential in using MATLAB to generate arbitrary Gaussian random vectors.

Quiz 8.5

$\mathbf{Z}$ is the two-dimensional standard normal random vector. The Gaussian random vector $\mathbf{X}$ has components

$$X_1 = 2Z_1 + Z_2 + 2 \quad \text{and} \quad X_2 = Z_1 - Z_2. \tag{8.65}$$

Calculate the expected value vector $\boldsymbol{\mu}_\mathbf{X}$ and the covariance matrix $\mathbf{C_X}$.

8.6 MATLAB

MATLAB is especially useful for random vectors. We use a sample space grid to calculate properties of a probability model of a discrete random vector. We use the functions `randn` and `svd` to generate samples of Gaussian random vectors.

As in Section 5.11, we demonstrate two ways of using MATLAB to study random vectors. We first present examples of programs that calculate values of probability functions, in this case the PMF of a discrete random vector and the PDF of a Gaussian random vector. Then we present a program that generates sample values of the Gaussian (μ_X, C_X) random vector given any μ_X and C_X.

Probability Functions

The MATLAB approach of using a sample space grid, presented in Section 5.11, can also be applied to finite random vectors X described by a PMF $P_X(x)$.

Example 8.12

Finite random vector $X = \begin{bmatrix} X_1 & X_2, \cdots & X_5 \end{bmatrix}'$ has PMF

$$P_X(x) = \begin{cases} k\sqrt{x'x} & x_i \in \{-10, -9, \ldots, 10\}; \\ & i = 1, 2, \ldots, 5, \\ 0 & \text{otherwise}. \end{cases} \tag{8.66}$$

What is the constant k? Find the expected value and standard deviation of X_3.

Summing $P_X(x)$ over all possible values of x is the sort of tedious task that MATLAB handles easily. Here are the code and corresponding output:

```
%x5.m
sx=-10:10;
[SX1,SX2,SX3,SX4,SX5]...
    =ndgrid(sx,sx,sx,sx,sx);
P=sqrt(SX1.^2 +SX2.^2+SX3.^2+SX4.^2+SX5.^2);
k=1.0/(sum(sum(sum(sum(sum(P))))))
P=k*P;
EX3=sum(sum(sum(sum(sum(P.*SX3)))))
EX32=sum(sum(sum(sum(sum(P.*(SX3.^2))))));
sigma3=sqrt(EX32-(EX3)^2)
```

```
>> x5
k =
  1.8491e-008
EX3 =
  -3.2960e-017
sigma3 =
     6.3047
>>
```

In fact, by symmetry arguments, it should be clear that $E[X_3] = 0$. In adding 11^5 terms, MATLAB's finite precision led to a small error on the order of 10^{-17}.

Example 8.12 demonstrates the use of MATLAB to calculate properties of a probability model by performing lots of straightforward calculations. For a continuous random vector X, MATLAB could be used to calculate $E[g(X)]$ using Theorem 8.3 and numeric integration. One step in such a calculation is computing values of the PDF. The next example performs this function for any Gaussian (μ_X, C_X) random vector.

Example 8.13

Write a MATLAB function `f=gaussvectorpdf(mu,C,x)` that calculates $f_{\mathbf{X}}(\mathbf{x})$ for a Gaussian (μ, C) random vector.

```
function f=gaussvectorpdf(mu,C,x)
n=length(x);
z=x(:)-mu(:);
f=exp(-z'*inv(C)*z)/...
    sqrt((2*pi)^n*det(C));
```

`gaussvectorpdf` computes the Gaussian PDF $f_{\mathbf{X}}(\mathbf{x})$ of Definition 8.12. Of course, MATLAB makes the calculation simple by providing operators for matrix inverses and determinants.

Sample Values of Gaussian Random Vectors

Gaussian random vectors appear in a wide variety of experiments. Here we present a program that uses the built-in MATLAB function `randn` to generate sample values of Gaussian ($\mu_{\mathbf{X}}, \mathbf{C}_{\mathbf{X}}$) random vectors. The matrix notation lends itself to concise MATLAB coding. Our approach is based on Theorem 8.13. In particular, we generate a standard normal random vector $\mathbf{Z}$ and, given a covariance matrix $\mathbf{C}$, we use built-in MATLAB functions to calculate a matrix $\mathbf{A}$ such that $\mathbf{C} = \mathbf{A}\mathbf{A}'$. By Theorem 8.13, $\mathbf{X} = \mathbf{A}\mathbf{Z} + \mu_{\mathbf{X}}$ is a Gaussian ($\mu_{\mathbf{X}}, \mathbf{C}$) vector. Although the MATLAB code for this task will be quite short, it needs some explanation:

- `x=randn(m,n)` produces an $m \times n$ matrix, with each matrix element a Gaussian (0, 1) random variable. Thus each column of `x` is a sample vector of standard normal vector $\mathbf{Z}$.
- `[U,D,V]=svd(C)` is the singular value decomposition (SVD) of matrix `C`. In math notation, given $\mathbf{C}$, `svd` produces a diagonal matrix $\mathbf{D}$ of the same dimension as $\mathbf{C}$ and with nonnegative diagonal elements in decreasing order, and unitary matrices $\mathbf{U}$ and $\mathbf{V}$ so that $\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{V}'$. Singular value decomposition is a powerful technique that can be applied to any matrix. When $\mathbf{C}$ is a covariance matrix, the singular value decomposition yields $\mathbf{U} = \mathbf{V}$ and $\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{U}'$. Just as in the proof of Theorem 8.14, $\mathbf{A} = \mathbf{U}\mathbf{D}^{1/2}$.

```
function x=gaussvector(mu,C,m)
[U,D,V]=svd(C);
x=V*(D^(0.5))*randn(n,m)...
    +(mu(:)*ones(1,m));
```

Using MATLAB functions `randn` and `svd`, generating Gaussian random vectors is easy. The function `x=gaussvector(mu,C,1)` produces a Gaussian (`mu`, `C`) random vector.

The general form `gaussvector(mu,C,m)` produces an $n \times m$ matrix where each of the m columns is a Gaussian random vector with expected value `mu` and covariance `C`. The reason for defining `gaussvector` to return m vectors at the same time is that calculating the singular value decomposition is a computationally burdensome step. Instead, we perform the SVD just once, rather than m times.

Quiz 8.6

The daily noon temperature, measured in degrees Fahrenheit, in New Jersey in July can be modeled as a Gaussian random vector $\mathbf{T} = \begin{bmatrix} T_1 & \cdots & T_{31} \end{bmatrix}'$ where T_i

is the temperature on the ith day of the month. Suppose that $E[T_i] = 80$ for all i, and that T_i and T_j have covariance

$$\text{Cov}[T_i, T_j] = \frac{36}{1+|i-j|} \tag{8.67}$$

Define the daily average temperature as

$$Y = \frac{T_1 + T_2 + \cdots + T_{31}}{31}. \tag{8.68}$$

Based on this model, write a MATLAB program `p=julytemps(T)` that calculates $P[Y \geq T]$, the probability that the daily average temperature is at least T degrees.

Further Reading: [WS01] and [PP02] make extensive use of vectors and matrices. To go deeply into vector random variables, students can use [Str98] to gain a firm grasp of principles of linear algebra.

Problems

Difficulty: ● Easy ■ Moderate ♦ Difficult ♦♦ Experts Only

8.1.1● For random variables $X_1, \ldots, X_n$ in Problem 5.10.3, let $\mathbf{X} = [X_1 \ \cdots \ X_n]'$. What is $f_{\mathbf{X}}(\mathbf{x})$?

8.1.2● Random vector $\mathbf{X}$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} c\mathbf{a}'\mathbf{x} & \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathbf{a} = [a_1 \ \cdots \ a_n]'$ is a vector with each component $a_i > 0$. What is c?

8.1.3■ Given $f_{\mathbf{X}}(\mathbf{x})$ with $c = 2/3$ and $a_1 = a_2 = a_3 = 1$ in Problem 8.1.2, find the marginal PDF $f_{X_3}(x_3)$.

8.1.4■ $\mathbf{X} = [X_1 \ X_2 \ X_3]'$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 6 & 0 \leq x_1 \leq x_2 \leq x_3 \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathbf{U} = [X_1 \ X_2]'$, $\mathbf{V} = [X_1 \ X_3]'$ and $\mathbf{W} = [X_2 \ X_3]'$. Find the marginal PDFs $f_{\mathbf{U}}(\mathbf{u})$, $f_{\mathbf{V}}(\mathbf{v})$ and $f_{\mathbf{W}}(\mathbf{w})$.

8.1.5■ A wireless data terminal has three messages waiting for transmission. After sending a message, it expects an acknowledgement from the receiver. When it receives the acknowledgement, it transmits the next message. If the acknowledgement does not arrive, it sends the message again. The probability of successful transmission of a message is p independent of other transmissions. Let $\mathbf{K} = [K_1 \ K_2 \ K_3]'$ be the three-dimensional random vector in which K_i is the total number of transmissions when message i is received successfully. (K_3 is the total number of transmissions used to send all three messages.) Show that

$$P_{\mathbf{K}}(\mathbf{k}) = \begin{cases} p^3(1-p)^{k_3-3} & k_1 < k_2 < k_3; \\ & k_i \in \{1, 2\ldots\}, \\ 0 & \text{otherwise.} \end{cases}$$

8.1.6■ From the joint PMF $P_{\mathbf{K}}(\mathbf{k})$ in Problem 8.1.5, find the marginal PMFs

(a) $P_{K_1,K_2}(k_1, k_2)$,

(b) $P_{K_1,K_3}(k_1, k_3)$,

(c) $P_{K_2,K_3}(k_2, k_3)$,

(d) $P_{K_1}(k_1)$, $P_{K_2}(k_2)$, and $P_{K_3}(k_3)$.

8.1.7■ Let $\mathbf{N}$ be the r-dimensional random vector with the multinomial PMF given in

Example 5.21 with $n > r > 2$:

$$P_{\mathbf{N}}(\mathbf{n}) = \binom{n}{n_1, \ldots, n_r} p_1^{n_1} \cdots p_r^{n_r}.$$

(a) What is the joint PMF of N_1 and N_2? Hint: Consider a new classification scheme with categories: s_1, s_2, and "other."

(b) Let $T_i = N_1 + \cdots + N_i$. What is the PMF of T_i?

(c) What is the joint PMF of T_1 and T_2?

8.1.8■ The random variables $Y_1, \ldots, Y_4$ have the joint PDF

$$f_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} 24 & 0 \le y_1 \le y_2 \le y_3 \le y_4 \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the marginal PDFs $f_{Y_1,Y_4}(y_1, y_4)$, $f_{Y_1,Y_2}(y_1, y_2)$, and $f_{Y_1}(y_1)$.

8.1.9♦ As a generalization of the message transmission system in Problem 8.1.5, consider a terminal that has n messages to transmit. The components k_i of the n-dimensional random vector $\mathbf{K}$ are the total number of messages transmitted when message i is received successfully.

(a) Find the PMF of $\mathbf{K}$.

(b) For each $j \in \{1, 2, \ldots, n-1\}$, find the marginal PMF $P_{K_1,\ldots,K_j}(k_1, \ldots, k_j)$.

(c) For each $i \in \{1, 2, \ldots, n\}$, find the marginal PMF $P_{K_i}(k_i)$.

Hint: These PMFs are members of a family of discrete random variables in Appendix A.

8.2.1● The n components X_i of random vector $\mathbf{X}$ have $E[X_i] = 0$ $\text{Var}[X_i] = \sigma^2$. What is the covariance matrix $\mathbf{C_X}$?

8.2.2● The 4-dimensional random vector $\mathbf{X}$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 1 & 0 \le x_i \le 1, i = 1, 2, 3, 4, \\ 0 & \text{otherwise.} \end{cases}$$

Are the four components of $\mathbf{X}$ independent random variables?

8.2.3● As in Example 8.1, the random vector $\mathbf{X}$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 6e^{-\mathbf{a}'\mathbf{x}} & \mathbf{x} \ge \mathbf{0}, \\ 0 & \text{otherwise.} \end{cases}$$

where $\mathbf{a} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}'$. Are the components of $\mathbf{X}$ independent random variables?

8.2.4● The PDF of the 3-dimensional random vector $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} e^{-x_3} & 0 \le x_1 \le x_2 \le x_3, \\ 0 & \text{otherwise.} \end{cases}$$

Are the components of $\mathbf{X}$ independent random variables?

8.2.5■ The random vector $\mathbf{X}$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} e^{-x_3} & 0 \le x_1 \le x_2 \le x_3, \\ 0 & \text{otherwise.} \end{cases}$$

Find the marginal PDFs $f_{X_1}(x_1)$, $f_{X_2}(x_2)$, and $f_{X_3}(x_3)$.

8.3.1● Discrete random vector $\mathbf{X}$ has PMF $P_{\mathbf{X}}(\mathbf{x})$. Prove that for an invertible matrix $\mathbf{A}$, $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ has PMF

$$P_{\mathbf{Y}}(\mathbf{y}) = P_{\mathbf{X}}\left(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})\right).$$

8.3.2■ In the message transmission problem, Problem 8.1.5, the PMF for the number of transmissions when message i is received successfully is

$$P_{\mathbf{K}}(\mathbf{k}) = \begin{cases} p^3(1-p)^{k_3-3} & k_1 < k_2 < k_3; \\ & k_i \in \{1, 2 \ldots\}, \\ 0 & \text{otherwise.} \end{cases}$$

Let $J_3 = K_3 - K_2$, the number of transmissions of message 3; $J_2 = K_2 - K_1$, the number of transmissions of message 2; and $J_1 = K_1$, the number of transmissions of message 1. Derive a formula for $P_{\mathbf{J}}(\mathbf{j})$, the PMF of the number of transmissions of individual messages.

8.3.3■ In an automatic geolocation system, a dispatcher sends a message to six trucks in a fleet asking their locations. The waiting times for responses from the six trucks are iid exponential random variables, each with expected value 2 seconds.

(a) What is the probability that all six responses will arrive within 5 seconds?

(b) If the system has to locate all six vehicles within 3 seconds, it has to reduce the expected response time of each vehicle. What is the maximum expected response time that will produce a location time for all six vehicles of 3 seconds or less with probability of at least 0.9?

8.3.4◆◆ Let $X_1, \ldots, X_n$ denote n iid random variables with PDF $f_X(x)$ and CDF $F_X(x)$. What is the probability $P[X_n = \max\{X_1, \ldots, X_n\}]$?

8.4.1● Random variables X_1 and X_2 have zero expected value and variances $\text{Var}[X_1] = 4$ and $\text{Var}[X_2] = 9$. Their covariance is $\text{Cov}[X_1, X_2] = 3$.

(a) Find the covariance matrix of $\mathbf{X} = [X_1 \quad X_2]'$.

(b) Find the covariance matrix of $\mathbf{Y} = [Y_1 \quad Y_2]'$ given by

$$Y_1 = X_1 - 2X_2,$$
$$Y_2 = 3X_1 + 4X_2.$$

8.4.2● Let $X_1, \ldots, X_n$ be iid random variables with expected value 0, variance 1, and covariance $\text{Cov}[X_i, X_j] = \rho$, for $i \neq j$. Use Theorem 8.8 to find the expected value and variance of the sum $Y = X_1 + \cdots + X_n$.

8.4.3● The two-dimensional random vector $\mathbf{X}$ and the three-dimensional random vector $\mathbf{Y}$ are independent and $E[\mathbf{Y}] = \mathbf{0}$. What is the vector cross-correlation $\mathbf{R}_{\mathbf{XY}}$?

8.4.4● The four-dimensional random vector $\mathbf{X}$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 1 & 0 \leq x_i \leq 1, i = 1, 2, 3, 4 \\ 0 & \text{otherwise.} \end{cases}$$

Find the expected value vector $E[\mathbf{X}]$, the correlation matrix $\mathbf{R}_{\mathbf{X}}$, and the covariance matrix $\mathbf{C}_{\mathbf{X}}$.

8.4.5● The random vector $\mathbf{Y} = [Y_1 \quad Y_2]'$ has covariance matrix $\mathbf{C}_{\mathbf{Y}} = \left[\begin{smallmatrix} 25 & \gamma \\ \gamma & 4 \end{smallmatrix}\right]$ where γ is a constant. In terms of γ, what is the correlation coefficient ρ_{Y_1,Y_2} of Y_1 and Y_2? For what values of γ is $\mathbf{C}_{\mathbf{Y}}$ a valid covariance matrix?

8.4.6● In the message transmission system in Problem 8.1.5, the solution to Problem 8.3.2 is a formula for the PMF of $\mathbf{J}$, the number of transmissions of individual messages. For $p = 0.8$, find the expected value vector $E[\mathbf{J}]$, the correlation matrix $\mathbf{R}_{\mathbf{J}}$, and the covariance matrix $\mathbf{C}_{\mathbf{J}}$.

8.4.7■ In the message transmission system in Problem 8.1.5,

$$P_{\mathbf{K}}(\mathbf{k}) = \begin{cases} p^3(1-p)^{k_3-3}; & k_1 < k_2 < k_3; \\ & k_i \in \{1, 2, \ldots\}, \\ 0 & \text{otherwise.} \end{cases}$$

For $p = 0.8$, find the expected value vector $E[\mathbf{K}]$, the covariance matrix $\mathbf{C}_{\mathbf{K}}$, and the correlation matrix $\mathbf{R}_{\mathbf{K}}$.

8.4.8■ Random vector $\mathbf{X} = [X_1 \quad X_2]'$ has PDF

$$f_{\mathbf{X}}(\mathbf{x}) = \begin{cases} 10e^{-5x_1 - 2x_2} & x_1 \geq 0, x_2 \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$.

(b) Derive the expected value vector $\mu_{\mathbf{X}}$ and covariance matrix $\mathbf{C}_{\mathbf{X}}$.

(c) Let $\mathbf{Z} = \mathbf{AX}$, where $\mathbf{A} = \left(\begin{smallmatrix} -1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Find the covariance matrix of $\mathbf{Z}$.

8.4.9■ As in Quiz 5.10 and Example 5.23, the 4-dimensional random vector $\mathbf{Y}$ has PDF

$$f_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} 4 & 0 \leq y_1 \leq y_2 \leq 1; \\ & 0 \leq y_3 \leq y_4 \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the expected value vector $E[\mathbf{Y}]$, the correlation matrix $\mathbf{R}_{\mathbf{Y}}$, and the covariance matrix $\mathbf{C}_{\mathbf{Y}}$.

8.4.10■ $\mathbf{X} = [X_1 \quad X_2]'$ is a random vector with $E[\mathbf{X}] = [0 \quad 0]'$ and covariance matrix

$$\mathbf{C_X} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

For some ω satisfying $0 \leq \omega \leq 1$, let $Y = \sqrt{\omega}X_1 + \sqrt{1-\omega}X_2$. What value (or values) of ω will maximize $E[Y^2]$?

8.4.11■ The two-dimensional random vector $\mathbf{Y}$ has PDF

$$f_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} 2 & \mathbf{y} \geq \mathbf{0}, \begin{bmatrix} 1 & 1 \end{bmatrix} \mathbf{y} \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Find the expected value vector $E[\mathbf{Y}]$, the correlation matrix $\mathbf{R_Y}$, and the covariance matrix $\mathbf{C_Y}$.

8.4.12◆ Let $\mathbf{X}$ be a random vector with correlation matrix $\mathbf{R_X}$ and covariance matrix $\mathbf{C_X}$. Show that $\mathbf{R_X}$ and $\mathbf{C_X}$ are both positive semidefinite by showing that for any nonzero vector $\mathbf{a}$,

$$\mathbf{a}'\mathbf{R_X}\mathbf{a} \geq 0,$$
$$\mathbf{a}'\mathbf{C_X}\mathbf{a} \geq 0.$$

8.5.1● $\mathbf{X}$ is the 3-dimensional Gaussian random vector with expected value $\boldsymbol{\mu}_\mathbf{X} = [4 \quad 8 \quad 6]'$ and covariance

$$\mathbf{C_X} = \begin{bmatrix} 4 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 4 \end{bmatrix}.$$

Calculate

(a) the correlation matrix, $\mathbf{R_X}$,

(b) the PDF of the first two components of $\mathbf{X}$, $f_{X_1,X_2}(x_1, x_2)$,

(c) the probability that $X_1 > 8$.

8.5.2● $\mathbf{X} = [X_1 \quad X_2]'$ is the Gaussian random vector with $E[\mathbf{X}] = [0 \quad 0]'$ and covariance matrix

$$\mathbf{C_X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

What is the PDF of $Y = [2 \quad 1]\mathbf{X}$?

8.5.3● Given the Gaussian random vector $\mathbf{X}$ in Problem 8.5.1, $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 1/2 & 2/3 \\ 1 & -1/2 & 2/3 \end{bmatrix}$$

and $\mathbf{b} = [-4 \quad -4]'$. Calculate

(a) the expected value $\boldsymbol{\mu}_\mathbf{Y}$,

(b) the covariance $\mathbf{C_Y}$,

(c) the correlation $\mathbf{R_Y}$,

(d) the probability that $-1 \leq Y_2 \leq 1$.

8.5.4● Let $\mathbf{X}$ be a Gaussian $(\boldsymbol{\mu}_\mathbf{X}, \mathbf{C_X})$ random vector. Given a vector $\mathbf{a}$, find the expected value and variance of $Y = \mathbf{a}'\mathbf{X}$. Is Y a Gaussian random variable?

8.5.5■ Random variables X_1 and X_2 have zero expected value. The random vector $\mathbf{X} = [X_1 \quad X_2]'$ has a covariance matrix of the form

$$\mathbf{C} = \begin{bmatrix} 1 & \alpha \\ \beta & 4 \end{bmatrix}.$$

(a) For what values of α and β is $\mathbf{C}$ a valid covariance matrix?

(b) For what values of α and β can $\mathbf{X}$ be a Gaussian random vector?

(c) Suppose now that α and β satisfy the conditions in part (b) and $\mathbf{X}$ is a Gaussian random vector. What is the PDF of X_2? What is the PDF of $W = 2X_1 - X_2$?

8.5.6■ The Gaussian random vector $\mathbf{X} = [X_1 \quad X_2]'$ has expected value $E[\mathbf{X}] = \mathbf{0}$ and covariance matrix $\mathbf{C}_X = \left[\begin{smallmatrix} \sigma_1^2 & 1 \\ 1 & \sigma_2^2 \end{smallmatrix}\right]$.

(a) Under what conditions on σ_1^2 and σ_2^2 is $\mathbf{C}_X$ a valid covariance matrix?

(b) Suppose $\mathbf{Y} = [Y_1 \quad Y_2]' = \mathbf{AX}$ where $\mathbf{A} = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. For what values (if any) of σ_1^2 and σ_2^2 are the components Y_1 and Y_2 independent?

8.5.7■ The Gaussian random vector $\mathbf{X} = [X_1 \quad X_2]'$ has expected value $E[\mathbf{X}] = \mathbf{0}$ and covariance matrix $\mathbf{C}_X = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 1 \end{smallmatrix}\right]$.

(a) Find the PDF of $W = X_1 + 2X_2$.

(b) Find the PDF $f_{\mathbf{Y}}(\mathbf{y})$ of $\mathbf{Y} = \mathbf{AX}$ where $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.

8.5.8■ Let $\mathbf{X}$ be a Gaussian random vector with expected value $[\mu_1 \quad \mu_2]'$ and covariance matrix

$$\mathbf{C_X} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Show that $\mathbf{X}$ has bivariate Gaussian PDF $f_{\mathbf{X}}(\mathbf{x}) = f_{X_1,X_2}(x_1, x_2)$ given by Definition 5.10.

8.5.9■ $X = [X_1 \quad X_2]'$ is a Gaussian random vector with $E[X] = [0 \quad 0]'$ and covariance matrix $\mathbf{C_X} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

(a) What conditions must a, b, c, and d satisfy?

(b) Under what conditions (in addition to those in part (a)) are X_1 and X_2 independent?

(c) Under what conditions (in addition to those in part (a)) are X_1 and X_2 identical?

8.5.10■ Let $\mathbf{X}$ be a Gaussian $(\boldsymbol{\mu}_{\mathbf{X}}, \mathbf{C_X})$ random vector. Let $\mathbf{Y} = \mathbf{AX}$ where $\mathbf{A}$ is an $m \times n$ matrix of rank m. By Theorem 8.11, $\mathbf{Y}$ is a Gaussian random vector. Is

$$\mathbf{W} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$$

a Gaussian random vector?

8.5.11■ The 2×2 matrix

$$\mathbf{Q} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

is called a rotation matrix because $\mathbf{y} = \mathbf{Qx}$ is the rotation of $\mathbf{x}$ by the angle θ. Suppose $\mathbf{X} = [X_1 \quad X_2]'$ is a Gaussian $(\mathbf{0}, \mathbf{C_X})$ vector where $\mathbf{C_X} = \text{diag}[\sigma_1^2, \sigma_2^2]$ and $\sigma_2^2 \geq \sigma_1^2$. Let $\mathbf{Y} = \mathbf{QX}$.

(a) Find the covariance of Y_1 and Y_2. Show that Y_1 and Y_2 are independent for all θ if $\sigma_1^2 = \sigma_2^2$.

(b) Suppose $\sigma_2^2 > \sigma_1^2$. For what values θ are Y_1 and Y_2 independent?

8.5.12♦ $\mathbf{X} = [X_1 \quad X_2]'$ is a Gaussian $(\mathbf{0}, \mathbf{C_X})$ vector where

$$\mathbf{C_X} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Thus, depending on the value of the correlation coefficient ρ, the joint PDF of X_1 and X_2 may resemble one of the graphs of Figure 5.6 with $X_1 = X$ and $X_2 = Y$. Show that $\mathbf{X} = \mathbf{QY}$, where $\mathbf{Q}$ is the $\theta = 45°$ rotation matrix (see Problem 8.5.11) and $\mathbf{Y}$ is a Gaussian $(\mathbf{0}, \mathbf{C_Y})$ vector such that

$$\mathbf{C_Y} = \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix}.$$

This result verifies, for $\rho \neq 0$, that the PDF of X_1 and X_2 shown in Figure 5.6 is the joint PDF of two independent Gaussian random variables (with variances $1+\rho$ and $1-\rho$) rotated by 45°.

8.5.13♦ An n-dimensional Gaussian vector $\mathbf{W}$ has a block diagonal covariance matrix

$$\mathbf{C_W} = \begin{bmatrix} \mathbf{C_X} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_Y} \end{bmatrix},$$

where $\mathbf{C_X}$ is $m \times m$, $\mathbf{C_Y}$ is $(n-m) \times (n-m)$. Show that $\mathbf{W}$ can be written in terms of component vectors $\mathbf{X}$ and $\mathbf{Y}$ in the form

$$\mathbf{W} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix},$$

such that $\mathbf{X}$ and $\mathbf{Y}$ are independent Gaussian random vectors.

8.5.14♦♦ In this problem, we extend the proof of Theorem 8.11 to the case when $\mathbf{A}$ is $m \times n$ with $m < n$. For this proof, we assume $\mathbf{X}$ is an n-dimensional Gaussian vector and that we have proved Theorem 8.11 for the case $m = n$. Since the case $m = n$ is sufficient to prove that $\mathbf{Y} = \mathbf{X} + \mathbf{b}$ is Gaussian, it is sufficient to show for $m < n$ that $\mathbf{Y} = \mathbf{AX}$ is Gaussian in the case when $\boldsymbol{\mu}_{\mathbf{X}} = \mathbf{0}$.

(a) Prove there exists an $(n-m) \times n$ matrix $\tilde{\mathbf{A}}$ of rank $n - m$ with the property that $\tilde{\mathbf{A}}\mathbf{A}' = \mathbf{0}$. Hint: Review the Gram-Schmidt procedure.

(b) Let $\hat{\mathbf{A}} = \hat{\mathbf{A}}\mathbf{C}_{\mathbf{X}}^{-1}$ and define the random vector

$$\bar{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y} \\ \hat{\mathbf{Y}} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \hat{\mathbf{A}} \end{bmatrix} \mathbf{X}.$$

Use Theorem 8.11 for the case $m = n$ to argue that $\bar{\mathbf{Y}}$ is a Gaussian random vector.

(c) Find the covariance matrix $\bar{\mathbf{C}}$ of $\bar{\mathbf{Y}}$. Use the result of Problem 8.5.13 to show that $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ are independent Gaussian random vectors.

8.6.1• Consider the vector $\mathbf{X}$ in Problem 8.5.1 and define $Y = (X_1+X_2+X_3)/3$. What is the probability that $Y > 4$?

8.6.2■ A better model for the sailboat race of Problem 5.10.8 accounts for the fact that all boats are subject to the same randomness of wind and tide. Suppose in the race of ten sailboats, the finishing times X_i are identical Gaussian random variables with expected value 35 minutes and standard deviation 5 minutes. However, for every pair of boats i and j, the finish times X_i and X_j have correlation coefficient $\rho = 0.8$.

(a) What is the covariance matrix of $\mathbf{X} = [X_1 \ \cdots \ X_{10}]'$?

(b) Let

$$Y = \frac{X_1 + X_2 + \cdots + X_{10}}{10}.$$

What are the expected value and variance of Y? What is $P[Y \leq 25]$?

8.6.3■ For the vector of daily temperatures $[T_1 \ \cdots \ T_{31}]'$ and average temperature Y modeled in Quiz 8.6, we wish to estimate the probability of the event

$$A = \left\{Y \leq 82, \min_i T_i \geq 72\right\}.$$

To form an estimate of A, generate 10,000 independent samples of the vector $\mathbf{T}$ and calculate the relative frequency of A in those trials.

8.6.4♦ We continue Problem 8.6.2 where the vector $\mathbf{X}$ of finish times has correlated components. Let W denote the finish time of the winning boat. We wish to estimate $P[W \leq 25]$, the probability that the winning boat finishes in under 25 minutes. To do this, simulate $m = 10{,}000$ races by generating m samples of the vector $\mathbf{X}$ of finish times. Let $Y_j = 1$ if the winning time in race i is under 25 minutes; otherwise, $Y_j = 0$. Calculate the estimate

$$P[W \leq 25] \approx \frac{1}{m} \sum_{j=1}^{m} Y_j.$$

8.6.5♦♦ Write a MATLAB program that simulates m runs of the weekly lottery of Problem 7.5.9. For $m = 1000$ sample runs, form a histogram for the jackpot J.

9

Sums of Random Variables

Random variables of the form

$$W_n = X_1 + \cdots + X_n \tag{9.1}$$

appear repeatedly in probability theory and applications. We could in principle derive the probability model of W_n from the PMF or PDF of $X_1, \ldots, X_n$. However, in many practical applications, the nature of the analysis or the properties of the random variables allow us to apply techniques that are simpler than analyzing a general n-dimensional probability model. In Section 9.1 we consider applications in which our interest is confined to expected values related to W_n, rather than a complete model of W_n. Subsequent sections emphasize techniques that apply when $X_1, \ldots, X_n$ are mutually independent. A useful way to analyze the sum of independent random variables is to transform the PDF or PMF of each random variable to a *moment generating function*.

The central limit theorem reveals a fascinating property of the sum of independent random variables. It states that the CDF of the sum converges to a Gaussian CDF as the number of terms grows without limit. This theorem allows us to use the properties of Gaussian random variables to obtain accurate estimates of probabilities associated with sums of other random variables. In many cases exact calculation of these probabilities is extremely difficult.

9.1 Expected Values of Sums

> The expected value of a sum of *any* random variables is the sum of the expected values. The variance of the sum of any random variable is the sum of all the covariances. The variance of the sum of *independent* random variables is the sum of the variances.

The theorems of Section 5.7 can be generalized in a straightforward manner to describe expected values and variances of sums of more than two random variables.

Theorem 9.1

For any set of random variables $X_1, \ldots, X_n$, the sum $W_n = X_1 + \cdots + X_n$ has expected value

$$E[W_n] = E[X_1] + E[X_2] + \cdots + E[X_n].$$

Proof We prove this theorem by induction on n. In Theorem 5.11, we proved $E[W_2] = E[X_1] + E[X_2]$. Now we assume $E[W_{n-1}] = E[X_1] + \cdots + E[X_{n-1}]$. Notice that $W_n = W_{n-1} + X_n$. Since W_n is a sum of the two random variables W_{n-1} and X_n, we know that $E[W_n] = E[W_{n-1}] + E[X_n] = E[X_1] + \cdots + E[X_{n-1}] + E[X_n]$.

Keep in mind that the expected value of the sum equals the sum of the expected values whether or not $X_1, \ldots, X_n$ are independent. For the variance of W_n, we have the generalization of Theorem 5.12:

Theorem 9.2

The variance of $W_n = X_1 + \cdots + X_n$ is

$$\text{Var}[W_n] = \sum_{i=1}^{n} \text{Var}[X_i] + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{Cov}[X_i, X_j].$$

Proof From the definition of the variance, we can write $\text{Var}[W_n] = E[(W_n - E[W_n])^2]$. For convenience, let μ_i denote $E[X_i]$. Since $W_n = \sum_{i=1}^{n} X_n$ and $E[W_n] = \sum_{i=1}^{n} \mu_i$, we can write

$$\text{Var}[W_n] = E\left[\left(\sum_{i=1}^{n}(X_i - \mu_i)\right)^2\right] = E\left[\sum_{i=1}^{n}(X_i - \mu_i)\sum_{j=1}^{n}(X_j - \mu_j)\right] \tag{9.2}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}[X_i, X_j]. \tag{9.3}$$

In terms of the random vector $\mathbf{X} = [X_1 \ \cdots \ X_n]'$, we see that $\text{Var}[W_n]$ is the sum of all the elements of the covariance matrix $\mathbf{C_X}$. Recognizing that $\text{Cov}[X_i, X_i] = \text{Var}[X]$ and $\text{Cov}[X_i, X_j] = \text{Cov}[X_j, X_i]$, we place the diagonal terms of $\mathbf{C_X}$ in one sum and the off-diagonal terms (which occur in pairs) in another sum to arrive at the formula in the theorem.

When $X_1, \ldots, X_n$ are uncorrelated, $\text{Cov}[X_i, X_j] = 0$ for $i \neq j$ and the variance of the sum is the sum of the variances.

Theorem 9.3

When $X_1, \ldots, X_n$ are uncorrelated,

$$\mathrm{Var}[W_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n].$$

Example 9.1

$X_0, X_1, X_2, \ldots$ is a sequence of random variables with expected values $E[X_i] = 0$ and covariances, $\mathrm{Cov}[X_i, X_j] = 0.8^{|i-j|}$. Find the expected value and variance of a random variable Y_i defined as the sum of three consecutive values of the random sequence

$$Y_i = X_i + X_{i-1} + X_{i-2}. \tag{9.4}$$

Theorem 9.1 implies that

$$E[Y_i] = E[X_i] + E[X_{i-1}] + E[X_{i-2}] = 0. \tag{9.5}$$

Applying Theorem 9.2, we obtain for each i,

$$\begin{aligned}\mathrm{Var}[Y_i] = {} & \mathrm{Var}[X_i] + \mathrm{Var}[X_{i-1}] + \mathrm{Var}[X_{i-2}] \\ & + 2\mathrm{Cov}[X_i, X_{i-1}] + 2\mathrm{Cov}[X_i, X_{i-2}] + 2\mathrm{Cov}[X_{i-1}, X_{i-2}].\end{aligned} \tag{9.6}$$

We next note that $\mathrm{Var}[X_i] = \mathrm{Cov}[X_i, X_i] = 0.8^{i-i} = 1$ and that

$$\mathrm{Cov}[X_i, X_{i-1}] = \mathrm{Cov}[X_{i-1}, X_{i-2}] = 0.8^1, \qquad \mathrm{Cov}[X_i, X_{i-2}] = 0.8^2. \tag{9.7}$$

Therefore,

$$\mathrm{Var}[Y_i] = 3 \times 0.8^0 + 4 \times 0.8^1 + 2 \times 0.8^2 = 7.48. \tag{9.8}$$

The following example shows how a puzzling problem can be formulated as a question about the sum of a set of dependent random variables.

Example 9.2

At a party of $n \geq 2$ people, each person throws a hat in a common box. The box is shaken and each person blindly draws a hat from the box without replacement. We say a match occurs if a person draws his own hat. What are the expected value and variance of V_n, the number of matches?

Let X_i denote an indicator random variable such that

$$X_i = \begin{cases} 1 & \text{person } i \text{ draws his hat,} \\ 0 & \text{otherwise.} \end{cases} \tag{9.9}$$

The number of matches is $V_n = X_1 + \cdots + X_n$. Note that the X_i are generally not independent. For example, with $n = 2$ people, if the first person draws his own hat, then the second person must also draw her own hat. Note that the ith person is equally likely to draw any of the n hats, thus $P_{X_i}(1) = 1/n$ and $E[X_i] = P_{X_i}(1) = 1/n$. Since the expected value of the sum always equals the sum of the expected values,

$$E[V_n] = E[X_1] + \cdots + E[X_n] = n(1/n) = 1. \tag{9.10}$$

To find the variance of V_n, we will use Theorem 9.2. The variance of X_i is

$$\mathrm{Var}[X_i] = E\left[X_i^2\right] - (E[X_i])^2 = \frac{1}{n} - \frac{1}{n^2}. \tag{9.11}$$

To find $\mathrm{Cov}[X_i, X_j]$, we observe that

$$\mathrm{Cov}[X_i, X_j] = E[X_i X_j] - E[X_i]E[X_j]. \tag{9.12}$$

Note that $X_i X_j = 1$ if and only if $X_i = 1$ and $X_j = 1$, and $X_i X_j = 0$ otherwise. Thus

$$E[X_i X_j] = P_{X_i,X_j}(1,1) = P_{X_i|X_j}(1|1)\, P_{X_j}(1). \tag{9.13}$$

Given $X_j = 1$, that is, the jth person drew his own hat, then $X_i = 1$ if and only if the ith person draws his own hat from the $n-1$ other hats. Hence $P_{X_i|X_j}(1|1) = 1/(n-1)$ and

$$E[X_i X_j] = \frac{1}{n(n-1)}, \qquad \mathrm{Cov}[X_i, X_j] = \frac{1}{n(n-1)} - \frac{1}{n^2}. \tag{9.14}$$

Finally, we can use Theorem 9.2 to calculate

$$\mathrm{Var}[V_n] = n\,\mathrm{Var}[X_i] + n(n-1)\mathrm{Cov}[X_i, X_j] = 1. \tag{9.15}$$

That is, both the expected value and variance of V_n are 1, no matter how large n is!

Example 9.3

Continuing Example 9.2, suppose each person immediately returns to the box the hat that he or she drew. What is the expected value and variance of V_n, the number of matches?

In this case the indicator random variables X_i are independent and identically distributed (iid) because each person draws from the same bin containing all n hats. The number of matches $V_n = X_1 + \cdots + X_n$ is the sum of n iid random variables. As before, the expected value of V_n is

$$E[V_n] = n\,E[X_i] = 1. \tag{9.16}$$

In this case, the variance of V_n equals the sum of the variances,

$$\mathrm{Var}[V_n] = n\,\mathrm{Var}[X_i] = n\left(\frac{1}{n} - \frac{1}{n^2}\right) = 1 - \frac{1}{n}. \tag{9.17}$$

The remainder of this chapter examines tools for analyzing complete probability models of sums of random variables, with the emphasis on sums of independent random variables.

Quiz 9.1

Let W_n denote the sum of n independent throws of a fair four-sided die. Find the expected value and variance of W_n.

9.2 Moment Generating Functions

$\phi_X(s)$, the moment generating function (MGF) of a random variable X, is a probability model of X. If X is discrete, the MGF is a transform of the PMF. The MGF of a continuous random variable is a transform of the PDF, similar to a Laplace transform. The n-th moment of X is the n-th derivative of $\phi_X(s)$ evaluated at $s = 0$.

In Section 6.5, we learned in Theorem 6.9 that the PDF of the sum $W_2 = X_1 + X_2$ of independent random variables can be written as the convolution $f_{W_2}(w_2) = \int_{-\infty}^{\infty} f_{X_1}(w_2 - x_2) f_{X_2}(x_2)\, dx_2$. To find the PDF a sum of three independent random variables, $W_3 = X_1 + X_2 + X_3$, we could use Theorem 6.9 to find the PDF of $W_2 = X_1 + X_2$. and then, because $W_3 = W_2 + X_3$ and W_2 and X_3 are independent, we could use Theorem 6.9 again to find the PDF of W_3 from the convolution $f_{W_3}(w_3) = \int_{-\infty}^{\infty} f_{W_2}(w_3 - x_3) f_{X_3}(x_3)\, dx_3$. In principle, we could continue this sequence of convolutions to find the PDF of $W_n = X_1 + \cdots + X_n$ for any n. While this procedure is sound in theory, convolutional integrals are generally tricky, and a sequence of n convolutions is often prohibitively difficult to evaluate by hand. Even MATLAB typically fails to simplify the evaluation of a sequence of convolutions.

In linear system theory, however, convolution in the time domain corresponds to multiplication in the frequency domain with time functions and frequency functions related by the Fourier transform. In probability theory, we can, in a similar way, use transform methods to replace the convolution of PDFs by multiplication of transforms. In the language of probability theory, the transform of a PDF or a PMF is a *moment generating function*.

Definition 9.1 Moment Generating Function (MGF)

For a random variable X, the moment generating function (MGF) of X is

$$\phi_X(s) = \mathrm{E}\left[e^{sX}\right].$$

Definition 9.1 applies to both discrete and continuous random variables X. What changes in going from discrete X to continuous X is the method of calculating the expected value. When X is a continuous random variable,

$$\phi_X(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx. \tag{9.18}$$

For a discrete random variable Y, the MGF is

$$\phi_Y(s) = \sum_{y_i \in S_Y} e^{sy_i} P_Y(y_i). \tag{9.19}$$

Equation (9.18) indicates that the MGF of a continuous random variable is similar to the Laplace transform of a time function. The primary difference between an MGF and a Laplace transform is that the MGF is defined only for real values of s. For a given random variable X, there is a range of possible values of s for which $\phi_X(s)$ exists. The set of values of s for which $\phi_X(s)$ exists is called the *region of convergence*. The definition of the MGF implies that $\phi_X(0) = E[e^0] = 1$. Thus $s = 0$ is always in the region of convergence. If X is a nonnegative random variable, the region of convergence includes all $s \leq 0$. If X is bounded so that $P[a < X \leq b] = 1$, then $\phi_X(s)$ exists for all real s. Typically, the region of convergence is an interval around the $s = 0$.

Because the MGF and PMF or PDF form a transform pair, the MGF is also a complete probability model of a random variable. Given the MGF, it is possible to compute the PDF or PMF. Moreover, the derivatives of $\phi_X(s)$ evaluated at $s = 0$ are the moments of X.

Theorem 9.4

A random variable X with MGF $\phi_X(s)$ has nth moment

$$E[X^n] = \left.\frac{d^n \phi_X(s)}{ds^n}\right|_{s=0}.$$

Proof The first derivative of $\phi_X(s)$ is

$$\frac{d\phi_X(s)}{ds} = \frac{d}{ds}\left(\int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx\right) = \int_{-\infty}^{\infty} x e^{sx} f_X(x)\, dx. \tag{9.20}$$

Evaluating this derivative at $s = 0$ proves the theorem for $n = 1$.

$$\left.\frac{d\phi_X(s)}{ds}\right|_{s=0} = \int_{-\infty}^{\infty} x f_X(x)\, dx = E[X]. \tag{9.21}$$

Similarly, the nth derivative of $\phi_X(s)$ is

$$\frac{d^n \phi_X(s)}{ds^n} = \int_{-\infty}^{\infty} x^n e^{sx} f_X(x)\, dx. \tag{9.22}$$

The integral evaluated at $s = 0$ is the formula in the theorem statement.

Typically it is easier to calculate the moments of X by finding the MGF and

Random Variable	PMF or PDF	MGF $\phi_X(s)$
Bernoulli (p)	$P_X(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & \text{otherwise} \end{cases}$	$1-p+pe^s$
Binomial (n,p)	$P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$	$(1-p+pe^s)^n$
Geometric (p)	$P_X(x) = \begin{cases} p(1-p)^{x-1} & x=1,2,\ldots \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{pe^s}{1-(1-p)e^s}$
Pascal (k,p)	$P_X(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$	$\left(\dfrac{pe^s}{1-(1-p)e^s}\right)^k$
Poisson (α)	$P_X(x) = \begin{cases} \alpha^x e^{-\alpha}/x! & x=0,1,2,\ldots \\ 0 & \text{otherwise} \end{cases}$	$e^{\alpha(e^s-1)}$
Disc. Uniform (k,l)	$P_X(x) = \begin{cases} \frac{1}{l-k+1} & x=k,\ldots,l \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{sk}-e^{s(l+1)}}{1-e^s}$
Constant (a)	$f_X(x) = \delta(x-a)$	e^{sa}
Uniform (a,b)	$f_X(x) = \begin{cases} \frac{1}{b-a} & a<x<b \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{e^{bs}-e^{as}}{s(b-a)}$
Exponential (λ)	$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{\lambda}{\lambda-s}$
Erlang (n,λ)	$f_X(x) = \begin{cases} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$	$\left(\dfrac{\lambda}{\lambda-s}\right)^n$
Gaussian (μ,σ)	$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$	$e^{s\mu+s^2\sigma^2/2}$

Table 9.1 Moment generating function for families of random variables.

differentiating than by integrating $x^n f_X(x)$.

Example 9.4

X is an exponential random variable with MGF $\phi_X(s) = \lambda/(\lambda - s)$. What are the first and second moments of X? Write a general expression for the nth moment.

The first moment is the expected value:

$$E[X] = \left.\frac{d\phi_X(s)}{ds}\right|_{s=0} = \left.\frac{\lambda}{(\lambda-s)^2}\right|_{s=0} = \frac{1}{\lambda}. \tag{9.23}$$

The second moment of X is the mean square value:

$$E[X^2] = \left.\frac{d^2\phi_X(s)}{ds^2}\right|_{s=0} = \left.\frac{2\lambda}{(\lambda-s)^3}\right|_{s=0} = \frac{2}{\lambda^2}. \tag{9.24}$$

Proceeding in this way, it should become apparent that the nth moment of X is

$$E[X^n] = \left.\frac{d^n\phi_X(s)}{ds^n}\right|_{s=0} = \left.\frac{n!\lambda}{(\lambda-s)^{n+1}}\right|_{s=0} = \frac{n!}{\lambda^n}. \tag{9.25}$$

Table 9.1 presents the MGF for the families of random variables defined in Chapters 3 and 4. The following theorem derives the MGF of a linear transformation of a random variable X in terms of $\phi_X(s)$.

Theorem 9.5

The MGF of $Y = aX + b$ is $\phi_Y(s) = e^{sb}\phi_X(as)$.

Proof From the definition of the MGF,

$$\phi_Y(s) = E\left[e^{s(aX+b)}\right] = e^{sb} E\left[e^{(as)X}\right] = e^{sb}\phi_X(as). \tag{9.26}$$

Quiz 9.2

Random variable K has PMF

$$P_K(k) = \begin{cases} 0.2 & k = 0, \ldots, 4, \\ 0 & \text{otherwise.} \end{cases} \tag{9.27}$$

Use $\phi_K(s)$ to find the first, second, third, and fourth moments of K.

9.3 MGF of the Sum of Independent Random Variables

Moment generating functions provide a convenient way to determine the probability model of a sum of iid random variables. Using MGFs, we determine that when $W = X_1 + \cdots + X_n$ is a sum of n iid random variables:

- If X_i is Bernoulli (p), W is binomial (n, p).
- If X_i is Poisson (α), W is Poisson $(n\alpha)$.
- If X_i is geometric (p), W is Pascal (n, p).
- If X_i is exponential (λ), W is Erlang (n, λ).
- If X_i is Gaussian (μ, σ), W is Gaussian $(n\mu, \sqrt{n}\sigma)$.

Moment generating functions are particularly useful for analyzing sums of independent random variables, because if X and Y are independent, the MGF of $W = X+Y$ is the product

$$\phi_W(s) = \mathrm{E}\left[e^{sX}e^{sY}\right] = \mathrm{E}\left[e^{sX}\right]\mathrm{E}\left[e^{sY}\right] = \phi_X(s)\phi_Y(s). \tag{9.28}$$

Theorem 9.6 generalizes this result to a sum of n independent random variables.

Theorem 9.6

For a set of independent random variables $X_1, \ldots, X_n$, the moment generating function of $W = X_1 + \cdots + X_n$ is

$$\phi_W(s) = \phi_{X_1}(s)\phi_{X_2}(s)\cdots\phi_{X_n}(s).$$

When $X_1, \ldots, X_n$ are iid, each with MGF $\phi_{X_i}(s) = \phi_X(s)$,

$$\phi_W(s) = [\phi_X(s)]^n.$$

Proof From the definition of the MGF,

$$\phi_W(s) = \mathrm{E}\left[e^{s(X_1+\cdots+X_n)}\right] = \mathrm{E}\left[e^{sX_1}e^{sX_2}\cdots e^{sX_n}\right]. \tag{9.29}$$

Here, we have the expected value of a product of functions of independent random variables. Theorem 8.4 states that this expected value is the product of the individual expected values:

$$\mathrm{E}[g_1(X_1)g_2(X_2)\cdots g_n(X_n)] = \mathrm{E}[g_1(X_1)]\mathrm{E}[g_2(X_2)]\cdots\mathrm{E}[g_n(X_n)]. \tag{9.30}$$

By Equation (9.30) with $g_i(X_i) = e^{sX_i}$, the expected value of the product is

$$\phi_W(s) = \mathrm{E}\left[e^{sX_1}\right]\mathrm{E}\left[e^{sX_2}\right]\cdots\mathrm{E}\left[e^{sX_n}\right] = \phi_{X_1}(s)\phi_{X_2}(s)\cdots\phi_{X_n}(s). \tag{9.31}$$

When $X_1, \ldots, X_n$ are iid, $\phi_{X_i}(s) = \phi_X(s)$ and thus $\phi_W(s) = (\phi_W(s))^n$.

Moment generating functions provide a convenient way to study the properties of sums of independent finite discrete random variables.

Example 9.5

J and K are independent random variables with probability mass functions

$$\begin{array}{c|ccc} j & 1 & 2 & 3 \\ \hline P_J(j) & 0.2 & 0.6 & 0.2 \end{array}, \qquad \begin{array}{c|cc} k & -1 & 1 \\ \hline P_K(k) & 0.5 & 0.5 \end{array}. \tag{9.32}$$

Find the MGF of $M = J + K$. What are $P_M(m)$ and $E[M^3]$?

J and K have have moment generating functions

$$\phi_J(s) = 0.2e^s + 0.6e^{2s} + 0.2e^{3s}, \qquad \phi_K(s) = 0.5e^{-s} + 0.5e^s. \tag{9.33}$$

Therefore, by Theorem 9.6, $M = J + K$ has MGF

$$\phi_M(s) = \phi_J(s)\phi_K(s) = 0.1 + 0.3e^s + 0.2e^{2s} + 0.3e^{3s} + 0.1e^{4s}. \tag{9.34}$$

The value of $P_M(m)$ at any value of m is the coefficient of e^{ms} in $\phi_M(s)$:

$$\phi_M(s) = \mathrm{E}\left[e^{sM}\right] = \underbrace{0.1}_{P_M(0)} + \underbrace{0.3}_{P_M(1)} e^s + \underbrace{0.2}_{P_M(2)} e^{2s} + \underbrace{0.3}_{P_M(3)} e^{3s} + \underbrace{0.1}_{P_M(4)} e^{4s}.$$

From the coefficients of $\phi_M(s)$, we construct the table for the PMF of M:

$$\begin{array}{c|ccccc} m & 0 & 1 & 2 & 3 & 4 \\ \hline P_M(m) & 0.1 & 0.3 & 0.2 & 0.3 & 0.1 \end{array}.$$

To find the third moment of M, we differentiate $\phi_M(s)$ three times:

$$\begin{aligned} \mathrm{E}\left[M^3\right] &= \left.\frac{d^3\phi_M(s)}{ds^3}\right|_{s=0} \\ &= 0.3e^s + 0.2(2^3)e^{2s} + 0.3(3^3)e^{3s} + 0.1(4^3)e^{4s}\Big|_{s=0} = 16.4. \end{aligned} \tag{9.35}$$

Besides enabling us to calculate probabilities and moments for sums of discrete random variables, we can also use Theorem 9.6 to derive the PMF or PDF of certain sums of iid random variables. In particular, we use Theorem 9.6 to prove that the sum of independent Poisson random variables is a Poisson random variable, and the sum of independent Gaussian random variables is a Gaussian random variable.

Theorem 9.7

If $K_1, \ldots, K_n$ are independent Poisson random variables, $W = K_1 + \cdots + K_n$ is a Poisson random variable.

Proof We adopt the notation $\mathrm{E}[K_i] = \alpha_i$ and note in Table 9.1 that K_i has MGF

$$\phi_{K_i}(s) = e^{\alpha_i(e^s-1)}. \tag{9.36}$$

By Theorem 9.6,

$$\phi_W(s) = e^{\alpha_1(e^s-1)}e^{\alpha_2(e^s-1)}\cdots e^{\alpha_n(e^s-1)} = e^{(\alpha_1+\cdots+\alpha_n)(e^s-1)} = e^{(\alpha_T)(e^s-1)} \tag{9.37}$$

where $\alpha_T = \alpha_1 + \cdots + \alpha_n$. Examining Table 9.1, we observe that $\phi_W(s)$ is the moment generating function of the Poisson (α_T) random variable. Therefore,

$$P_W(w) = \begin{cases} \alpha_T^w e^{-\alpha}/w! & w = 0, 1, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{9.38}$$

Theorem 9.8

The sum of n independent Gaussian random variables $W = X_1 + \cdots + X_n$ is a Gaussian random variable.

Proof For convenience, let $\mu_i = E[X_i]$ and $\sigma_i^2 = \text{Var}[X_i]$. Since the X_i are independent, we know that

$$\begin{aligned} \phi_W(s) &= \phi_{X_1}(s)\phi_{X_2}(s)\cdots\phi_{X_n}(s) \\ &= e^{s\mu_1+\sigma_1^2 s^2/2}e^{s\mu_2+\sigma_2^2 s^2/2}\cdots e^{s\mu_n+\sigma_n^2 s^2/2} \\ &= e^{s(\mu_1+\cdots+\mu_n)+(\sigma_1^2+\cdots+\sigma_n^2)s^2/2}. \end{aligned} \tag{9.39}$$

From Equation (9.39), we observe that $\phi_W(s)$ is the moment generating function of a Gaussian random variable with expected value $\mu_1 + \cdots + \mu_n$ and variance $\sigma_1^2 + \cdots + \sigma_n^2$.

In general, the sum of independent random variables in one family is a different kind of random variable. The following theorem shows that the Erlang (n, λ) random variable is the sum of n independent exponential (λ) random variables.

Theorem 9.9

If $X_1, \ldots, X_n$ are iid exponential (λ) random variables, then $W = X_1 + \cdots + X_n$ has the Erlang PDF

$$f_W(w) = \begin{cases} \frac{\lambda^n w^{n-1} e^{-\lambda w}}{(n-1)!} & w \geq 0, \\ 0 & \textit{otherwise.} \end{cases}$$

Proof In Table 9.1 we observe that each X_i has MGF $\phi_X(s) = \lambda/(\lambda - s)$. By Theorem 9.6, W has MGF

$$\phi_W(s) = \left(\frac{\lambda}{\lambda - s}\right)^n. \tag{9.40}$$

Returning to Table 9.1, we see that W has the MGF of an Erlang (n, λ) random variable.

Similar reasoning demonstrates that the sum of n Bernoulli (p) random variables is the binomial (n,p) random variable, and that the sum of k geometric (p) random variables is a Pascal (k,p) random variable.

Quiz 9.3

(A) Let $K_1, K_2, \ldots, K_m$ be iid discrete uniform random variables with PMF

$$P_K(k) = \begin{cases} 1/n & k = 1,2,\ldots,n, \\ 0 & \text{otherwise.} \end{cases} \tag{9.41}$$

Find the MGF of $J = K_1 + \cdots + K_m$.

(B) Let $X_1, \ldots, X_n$ be independent Gaussian random variables with $E[X_i = 0]$ and $\text{Var}[X_i] = i$. Find the PDF of

$$W = \alpha X_1 + \alpha^2 X_2 + \cdots + \alpha^n X_n. \tag{9.42}$$

9.4 Random Sums of Independent Random Variables

> $R = X_1 + \cdots + R_N$ is a random sum of random variables when N, the number of terms in the sum, is a random variable. When N is independent of each X_i and the X_i are iid, there are concise formulas for the MGF, the expected value, and the variance of R.

Many practical problems can be analyzed by reference to a sum of iid random variables in which the number of terms in the sum is also a random variable. We refer to the resultant random variable, R, as a *random sum* of iid random variables. Thus, given a random variable N and a sequence of iid random variables $X_1, X_2, \ldots$, let

$$R = X_1 + \cdots + X_N. \tag{9.43}$$

The following two examples describe experiments in which the observations are random sums of random variables.

Example 9.6

At a bus terminal, count the number of people arriving on buses during one minute. If the number of people on the ith bus is K_i and the number of arriving buses is N, then the number of people arriving during the minute is

$$R = K_1 + \cdots + K_N. \tag{9.44}$$

In general, the number N of buses that arrive is a random variable. Therefore, R is a random sum of random variables.

Example 9.7

Count the number N of data packets transmitted over a communications link in one minute. Suppose each packet is successfully decoded with probability p, independent of the decoding of any other packet. The number of successfully decoded packets in the one-minute span is

$$R = X_1 + \cdots + X_N. \tag{9.45}$$

where X_i is 1 if the ith packet is decoded correctly and 0 otherwise. When N is a known constant, R is a binomial random variable. By contrast, when N, the number of packets transmitted, is random, R is a random sum.

In the preceding examples we can use the methods of Chapter 5 to find the joint PMF $P_{N,R}(n,r)$. However, we are not able to find a simple closed form expression for the PMF $P_R(r)$. On the other hand, we see in the next theorem that it is possible to express the probability model of R as a formula for the moment generating function $\phi_R(s)$.

Theorem 9.10

Let $\{X_1, X_2, \ldots\}$ be a collection of iid random variables, each with MGF $\phi_X(s)$, and let N be a nonnegative integer-valued random variable that is independent of $\{X_1, X_2, \ldots\}$. The random sum $R = X_1 + \cdots + X_N$ has moment generating function

$$\phi_R(s) = \phi_N(\ln \phi_X(s)).$$

Proof To find $\phi_R(s) = \mathrm{E}[e^{sR}]$, we first find the conditional expected value $\mathrm{E}[e^{sR}|N=n]$. Because this expected value is a function of n, it is a random variable. Theorem 7.14 states that $\phi_R(s)$ is the expected value, with respect to N, of $\mathrm{E}[e^{sR}|N=n]$:

$$\phi_R(s) = \sum_{n=0}^{\infty} \mathrm{E}\left[e^{sR}|N=n\right] P_N(n) = \sum_{n=0}^{\infty} \mathrm{E}\left[e^{s(X_1+\cdots+X_N)}|N=n\right] P_N(n). \tag{9.46}$$

Because the X_i are independent of N,

$$\mathrm{E}\left[e^{s(X_1+\cdots+X_N)}|N=n\right] = \mathrm{E}\left[e^{s(X_1+\cdots+X_n)}\right] = \mathrm{E}\left[e^{sW}\right] = \phi_W(s). \tag{9.47}$$

In Equation (9.46), $W = X_1 + \cdots + X_n$. From Theorem 9.6, we know that $\phi_W(s) = [\phi_X(s)]^n$, implying

$$\phi_R(s) = \sum_{n=0}^{\infty} [\phi_X(s)]^n P_N(n). \tag{9.48}$$

We observe that we can write $[\phi_X(s)]^n = [e^{\ln \phi_X(s)}]^n = e^{[\ln \phi_X(s)]n}$. This implies

$$\phi_R(s) = \sum_{n=0}^{\infty} e^{[\ln \phi_X(s)]n} P_N(n). \tag{9.49}$$

Recognizing that this sum has the same form as the sum in Equation (9.10), we infer that the sum is $\phi_N(s)$ evaluated at $s = \ln \phi_X(s)$. Therefore, $\phi_R(s) = \phi_N(\ln \phi_X(s))$.

In the following example, we find the MGF of a random sum and then transform it to the PMF.

Example 9.8

The number of pages, N, viewed in a Web search has a geometric PMF with expected value $1/q = 4$. The number of bytes K in a Web page has a geometric distribution with expected value $1/p = 10^5$ bytes, independent of the number of bytes in any other page and independent of the number of pages. Find the MGF and the PMF of B, the total number of bytes transmitted in a Web search.

When the ith page has K_i bytes, the total number of bytes is the random sum $B = K_1 + \cdots + K_N$. Thus $\phi_B(s) = \phi_N(\ln \phi_K(s))$. From Table 9.1,

$$\phi_N(s) = \frac{qe^s}{1-(1-q)e^s}, \qquad \phi_K(s) = \frac{pe^s}{1-(1-p)e^s}. \tag{9.50}$$

To calculate $\phi_B(s)$, we substitute $\ln \phi_K(s)$ for every occurrence of s in $\phi_N(s)$. Equivalently, we can substitute $\phi_K(s)$ for every occurrence of e^s in $\phi_N(s)$. This substitution yields

$$\phi_B(s) = \frac{q\left(\frac{pe^s}{1-(1-p)e^s}\right)}{1-(1-q)\left(\frac{pe^s}{1-(1-p)e^s}\right)} = \frac{pqe^s}{1-(1-pq)e^s}. \tag{9.51}$$

By comparing $\phi_K(s)$ and $\phi_B(s)$, we see that B has the MGF of a geometric ($pq = 2.5 \times 10^{-5}$) random variable with expected value $1/(pq) = 400{,}000$ bytes. Therefore, B has the geometric PMF

$$P_B(b) = \begin{cases} pq(1-pq)^{b-1} & b = 1, 2, \ldots, \\ 0 & \text{otherwise}. \end{cases} \tag{9.52}$$

Using Theorem 9.10, we can take derivatives of $\phi_N(\ln \phi_X(s))$ to find simple expressions for the expected value and variance of a random sum R.

Theorem 9.11

For the random sum of iid random variables $R = X_1 + \cdots + X_N$,

$$E[R] = E[N]\,E[X], \qquad \operatorname{Var}[R] = E[N]\operatorname{Var}[X] + \operatorname{Var}[N](E[X])^2.$$

Proof By the chain rule for derivatives,

$$\phi_R'(s) = \phi_N'(\ln \phi_X(s))\frac{\phi_X'(s)}{\phi_X(s)}. \tag{9.53}$$

Scanned with CamScanner

Since $\phi_X(0) = 1$, $\phi'_N(0) = E[N]$, and $\phi'_X(0) = E[X]$, evaluating the equation at $s = 0$ yields

$$E[R] = \phi'_R(0) = \phi'_N(0)\frac{\phi'_X(0)}{\phi_X(0)} = E[N]E[X]. \tag{9.54}$$

For the second derivative of $\phi_X(s)$, we have

$$\phi''_R(s) = \phi''_N(\ln\phi_X(s))\left(\frac{\phi'_X(s)}{\phi_X(s)}\right)^2 + \phi'_N(\ln\phi_X(s))\frac{\phi_X(s)\phi''_X(s) - [\phi'_X(s)]^2}{[\phi_X(s)]^2}. \tag{9.55}$$

The value of this derivative at $s = 0$ is

$$E[R^2] = E[N^2]\mu_X^2 + E[N](E[X^2] - \mu_X^2). \tag{9.56}$$

Subtracting $(E[R])^2 = (\mu_N\mu_X)^2$ from both sides of this equation completes the proof.

We observe that $\text{Var}[R]$ contains two terms: the first term, $\mu_N \text{Var}[X]$, results from the randomness of X, while the second term, $\text{Var}[N]\mu_X^2$, is a consequence of the randomness of N. To see this, consider these two cases.

- Suppose N is deterministic such that $N = n$ every time. In this case, $\mu_N = n$ and $\text{Var}[N] = 0$. The random sum R is an ordinary deterministic sum $R = X_1 + \cdots + X_n$ and $\text{Var}[R] = n\,\text{Var}[X]$.
- Suppose N is random, but each X_i is a deterministic constant x. In this instance, $\mu_X = x$ and $\text{Var}[X] = 0$. Moreover, the random sum becomes $R = Nx$ and $\text{Var}[R] = x^2\,\text{Var}[N]$.

We emphasize that Theorems 9.10 and 9.11 require that N be independent of the random variables $X_1, X_2, \ldots$. That is, the number of terms in the random sum cannot depend on the actual values of the terms in the sum.

Example 9.9

Let $X_1, X_2 \ldots$ be a sequence of independent Gaussian $(100, 10)$ random variables. If K is a Poisson (1) random variable independent of $X_1, X_2 \ldots$, find the expected value and variance of $R = X_1 + \cdots + X_K$.

The PDF and MGF of R are complicated. However, Theorem 9.11 simplifies the calculation of the expected value and the variance. From Appendix A, we observe that a Poisson (1) random variable has variance 1. Thus

$$E[R] = E[X]E[K] = 100, \tag{9.57}$$

and

$$\text{Var}[R] = E[K]\,\text{Var}[X] + \text{Var}[K](E[X])^2 = 100 + (100)^2 = 10,100. \tag{9.58}$$

We see that most of the variance is contributed by the randomness in K. This is true because K is very likely to take on the values 0 and 1 ($P_K(0) = P_K(1) = e^{-1} = 0.368$), and there is a dramatic difference between a sum with no terms and a sum of one or more Gaussian(100,10) random variables.