Probability, Statistics, and Random Processes for Electrical Engineering

Third Edition

Alberto Leon-Garcia

University of Toronto

Contents

Basic Concepts of Probability Theory

This chapter presents the basic concepts of probability theory. In the remainder of the book, we will usually be further developing or elaborating the basic concepts presented here. You will be well prepared to deal with the rest of the book if you have a good understanding of these basic concepts when you complete the chapter.

The following basic concepts will be presented. First, set theory is used to specify the sample space and the events of a random experiment. Second, the axioms of probability specify rules for computing the probabilities of events. Third, the notion of conditional probability allows us to determine how partial information about the outcome of an experiment affects the probabilities of events. Conditional probability also allows us to formulate the notion of "independence" of events and of experiments. Finally, we consider "sequential" random experiments that consist of performing a sequence of simple random subexperiments. We show how the probabilities of events in these experiments can be derived from the probabilities of the simpler subexperiments. Throughout the book it is shown that complex random experiments can be analyzed by decomposing them into simple subexperiments.

2.1   SPECIFYING RANDOM EXPERIMENTS

A random experiment is an experiment in which the outcome varies in an unpredictable fashion when the experiment is repeated under the same conditions. *A random experiment is specified by stating an experimental procedure and a set of one or more measurements or observations.*

Example 2.1

Experiment E_1: Select a ball from an urn containing balls numbered 1 to 50. Note the number of the ball.

Experiment E_2: Select a ball from an urn containing balls numbered 1 to 4. Suppose that balls 1 and 2 are black and that balls 3 and 4 are white. Note the number and color of the ball you select.

Experiment E_3: Toss a coin three times and note the sequence of heads and tails.

Experiment E_4: Toss a coin three times and note the number of heads.

Experiment E_5: Count the number of voice packets containing only silence produced from a group of N speakers in a 10-ms period.

Experiment E_6: A block of information is transmitted repeatedly over a noisy channel until an error-free block arrives at the receiver. Count the number of transmissions required.

Experiment E_7: Pick a number at random between zero and one.

Experiment E_8: Measure the time between page requests in a Web server.

Experiment E_9: Measure the lifetime of a given computer memory chip in a specified environment.

Experiment E_{10}: Determine the value of an audio signal at time t_1.

Experiment E_{11}: Determine the values of an audio signal at times t_1 and t_2.

Experiment E_{12}: Pick two numbers at random between zero and one.

Experiment E_{13}: Pick a number X at random between zero and one, then pick a number Y at random between zero and X.

Experiment E_{14}: A system component is installed at time $t = 0$. For $t \geq 0$ let $X(t) = 1$ as long as the component is functioning, and let $X(t) = 0$ after the component fails.

The specification of a random experiment must include an unambiguous statement of exactly what is measured or observed. For example, random experiments may consist of the same procedure but differ in the observations made, as illustrated by E_3 and E_4.

A random experiment may involve more than one measurement or observation, as illustrated by E_2, E_3, E_{11}, E_{12}, and E_{13}. A random experiment may even involve a continuum of measurements, as shown by E_{14}.

Experiments $E_3, E_4, E_5, E_6, E_{12}$, and E_{13} are examples of sequential experiments that can be viewed as consisting of a sequence of simple subexperiments. Can you identify the subexperiments in each of these? Note that in E_{13} the second subexperiment depends on the outcome of the first subexperiment.

2.1.1 The Sample Space

Since random experiments do not consistently yield the same result, it is necessary to determine the set of possible results. We define an **outcome** or **sample point** of a random experiment as a result that cannot be decomposed into other results. When we perform a random experiment, one and only one outcome occurs. Thus outcomes are mutually exclusive in the sense that they cannot occur simultaneously. The **sample space** S of a random experiment is defined as the set of all possible outcomes.

We will denote an outcome of an experiment by ζ, where ζ is an element or point in S. Each performance of a random experiment can then be viewed as the selection at random of a single point (outcome) from S.

The sample space S can be specified compactly by using set notation. It can be visualized by drawing tables, diagrams, intervals of the real line, or regions of the plane. There are two basic ways to specify a set:

1. *List all the elements*, separated by commas, inside a pair of braces:

$$A = \{0, 1, 2, 3\},$$

2. *Give a property* that specifies the elements of the set:

$$A = \{x : x \text{ is an integer such that } 0 \leq x \leq 3\}.$$

Note that the order in which items are listed does not change the set, e.g., $\{0, 1, 2, 3\}$ and $\{1, 2, 3, 0\}$ are the same set.

Example 2.2

The sample spaces corresponding to the experiments in Example 2.1 are given below using set notation:

$$S_1 = \{1, 2, \ldots, 50\}$$

$$S_2 = \{(1, b), (2, b), (3, w), (4, w)\}$$

$$S_3 = \{\text{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}\}$$

$$S_4 = \{0, 1, 2, 3\}$$

$$S_5 = \{0, 1, 2, \ldots, N\}$$

$$S_6 = \{1, 2, 3, \ldots\}$$

$$S_7 = \{x : 0 \leq x \leq 1\} = [0, 1] \qquad \text{See Fig. 2.1(a).}$$

$$S_8 = \{t : t \geq 0\} = [0, \infty)$$

$$S_9 = \{t : t \geq 0\} = [0, \infty) \qquad \text{See Fig. 2.1(b).}$$

$$S_{10} = \{v : -\infty < v < \infty\} = (-\infty, \infty)$$

$$S_{11} = \{(v_1, v_2) : -\infty < v_1 < \infty \text{ and } -\infty < v_2 < \infty\}$$

$$S_{12} = \{(x, y) : 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1\} \qquad \text{See Fig. 2.1(c).}$$

$$S_{13} = \{(x, y) : 0 \leq y \leq x \leq 1\} \qquad \text{See Fig. 2.1(d).}$$

S_{14} = set of functions $X(t)$ for which $X(t) = 1$ for $0 \leq t < t_0$ and $X(t) = 0$ for $t \geq t_0$, where $t_0 > 0$ is the time when the component fails.

Random experiments involving the same experimental procedure may have different sample spaces as shown by Experiments E_3 and E_4. Thus the purpose of an experiment affects the choice of sample space.

(a) Sample space for Experiment E_7.

(b) Sample space for Experiment E_9.

(c) Sample space for Experiment E_{12}.

(d) Sample space for Experiment E_{13}.
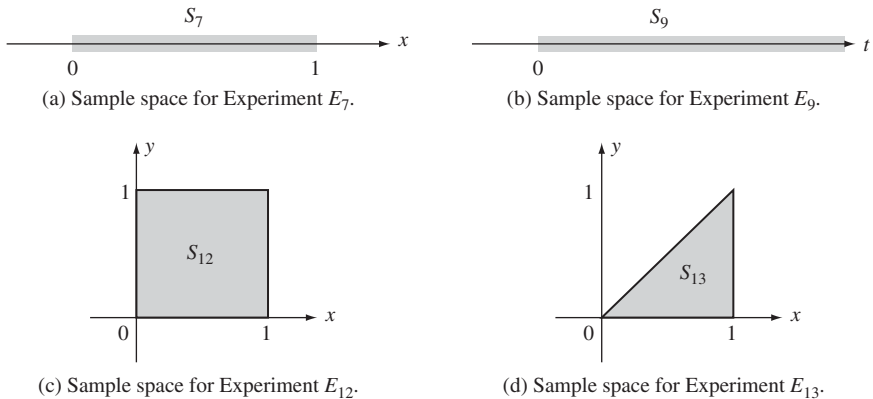
FIGURE 2.1
Sample spaces for Experiments E_7, E_9, E_{12}, and E_{13}.

There are three possibilities for the number of outcomes in a sample space. A sample space can be finite, countably infinite, or uncountably infinite. We call S a **discrete sample space** if S is countable; that is, its outcomes can be put into one-to-one correspondence with the positive integers. We call S a **continuous sample space** if S is not countable. Experiments E_1, E_2, E_3, E_4, and E_5 have finite discrete sample spaces. Experiment E_6 has a countably infinite discrete sample space. Experiments E_7 through E_{13} have continuous sample spaces.

Since an outcome of an experiment can consist of one or more observations or measurements, the sample space S can be multi-dimensional. For example, the outcomes in Experiments E_2, E_{11}, E_{12}, and E_{13} are two-dimensional, and those in Experiment E_3 are three-dimensional. In some instances, the sample space can be written as the Cartesian product of other sets.[1] For example, $S_{11} = R \times R$, where R is the set of real numbers, and $S_3 = S \times S \times S$, where $S = \{H, T\}$.

It is sometimes convenient to let the sample space include outcomes that are impossible. For example, in Experiment E_9 it is convenient to define the sample space as the positive real line, even though a device cannot have an infinite lifetime.

2.1.2   Events

We are usually not interested in the occurrence of specific outcomes, but rather in the occurrence of some event (i.e., whether the outcome satisfies certain conditions). This requires that we consider subsets of S. We say that A is a subset of B if every element of A also belongs to B. For example, in Experiment E_{10}, which involves the measurement of a voltage, we might be interested in the event "signal voltage is negative." The conditions of interest define a subset of the sample space, namely, the set of points ζ from S that satisfy the given conditions. For example, "voltage is negative" corresponds to the set $\{\zeta : -\infty < \zeta < 0\}$. The event occurs if and only if the outcome of the experiment ζ is in this subset. For this reason **events** correspond to subsets of S.

Two events of special interest are the **certain event**, S, which consists of all outcomes and hence always occurs, and the **impossible** or **null event**, $\varnothing$, which contains no outcomes and hence never occurs.

Example 2.3

In the following examples, A_k refers to an event corresponding to Experiment E_k in Example 2.1.

E_1: "An even-numbered ball is selected," $A_1 = \{2, 4, \ldots, 48, 50\}$.
E_2: "The ball is white and even-numbered," $A_2 = \{(4, w)\}$.
E_3: "The three tosses give the same outcome," $A_3 = \{HHH, TTT\}$.
E_4: "The number of heads equals the number of tails," $A_4 = \varnothing$.
E_5: "No active packets are produced," $A_5 = \{0\}$.

---

[1]The Cartesian product of the sets A and B consists of the set of all ordered pairs (a, b), where the first element is taken from A and the second from B.

E_6: "Fewer than 10 transmissions are required," $A_6 = \{1, \dots, 9\}$.

E_7: "The number selected is nonnegative," $A_7 = S_7$.

E_8: "Less than t_0 seconds elapse between page requests," $A_8 = \{t: 0 \le t < t_0\} = [0, t_0)$.

E_9: "The chip lasts more than 1000 hours but fewer than 1500 hours," $A_9 = \{t: 1000 < t < 1500\} = (1000, 1500)$.

E_{10}: "The absolute value of the voltage is less than 1 volt," $A_{10} = \{v: -1 < v < 1\} = (-1, 1)$.

E_{11}: "The two voltages have opposite polarities," $A_{11} = \{(v_1, v_2): (v_1 < 0 \text{ and } v_2 > 0) \text{ or } (v_1 > 0 \text{ and } v_2 < 0)\}$.

E_{12}: "The two numbers differ by less than 1/10," $A_{12} = \{(x, y): (x, y) \text{ in } S_{12} \text{ and } |x - y| < 1/10\}$.

E_{13}: "The two numbers differ by less than 1/10," $A_{13} = \{(x, y): (x, y) \text{ in } S_{13} \text{ and } |x - y| < 1/10\}$.

E_{14}: "The system is functioning at time t_1," $A_{14} = $ subset of S_{14} for which $X(t_1) = 1$.

An event may consist of a single outcome, as in A_2 and A_5. An event from a *discrete* sample space that consists of a single outcome is called an **elementary event**. Events A_2 and A_5 are elementary events. An event may also consist of the entire sample space, as in A_7. The null event, $\varnothing$, arises when none of the outcomes satisfy the conditions that specify a given event, as in A_4.

2.1.3   Review of Set Theory

In random experiments we are interested in the occurrence of events that are represented by sets. We can combine events using set operations to obtain other events. We can also express complicated events as combinations of simple events. Before proceeding with further discussion of events and random experiments, we present some essential concepts from set theory.

A set is a collection of objects and will be denoted by capital letters $S, A, B, \dots$. We define U as the **universal set** that consists of all possible objects of interest in a given setting or application. In the context of random experiments we refer to the universal set as the sample space. For example, the universal set in Experiment E_6 is $U = \{1, 2, \dots\}$. A set A is a collection of objects from U, and these objects are called the **elements** or **points** of the set A and will be denoted by lowercase letters, $\zeta, a, b, x, y, \dots$. We use the notation:

$$x \in A \qquad \text{and} \qquad x \notin A$$

to indicate that "x is an element of A" or "x is not an element of A," respectively.

We use Venn diagrams when discussing sets. A Venn diagram is an illustration of sets and their interrelationships. The universal set U is usually represented as the set of all points within a rectangle as shown in Fig. 2.2(a). The set A is then the set of points within an enclosed region inside the rectangle.

We say A is a **subset** of B if every element of A also belongs to B, that is, if $x \in A$ implies $x \in B$. We say that "A is contained in B" and we write:

$$A \subset B.$$

If A is a subset of B, then the Venn diagram shows the region for A to be inside the region for B as shown in Fig. 2.2(e).
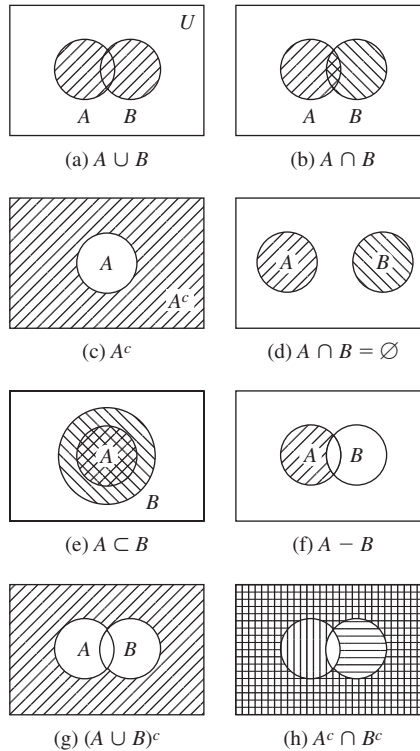
FIGURE 2.2
Set operations and set relations.

Example 2.4

In Experiment E_6 three sets of interest might be $A = \{x : x \geq 10\} = \{10, 11, \dots\}$, that is, 10 or more transmissions are required; $B = \{2, 4, 6, \dots\}$, the number of transmissions is an even number; and $C = \{x : x \geq 20\} = \{20, 21, \dots\}$. Which of these sets are subsets of the others?

Clearly, C is a subset of A ($C \subset A$). However, C is not a subset of B, and B is not a subset of C, because both sets contain elements the other set does not contain. Similarly, B is not a subset of A, and A is not a subset of B.

The **empty set** $\varnothing$ is defined as the set with no elements. The empty set $\varnothing$ is a subset of every set, that is, for any set $A, \varnothing \subset A$.

We say sets **A and B are equal** if they contain the same elements. Since every element in A is also in B, then $x \in A$ implies $x \in B$, so $A \subset B$. Similarly every element in B is also in A, so $x \in B$ implies $x \in A$ and so $B \subset A$. Therefore:

$$A = B \quad \text{if and only if} \quad A \subset B \quad \text{and} \quad B \subset A.$$

The standard method to show that two sets, A and B, are equal is to show that $A \subset B$ and $B \subset A$. A second method is to list all the items in A and all the items in B, and to show that the items are the same. A variation of this second method is to use a

Venn diagram to identify the region that corresponds to A and to then show that the Venn diagram for B occupies the same region. We provide examples of both methods shortly.

We will use three basic operations on sets. The *union* and the *intersection* operations are applied to two sets and produce a third set. The *complement* operation is applied to a single set to produce another set.

The **union** of two sets A and B is denoted by $A \cup B$ and is defined as the set of outcomes that are either in A or in B, or both:

$$A \cup B = \{x : x \in A \quad \text{or} \quad x \in B\}.$$

The operation $A \cup B$ corresponds to the logical "or" of the properties that define set A and set B, that is, x is in $A \cup B$ if x satisfies the property that defines A, or x satisfies the property that defines B, or both. The Venn diagram for $A \cup B$ consists of the shaded region in Fig. 2.2(a).

The **intersection** of two sets A and B is denoted by $A \cap B$ and is defined as the set of outcomes that are in both A and B:

$$A \cap B = \{x : x \in A \quad \text{and} \quad x \in B\}.$$

The operation $A \cap B$ corresponds to the logical "and" of the properties that define set A and set B. The Venn diagram for $A \cap B$ consists of the double shaded region in Fig. 2.2(b). Two sets are said to be **disjoint** or **mutually exclusive** if their intersection is the null set, $A \cap B = \varnothing$. Figure 2.2(d) shows two mutually exclusive sets A and B.

The **complement** of a set A is denoted by A^c and is defined as the set of all elements not in A:

$$A^c = \{x : x \notin A\}.$$

The operation A^c corresponds to the logical "not" of the property that defines set A. Figure 2.2(c) shows A^c. Note that $S^c = \varnothing$ and $\varnothing^c = S$.

The **relative complement** or **difference** of sets A and B is the set of elements in A that are not in B:

$$A - B = \{x : x \in A \text{ and } x \notin B\}.$$

$A - B$ is obtained by removing from A all the elements that are also in B, as illustrated in Fig. 2.2(f). Note that $A - B = A \cap B^c$. Note also that $B^c = S - B$.

Example 2.5

Let A, B, and C be the events from Experiment E_6 in Example 2.4. Find the following events: $A \cup B$, $A \cap B$, A^c, B^c, $A - B$, and $B - A$.

$$A \cup B = \{2, 4, 6, 8, 10, 11, 12, \dots\};$$

$$A \cap B = \{10, 12, 14, \dots\};$$

$$A^c = \{x : x < 10\} = \{1, 2, \dots, 9\};$$

$$B^c = \{1, 3, 5, \dots\};$$

$$A - B = \{11, 13, 15, \dots\};$$

$$\text{and } B - A = \{2, 4, 6, 8\}.$$

The three basic set operations can be combined to form other sets. The following properties of set operations are useful in deriving new expressions for combinations of sets:

Commutative properties:

$$A \cup B = B \cup A \quad \text{and} \quad A \cap B = B \cap A. \tag{2.1}$$

Associative properties:

$$A \cup (B \cup C) = (A \cup B) \cup C \quad \text{and} \quad A \cap (B \cap C) = (A \cap B) \cap C. \tag{2.2}$$

Distributive properties:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad \text{and}$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C). \tag{2.3}$$

By applying the above properties we can derive new identities. DeMorgan's rules provide an important such example:

DeMorgan's rules:

$$(A \cup B)^c = A^c \cap B^c \quad \text{and} \quad (A \cap B)^c = A^c \cup B^c \tag{2.4}$$

Example 2.6

Prove DeMorgan's rules by using Venn diagrams and by demonstrating set equality.

First we will use a Venn diagram to show the first equality. The shaded region in Fig. 2.2(g) shows the complement of $A \cup B$, the left-hand side of the equation. The cross-hatched region in Fig. 2.2(h) shows the intersection of A^c and B^c. The two regions are the same and so the sets are equal. Try sketching the Venn diagrams for the second equality in Eq. (2.4).

Next we prove DeMorgan's rules by proving set equality. The proof has two parts: First we show that $(A \cup B)^c \subset A^c \cap B^c$; then we show that $A^c \cap B^c \subset (A \cup B)^c$. Together these results imply $(A \cup B)^c = A^c \cap B^c$.

First, suppose that $x \in (A \cup B)^c$, then $x \notin A \cup B$. In particular, we have $x \notin A$, which implies $x \in A^c$. Similarly, we have $x \notin B$, which implies $x \in B^c$. Hence x is in both A^c and B^c, that is, $x \in A^c \cap B^c$. We have shown that $(A \cup B)^c \subset A^c \cap B^c$.

To prove inclusion in the other direction, suppose that $x \in A^c \cap B^c$. This implies that $x \in A^c$, so $x \notin A$. Similarly, $x \in B^c$ and so $x \notin B$. Therefore, $x \notin (A \cup B)$ and so $x \in (A \cup B)^c$. We have shown that $A^c \cap B^c \subset (A \cup B)^c$. This proves that $(A \cup B)^c = A^c \cap B^c$.

To prove the second DeMorgan rule, apply the first DeMorgan rule to A^c and B^c to obtain:

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

where we used the identity $A = (A^c)^c$. Now take complements of both sides of the above equation:

$$A^c \cup B^c = (A \cap B)^c.$$

Example 2.7

For Experiment E_{10}, let the sets A, B, and C be defined by

$$A = \{v: |v| > 10\}, \qquad \text{"magnitude of } v \text{ is greater than 10 volts,"}$$

$$B = \{v: v < -5\}, \qquad \text{"} v \text{ is less than } -5 \text{ volts,"}$$

$$C = \{v: v > 0\}, \qquad \text{"} v \text{ is positive."}$$

You should then verify that

$$A \cup B = \{v: v < -5 \text{ or } v > 10\},$$

$$A \cap B = \{v: v < -10\},$$

$$C^c = \{v: v \le 0\},$$

$$(A \cup B) \cap C = \{v: v > 10\},$$

$$A \cap B \cap C = \varnothing, \text{ and}$$

$$(A \cup B)^c = \{v: -5 \le v \le 10\}.$$

The union and intersection operations can be repeated for an arbitrary number of sets. Thus the union of n sets

$$\bigcup_{k=1}^{n} A_k = A_1 \cup A_2 \cup \cdots \cup A_n \qquad (2.5)$$

is the set that consists of all elements that are in A_k for at least one value of k. The same definition applies to the union of a countably infinite sequence of sets:

$$\bigcup_{k=1}^{\infty} A_k. \qquad (2.6)$$

The intersection of n sets

$$\bigcap_{k=1}^{n} A_k = A_1 \cap A_2 \cap \cdots \cap A_n \qquad (2.7)$$

is the set that consists of elements that are in *all* of the sets $A_1, \ldots, A_n$. The same definition applies to the intersection of a countably infinite sequence of sets:

$$\bigcap_{k=1}^{\infty} A_k. \qquad (2.8)$$

We will see that countable unions and intersections of sets are essential in dealing with sample spaces that are not finite.

2.1.4 Event Classes

We have introduced the sample space S as the set of all possible outcomes of the random experiment. We have also introduced events as subsets of S. Probability theory also requires that we state the *class* $\mathcal{F}$ of events of interest. Only events in this class

are assigned probabilities. We expect that any set operation on events in $\mathcal{F}$ will produce a set that is also an event in $\mathcal{F}$. In particular, we insist that complements, as well as countable unions and intersections of events in $\mathcal{F}$, i.e., Eqs. (2.1) and (2.5) through (2.8), result in events in $\mathcal{F}$. When the sample space S is finite or countable, we simply let $\mathcal{F}$ consist of all subsets of S and we can proceed without further concerns about $\mathcal{F}$. However, when S is the real line R (or an interval of the real line), we cannot let $\mathcal{F}$ be all possible subsets of R and still satisfy the axioms of probability. Fortunately, we can obtain all the events of practical interest by letting $\mathcal{F}$ be of the class of events obtained as complements and countable unions and intersections of intervals of the real line, e.g., $(a, b]$ or $(-\infty, b]$. We will refer to this class of events as the *Borel field*. In the remainder of the book, we will refer to the event class $\mathcal{F}$ from time to time. For the introductory-level course in probability you will not need to know more than what is stated in this paragraph.

When we speak of a class of events we are referring to a collection (set) of events (sets), that is, we are speaking of a "set of sets." We refer to the collection of sets as a class to remind us that the elements of the class are sets. We use script capital letters to refer to a class, e.g., $\mathcal{C}, \mathcal{F}, \mathcal{G}$. If the class $\mathcal{C}$ consists of the collection of sets $A_1, \ldots, A_k$, then we write $\mathcal{C} = \{A_1, \ldots, A_k\}$.

Example 2.8

Let $S = \{T, H\}$ be the outcome of a coin toss. Let every subset of S be an event. Find all possible events of S.

An event is a subset of S, so we need to find all possible subsets of S. These are:

$$\mathcal{S} = \{\varnothing, \{H\}, \{T\}, \{H, T\}\}.$$

Note that $\mathcal{S}$ includes both the empty set and S. Let i_T and i_H be binary numbers where $i = 1$ indicates that the corresponding element of S is in a given subset. We generate all possible subsets by taking all possible values of the pair i_T and i_H. Thus $i_T = 0, i_H = 1$ corresponds to the set $\{H\}$. Clearly there are 2^2 possible subsets as listed above.

For a *finite sample space*, $S = \{1, 2, \ldots, k\}$,[2] we usually allow all subsets of S to be events. This class of events is called the **power set of S** and we will denote it by $\mathcal{S}$. We can index all possible subsets of S with binary numbers $i_1, i_2, \ldots, i_k$, and we find that the power set of S has 2^k members. Because of this, the power set is also denoted by $\mathcal{S} = 2^S$.

Section 2.8 discusses some of the fine points on event classes.

2.2   THE AXIOMS OF PROBABILITY

Probabilities are numbers assigned to events that indicate how "likely" it is that the events will occur when an experiment is performed. A **probability law** for a random experiment is a rule that assigns probabilities to the events of the experiment that belong to the event class $\mathcal{F}$. Thus a probability law is a function that assigns a number to sets (events). In Section 1.3 we found a number of properties of relative frequency that any definition of probability should satisfy. The axioms of probability formally state that a

---

[2]The discussion applies to any finite sample space with arbitrary objects $S = \{x_1, \ldots, x_k\}$, but we consider $\{1, 2, \ldots, k\}$ for notational simplicity.

probability law must satisfy these properties. In this section, we develop a number of results that follow from this set of axioms.

Let E be a random experiment with sample space S and event class $\mathcal{F}$. A *probability law* for the experiment E is a rule that assigns to each event $A \in \mathcal{F}$ a number $P[A]$, called the **probability of A**, that satisfies the following axioms:

Axiom I $0 \leq P[A]$

Axiom II $P[S] = 1$

Axiom III   If $A \cap B = \varnothing$, then $P[A \cup B] = P[A] + P[B]$.

Axiom III′   If $A_1, A_2, \ldots$ is a sequence of events such that
$A_i \cap A_j = \varnothing$ for all $i \neq j$, then

$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k].$$

Axioms I, II, and III are enough to deal with experiments with finite sample spaces. In order to handle experiments with infinite sample spaces, Axiom III needs to be replaced by Axiom III′. Note that Axiom III′ includes Axiom III as a special case, by letting $A_k = \varnothing$ for $k \geq 3$. Thus we really only need Axioms I, II, and III′. Nevertheless we will gain greater insight by starting with Axioms I, II, and III.

The axioms allow us to view events as objects possessing a property (i.e., their probability) that has attributes similar to physical mass. Axiom I states that the probability (mass) is nonnegative, and Axiom II states that there is a fixed total amount of probability (mass), namely 1 unit. Axiom III states that the total probability (mass) in two disjoint objects is the sum of the individual probabilities (masses).

The axioms provide us with a set of consistency rules that any valid probability assignment must satisfy. We now develop several properties stemming from the axioms that are useful in the computation of probabilities.

The first result states that if we partition the sample space into two mutually exclusive events, A and A^c, then the probabilities of these two events add up to one.

Corollary 1

$P[A^c] = 1 - P[A]$

Proof: Since an event A and its complement A^c are mutually exclusive, $A \cap A^c = \varnothing$, we have from Axiom III that

$$P[A \cup A^c] = P[A] + P[A^c].$$

Since $S = A \cup A^c$, by Axiom II,

$$1 = P[S] = P[A \cup A^c] = P[A] + P[A^c].$$

The corollary follows after solving for $P[A^c]$.

The next corollary states that the probability of an event is always less than or equal to one. Corollary 2 combined with Axiom I provide good checks in problem

solving: If your probabilities are negative or are greater than one, you have made a mistake somewhere!

Corollary 2

$P[A] \leq 1$

Proof: From Corollary 1,

$$P[A] = 1 - P[A^c] \leq 1,$$

since $P[A^c] \geq 0$.

Corollary 3 states that the impossible event has probability zero.

Corollary 3

$P[\varnothing] = 0$

Proof: Let $A = S$ and $A^c = \varnothing$ in Corollary 1:

$$P[\varnothing] = 1 - P[S] = 0.$$

Corollary 4 provides us with the standard method for computing the probability of a complicated event A. The method involves decomposing the event A into the union of disjoint events $A_1, A_2, \ldots, A_n$. The probability of A is the sum of the probabilities of the A_k's.

Corollary 4

If $A_1, A_2, \ldots, A_n$ are pairwise mutually exclusive, then

$$P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} P[A_k] \qquad \text{for } n \geq 2.$$

Proof: We use mathematical induction. Axiom III implies that the result is true for $n = 2$. Next we need to show that if the result is true for some n, then it is also true for $n + 1$. This, combined with the fact that the result is true for $n = 2$, implies that the result is true for $n \geq 2$.

Suppose that the result is true for some $n > 2$; that is,

$$P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} P[A_k], \tag{2.9}$$

and consider the $n + 1$ case

$$P\left[\bigcup_{k=1}^{n+1} A_k\right] = P\left[\left\{\bigcup_{k=1}^{n} A_k\right\} \cup A_{n+1}\right] = P\left[\bigcup_{k=1}^{n} A_k\right] + P[A_{n+1}], \tag{2.10}$$

where we have applied Axiom III to the second expression after noting that the union of events A_1 to A_n is mutually exclusive with A_{n+1}. The distributive property then implies

$$\left\{\bigcup_{k=1}^{n} A_k\right\} \cap A_{n+1} = \bigcup_{k=1}^{n} \{A_k \cap A_{n+1}\} = \bigcup_{k=1}^{n} \varnothing = \varnothing.$$

Substitution of Eq. (2.9) into Eq. (2.10) gives the $n + 1$ case

$$P\left[\bigcup_{k=1}^{n+1} A_k\right] = \sum_{k=1}^{n+1} P[A_k].$$

Corollary 5 gives an expression for the union of two events that are not necessarily mutually exclusive.

Corollary 5

$P[A \cup B] = P[A] + P[B] - P[A \cap B]$

Proof: First we decompose $A \cup B$, A, and B as unions of disjoint events. From the Venn diagram in Fig. 2.3,

$$P[A \cup B] = P[A \cap B^c] + P[B \cap A^c] + P[A \cap B]$$

$$P[A] = P[A \cap B^c] + P[A \cap B]$$

$$P[B] = P[B \cap A^c] + P[A \cap B]$$

By substituting $P[A \cap B^c]$ and $P[B \cap A^c]$ from the two lower equations into the top equation, we obtain the corollary.

By looking at the Venn diagram in Fig. 2.3, you will see that the sum $P[A] + P[B]$ counts the probability (mass) of the set $A \cap B$ twice. The expression in Corollary 5 makes the appropriate correction.

Corollary 5 is easily generalized to three events,

$$P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[A \cap B]$$

$$- P[A \cap C] - P[B \cap C] + P[A \cap B \cap C], \quad (2.11)$$

and in general to n events, as shown in Corollary 6.
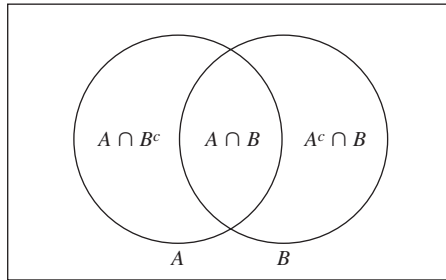
FIGURE 2.3
Decomposition of $A \cup B$ into three disjoint sets.

Corollary 6

$$P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{j=1}^{n} P[A_j] - \sum_{j<k} P[A_j \cap A_k] + \cdots$$

$$+ (-1)^{n+1} P[A_1 \cap \cdots \cap A_n].$$

Proof is by induction (see Problems 2.26 and 2.27).

Since probabilities are nonnegative, Corollary 5 implies that the probability of the union of two events is no greater than the sum of the individual event probabilities

$$P[A \cup B] \le P[A] + P[B]. \tag{2.12}$$

The above inequality is a special case of the fact that a subset of another set must have smaller probability. This result is frequently used to obtain upper bounds for probabilities of interest. In the typical situation, we are interested in an event A whose probability is difficult to find; so we find an event B for which the probability can be found and that includes A as a subset.

Corollary 7

If $A \subset B$, then $P[A] \le P[B]$.

Proof: In Fig. 2.4, B is the union of A and $A^c \cap B$, thus

$$P[B] = P[A] + P[A^c \cap B] \ge P[A],$$

since $P[A^c \cap B] \ge 0$.

The axioms together with the corollaries provide us with a set of rules for computing the probability of certain events in terms of other events. However, we still need an **initial probability assignment** for some basic set of events from which the probability of all other events can be computed. This problem is dealt with in the next two subsections.
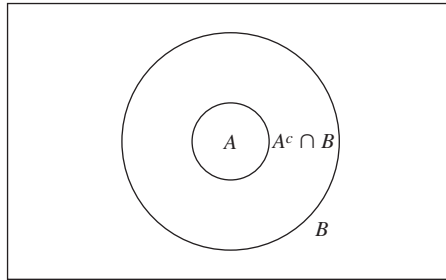
FIGURE 2.4
If $A \subset B$, then $P(A) \le P(B)$.

2.2.1   Discrete Sample Spaces

In this section we show that *the probability law for an experiment with a countable sample space can be specified by giving the probabilities of the elementary events*. First, suppose that the sample space is finite, $S = \{a_1, a_2, \ldots, a_n\}$ and let $\mathcal{F}$ consist of all subsets of S. All distinct elementary events are mutually exclusive, so by Corollary 4 the probability of any event $B = \{a'_1, a'_2, \ldots, a'_m\}$ is given by

$$
\begin{aligned}
P[B] &= P[\{a'_1, a'_2, \ldots, a'_m\}] \\
&= P[\{a'_1\}] + P[\{a'_2\}] + \cdots + P[\{a'_m\}];
\end{aligned} \tag{2.13}
$$

that is, the probability of an event is equal to the sum of the probabilities of the outcomes in the event. Thus we conclude that the probability law for a random experiment with a finite sample space is specified by giving the probabilities of the elementary events.

   If the sample space has n elements, $S = \{a_1, \ldots, a_n\}$, a probability assignment of particular interest is the case of **equally likely outcomes**. The probability of the elementary events is

$$
P[\{a_1\}] = P[\{a_2\}] = \cdots = P[\{a_n\}] = \frac{1}{n}. \tag{2.14}
$$

The probability of any event that consists of k outcomes, say $B = \{a'_1, \ldots, a'_k\}$, is

$$
P[B] = P[\{a'_1\}] + \cdots + P[\{a'_k\}] = \frac{k}{n}. \tag{2.15}
$$

Thus *if outcomes are equally likely, then the probability of an event is equal to the number of outcomes in the event divided by the total number of outcomes in the sample space*. Section 2.3 discusses counting methods that are useful in finding probabilities in experiments that have equally likely outcomes.

   Consider the case where the sample space is countably infinite, $S = \{a_1, a_2, \ldots\}$. Let the event class $\mathcal{F}$ be the class of all subsets of S. Note that $\mathcal{F}$ must now satisfy Eq. (2.8) because events can consist of countable unions of sets. Axiom III$'$ implies that the probability of an event such as $D = \{b_1, b_2, b_3, \ldots\}$ is given by

$$
P[D] = P[\{b'_1, b'_2, b'_3, \ldots\}] = P[\{b'_1\}] + P[\{b'_2\}] + P[\{b'_3\}] + \ldots
$$

The probability of an event with a countably infinite sample space is determined from the probabilities of the elementary events.

Example 2.9

An urn contains 10 identical balls numbered $0, 1, \ldots, 9$. A random experiment involves selecting a ball from the urn and noting the number of the ball. Find the probability of the following events:

$$A = \text{``number of ball selected is odd,''}$$

$$B = \text{``number of ball selected is a multiple of 3,''}$$

$$C = \text{``number of ball selected is less than 5,''}$$

and of $A \cup B$ and $A \cup B \cup C$.

The sample space is $S = \{0, 1, \ldots, 9\}$, so the sets of outcomes corresponding to the above events are

$$A = \{1, 3, 5, 7, 9\}, \qquad B = \{3, 6, 9\}, \qquad \text{and} \qquad C = \{0, 1, 2, 3, 4\}.$$

If we assume that the outcomes are equally likely, then

$$P[A] = P[\{1\}] + P[\{3\}] + P[\{5\}] + P[\{7\}] + P[\{9\}] = \frac{5}{10}.$$

$$P[B] = P[\{3\}] + P[\{6\}] + P[\{9\}] = \frac{3}{10}.$$

$$P[C] = P[\{0\}] + P[\{1\}] + P[\{2\}] + P[\{3\}] + P[\{4\}] = \frac{5}{10}.$$

From Corollary 5,

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] = \frac{5}{10} + \frac{3}{10} - \frac{2}{10} = \frac{6}{10},$$

where we have used the fact that $A \cap B = \{3, 9\}$, so $P[A \cap B] = 2/10$. From Corollary 6,

$$P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[A \cap B]$$

$$- P[A \cap C] - P[B \cap C] + P[A \cap B \cap C]$$

$$= \frac{5}{10} + \frac{3}{10} + \frac{5}{10} - \frac{2}{10} - \frac{2}{10} - \frac{1}{10} + \frac{1}{10}$$

$$= \frac{9}{10}.$$

You should verify the answers for $P[A \cup B]$ and $P[A \cup B \cup C]$ by enumerating the outcomes in the events.

Many probability models can be devised for the same sample space and events by varying the probability assignment; in the case of finite sample spaces all we need to do is come up with n nonnegative numbers that add up to one for the probabilities of the elementary events. Of course, in any particular situation, the probability assignment should be selected to reflect experimental observations to the extent possible. The following example shows that situations can arise where there is more than one "reasonable" probability assignment and where experimental evidence is required to decide on the appropriate assignment.

Example 2.10

Suppose that a coin is tossed three times. If we observe the sequence of heads and tails, then there are eight possible outcomes $S_3 = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$. If we assume that the outcomes of S_3 are equiprobable, then the probability of each of the eight elementary events is 1/8. This probability assignment implies that the probability of obtaining two heads in three tosses is, by Corollary 3,

$$P[\text{"2 heads in 3 tosses"}] = P[\{HHT, HTH, THH\}]$$

$$= P[\{HHT\}] + P[\{HTH\}] + P[\{THH\}] = \frac{3}{8}.$$

Now suppose that we toss a coin three times but we count the number of heads in three tosses instead of observing the sequence of heads and tails. The sample space is now $S_4 = \{0, 1, 2, 3\}$. If we assume the outcomes of S_4 to be equiprobable, then each of the elementary events of S_4 has probability 1/4. This second probability assignment predicts that the probability of obtaining two heads in three tosses is

$$P[\text{"2 heads in 3 tosses"}] = P[\{2\}] = \frac{1}{4}.$$

The first probability assignment implies that the probability of two heads in three tosses is 3/8, and the second probability assignment predicts that the probability is 1/4. Thus the two assignments are not consistent with each other. As far as the theory is concerned, either one of the assignments is acceptable. It is up to us to decide which assignment is more appropriate. Later in the chapter we will see that only the first assignment is consistent with the assumption that the coin is fair and that the tosses are "independent." This assignment correctly predicts the relative frequencies that would be observed in an actual coin tossing experiment.

Finally we consider an example with a countably infinite sample space.

Example 2.11

A fair coin is tossed repeatedly until the first heads shows up; the outcome of the experiment is the number of tosses required until the first heads occurs. Find a probability law for this experiment.

It is conceivable that an arbitrarily large number of tosses will be required until heads occurs, so the sample space is $S = \{1, 2, 3, \dots\}$. Suppose the experiment is repeated n times. Let N_j be the number of trials in which the jth toss results in the first heads. If n is very large, we expect N_1 to be approximately $n/2$ since the coin is fair. This implies that a second toss is necessary about $n - N_1 \approx n/2$ times, and again we expect that about half of these—that is, $n/4$—will result in heads, and so on, as shown in Fig. 2.5. Thus for large n, the relative frequencies are

$$f_j \approx \frac{N_j}{n} = \left(\frac{1}{2}\right)^j \qquad j = 1, 2, \dots.$$

We therefore conclude that a reasonable probability law for this experiment is

$$P[\,j \text{ tosses till first heads}] = \left(\frac{1}{2}\right)^j \qquad j = 1, 2, \dots. \tag{2.16}$$

We can verify that these probabilities add up to one by using the geometric series with $\alpha = 1/2$:

$$\sum_{j=1}^{\infty} \alpha^j = \left.\frac{\alpha}{1 - \alpha}\right|_{\alpha = 1/2} = 1.$$

2.2.2   Continuous Sample Spaces

Continuous sample spaces arise in experiments in which the outcomes are numbers that can assume a continuum of values, so we let the sample space S be the entire real line R (or some interval of the real line). We could consider letting the event class consist of all subsets of R. But it turns out that this class is "too large" and it is impossible
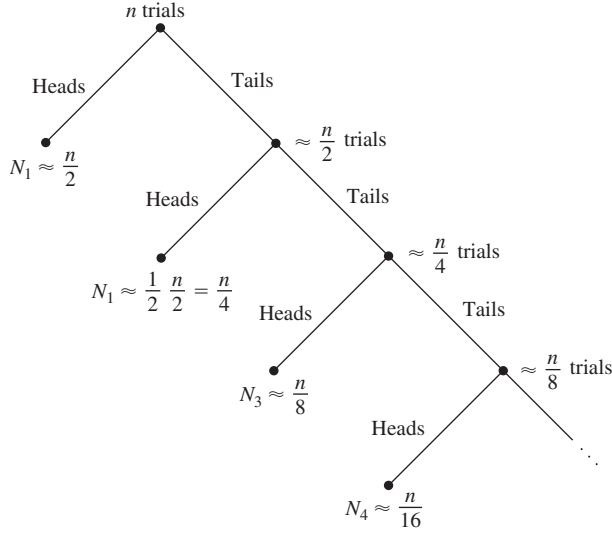
FIGURE 2.5
In n trials heads comes up in the first toss approximately $n/2$ times, in
the second toss approximately $n/4$ times, and so on.

to assign probabilities to all the subsets of R. Fortunately, it is possible to assign proba-
bilities to all events in a smaller class that includes all events of practical interest. This
class denoted by $\mathcal{B}$, is called the **Borel field** and it contains all open and closed intervals
of the real line as well as all events that can be obtained as countable unions, intersec-
tions, and complements.[3] Axiom III′ is once again the key to calculating probabilities of
events. Let $A_1, A_2, \ldots$ be a sequence of mutually exclusive events that are represented
by intervals of the real line, then

$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]$$

where each $P[A_k]$ is specified by the probability law. *For this reason, probability laws
in experiments with continuous sample spaces specify a rule for assigning numbers to in-
tervals of the real line.*

Example 2.12

Consider the random experiment "pick a number x at random between zero and one." The sample
space S for this experiment is the unit interval $[0, 1]$, which is uncountably infinite. If we suppose that
all the outcomes S are equally likely to be selected, then we would guess that the probability that the
outcome is in the interval $[0, 1/2]$ is the same as the probability that the outcome is in the interval
$[1/2, 1]$. We would also guess that the probability of the outcome being exactly equal to $1/2$ would be
zero since there are an uncountably infinite number of equally likely outcomes.

[3]Section 2.9 discusses $\mathcal{B}$ in more detail.

Consider the following probability law: "The probability that the outcome falls in a subinterval of S is equal to the length of the subinterval," that is,

$$P[[a, b]] = (b - a) \qquad \text{for } 0 \le a \le b \le 1, \tag{2.17}$$

where by $P[[a, b]]$ we mean the probability of the event corresponding to the interval $[a, b]$. Clearly, Axiom I is satisfied since $b \ge a \ge 0$. Axiom II follows from $S = [a, b]$ with $a = 0$ and $b = 1$.

We now show that the probability law is consistent with the previous guesses about the probabilities of the events $[0, 1/2]$, $[1/2, 1]$, and $\{1/2\}$:

$$P[[0, 0.5]] = 0.5 - 0 = .5$$

$$P[[0.5, 1]] = 1 - 0.5 = .5$$

In addition, if x_0 is any point in S, then $P[[x_0, x_0]] = 0$ since individual points have zero width.

Now suppose that we are interested in an event that is the union of several intervals; for example, "the outcome is at least 0.3 away from the center of the unit interval," that is, $A = [0, 0.2] \cup [0.8, 1]$. Since the two intervals are disjoint, we have by Axiom III

$$P[A] = P[[0, 0.2]] + P[[0.8, 1]] = .4.$$

The next example shows that an initial probability assignment that specifies the probability of semi-infinite intervals also suffices to specify the probabilities of all events of interest.

Example 2.13

Suppose that the lifetime of a computer memory chip is measured, and we find that "the proportion of chips whose lifetime exceeds t decreases exponentially at a rate α." Find an appropriate probability law.

Let the sample space in this experiment be $S = (0, \infty)$. If we interpret the above finding as "the probability that a chip's lifetime exceeds t decreases exponentially at a rate α," we then obtain the following assignment of probabilities to events of the form (t, ∞):

$$P[(t, \infty)] = e^{-\alpha t} \qquad \text{for } t > 0, \tag{2.18}$$

where $\alpha > 0$. Note that the exponential is a number between 0 and 1 for $t > 0$, so Axiom I is satisfied. Axiom II is satisfied since

$$P[S] = P[(0, \infty)] = 1.$$

The probability that the lifetime is in the interval $(r, s]$ is found by noting in Fig. 2.6 that $(r, s] \cup (s, \infty) = (r, \infty)$, so by Axiom III,

$$P[(r, \infty)] = P[(r, s]] + P[(s, \infty)].$$
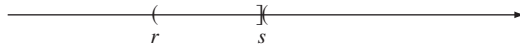
FIGURE 2.6
$(r, \infty) = (r, s] \cup (s, \infty)$.

By rearranging the above equation we obtain

$$P[(r, s]] = P[(r, \infty)] - P[(s, \infty)] = e^{-\alpha r} - e^{-\alpha s}.$$

We thus obtain the probability of arbitrary intervals in S.

In both Example 2.12 and Example 2.13, the probability that the outcome takes on a specific value is zero. You may ask: If an outcome (or event) has probability zero, doesn't that mean it cannot occur? And you may then ask: How can all the outcomes in a sample space have probability zero? We can explain this paradox by using the relative frequency interpretation of probability. An event that occurs only once in an infinite number of trials will have relative frequency zero. Hence the fact that an event or outcome has relative frequency zero does not imply that it cannot occur, but rather that it occurs *very infrequently*. In the case of continuous sample spaces, the set of possible outcomes is so rich that all outcomes occur infrequently enough that their relative frequencies are zero.

We end this section with an example where the events are regions in the plane.

Example 2.14

Consider Experiment E_{12}, where we picked two numbers x and y at random between zero and one. The sample space is then the unit square shown in Fig. 2.7(a). If we suppose that all pairs of numbers in the unit square are equally likely to be selected, then it is reasonable to use a probability assignment in which the probability of any region R inside the unit square is equal to the area of R. Find the probability of the following events: $A = \{x > 0.5\}$, $B = \{y > 0.5\}$, and $C = \{x > y\}$.

(a) Sample space   (b) Event $\{x > \frac{1}{2}\}$

(c) Event $\{y > \frac{1}{2}\}$ (d) Event $\{x > y\}$
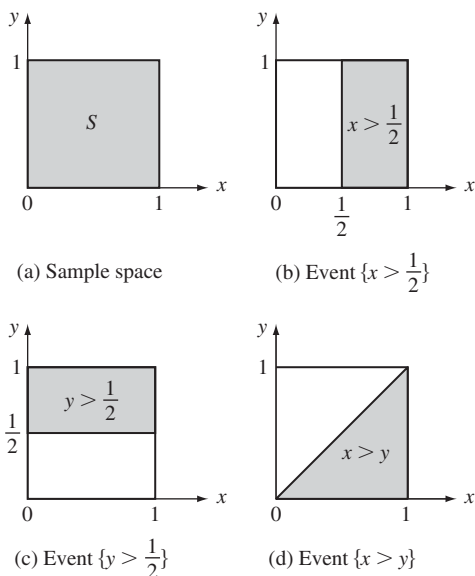
FIGURE 2.7
A two-dimensional sample space and three events.

Figures 2.7(b) through 2.7(d) show the regions corresponding to the events A, B, and C. Clearly each of these regions has area 1/2. Thus

$$P[A] = \frac{1}{2}, \qquad P[B] = \frac{1}{2}, \qquad P[C] = \frac{1}{2}.$$

We reiterate how to proceed from a problem statement to its probability model. The problem statement implicitly or explicitly defines a random experiment, which specifies an experimental procedure and a set of measurements and observations. These measurements and observations determine the set of all possible outcomes and hence the sample space S.

An initial probability assignment that specifies the probability of certain events must be determined next. This probability assignment must satisfy the axioms of probability. If S is discrete, then it suffices to specify the probabilities of elementary events. If S is continuous, it suffices to specify the probabilities of intervals of the real line or regions of the plane. The probability of other events of interest can then be determined from the initial probability assignment and the axioms of probability and their corollaries. Many probability assignments are possible, so the choice of probability assignment must reflect experimental observations and/or previous experience.

*2.3   COMPUTING PROBABILITIES USING COUNTING METHODS[4]

In many experiments with finite sample spaces, the outcomes can be assumed to be equiprobable. The probability of an event is then the ratio of the number of outcomes in the event of interest to the total number of outcomes in the sample space (Eq. (2.15)). The calculation of probabilities reduces to counting the number of outcomes in an event. In this section, we develop several useful counting (combinatorial) formulas.

Suppose that a multiple-choice test has k questions and that for question i the student must select one of n_i possible answers. What is the total number of ways of answering the entire test? The answer to question i can be viewed as specifying the ith component of a k-tuple, so the above question is equivalent to: How many distinct ordered k-tuples $(x_1, \ldots, x_k)$ are possible if x_i is an element from a set with n_i distinct elements?

Consider the $k = 2$ case. If we arrange all possible choices for x_1 and for x_2 along the sides of a table as shown in Fig. 2.8, we see that there are $n_1 n_2$ distinct ordered pairs. For triplets we could arrange the $n_1 n_2$ possible pairs (x_1, x_2) along the vertical side of the table and the n_3 choices for x_3 along the horizontal side. Clearly, the number of possible triplets is $n_1 n_2 n_3$.

In general, *the number of distinct ordered k-tuples $(x_1, \ldots, x_k)$ with components x_i from a set with n_i distinct elements is*

$$\text{number of distinct ordered } k\text{-tuples } = n_1 n_2 \ldots n_k. \qquad (2.19)$$

Many counting problems can be posed as sampling problems where we select "balls" from "urns" or "objects" from "populations." We will now use Eq. (2.19) to develop combinatorial formulas for various types of sampling.

---

[4]This section and all sections marked with an asterisk may be skipped without loss of continuity.

FIGURE 2.8
If there are n_1 distinct choices for x_1 and n_2 distinct choices
for x_2, then there are $n_1 n_2$ distinct ordered pairs (x_1, x_2).

2.3.1   Sampling with Replacement and with Ordering

Suppose we choose k objects from a set A that has n distinct objects, with replacement—that is, after selecting an object and noting its identity in an ordered list, the object is placed back in the set before the next choice is made. We will refer to the set A as the "population." The experiment produces an ordered k-tuple

$$(x_1, \ldots, x_k),$$

where $x_i \in A$ and $i = 1, \ldots, k$. Equation (2.19) with $n_1 = n_2 = \cdots = n_k = n$ implies that

$$\text{number of distinct ordered } k\text{-tuples} = n^k. \tag{2.20}$$

Example 2.15

An urn contains five balls numbered 1 to 5. Suppose we select two balls from the urn with replacement. How many distinct *ordered* pairs are possible? What is the probability that the two draws yield the same number?

Equation (2.20) states that the number of ordered pairs is $5^2 = 25$. Table 2.1 shows the 25 possible pairs. Five of the 25 outcomes have the two draws yielding the same number; if we suppose that all pairs are equiprobable, then the probability that the two draws yield the same number is $5/25 = .2$.

2.3.2   Sampling without Replacement and with Ordering

Suppose we choose k objects in succession without replacement from a population A of n distinct objects. Clearly, $k \leq n$. The number of possible outcomes in the first draw is $n_1 = n$; the number of possible outcomes in the second draw is $n_2 = n - 1$, namely all n objects except the one selected in the first draw; and so on, up to $n_k = n - (k - 1)$ in the final draw. Equation (2.19) then gives

$$\text{number of distinct ordered } k\text{-tuples} = n(n - 1) \ldots (n - k + 1). \tag{2.21}$$

TABLE 2.1   Enumeration of possible outcomes in various types of sampling of two balls from an urn containing five distinct balls.

(a) Ordered pairs for sampling with replacement.

$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$
$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$
$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$
$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$
$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$

(b) Ordered pairs for sampling without replacement.

	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$
$(2,1)$		$(2,3)$	$(2,4)$	$(2,5)$
$(3,1)$	$(3,2)$		$(3,4)$	$(3,5)$
$(4,1)$	$(4,2)$	$(4,3)$		$(4,5)$
$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	

(c) Pairs for sampling without replacement or ordering.

$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$
	$(2,3)$	$(2,4)$	$(2,5)$
		$(3,4)$	$(3,5)$
			$(4,5)$

Example 2.16

An urn contains five balls numbered 1 to 5. Suppose we select two balls in succession without replacement. How many distinct *ordered* pairs are possible? What is the probability that the first ball has a number larger than that of the second ball?

Equation (2.21) states that the number of ordered pairs is $5(4) = 20$. The 20 possible ordered pairs are shown in Table 2.1(b). Ten ordered pairs in Tab. 2.1(b) have the first number larger than the second number; thus the probability of this event is $10/20 = 1/2$.

Example 2.17

An urn contains five balls numbered $1, 2, \ldots, 5$. Suppose we draw three balls with replacement. What is the probability that all three balls are different?

From Eq. (2.20) there are $5^3 = 125$ possible outcomes, which we will suppose are equiprobable. The number of these outcomes for which the three draws are different is given by Eq. (2.21): $5(4)(3) = 60$. Thus the probability that all three balls are different is $60/125 = .48$.

2.3.3   Permutations of *n* Distinct Objects

Consider sampling without replacement with $k = n$. This is simply drawing objects from an urn containing n distinct objects until the urn is empty. Thus, the number of possible orderings (arrangements, permutations) of n distinct objects is equal to the

number of ordered n-tuples in sampling without replacement with $k = n$. From Eq. (2.21), we have

$$\text{number of permutations of } n \text{ objects} = n(n - 1)\ldots(2)(1) \triangleq n!. \qquad (2.22)$$

We refer to $n!$ as **n factorial**.

We will see that $n!$ appears in many of the combinatorial formulas. For large n, **Stirling's formula** is very useful:

$$n! \sim \sqrt{2\pi}\, n^{n+1/2} e^{-n}, \qquad (2.23)$$

where the sign $\sim$ indicates that the ratio of the two sides tends to unity as $n \to \infty$ [Feller, p. 52].

Example 2.18

Find the number of permutations of three distinct objects $\{1, 2, 3\}$. Equation (2.22) gives $3! = 3(2)(1) = 6$. The six permutations are

$$123 \quad 312 \quad 231 \quad 132 \quad 213 \quad 321.$$

Example 2.19

Suppose that 12 balls are placed at random into 12 cells, where more than 1 ball is allowed to occupy a cell. What is the probability that all cells are occupied?

The placement of each ball into a cell can be viewed as the selection of a cell number between 1 and 12. Equation (2.20) implies that there are 12^{12} possible placements of the 12 balls in the 12 cells. In order for all cells to be occupied, the first ball selects from any of the 12 cells, the second ball from the remaining 11 cells, and so on. Thus the number of placements that occupy all cells is $12!$. If we suppose that all 12^{12} possible placements are equiprobable, we find that the probability that all cells are occupied is

$$\frac{12!}{12^{12}} = \left(\frac{12}{12}\right)\left(\frac{11}{12}\right)\ldots\left(\frac{1}{12}\right) = 5.37(10^{-5}).$$

This answer is surprising if we reinterpret the question as follows. Given that 12 airplane crashes occur at random in a year, what is the probability that there is exactly 1 crash each month? The above result shows that this probability is very small. Thus a model that assumes that crashes occur randomly in time does *not* predict that they tend to occur uniformly over time [Feller, p. 32].

2.3.4   Sampling without Replacement and without Ordering

Suppose we pick k objects from a set of n distinct objects without replacement and that we record the result without regard to order. (You can imagine putting each selected object into another jar, so that when the k selections are completed we have no record of the order in which the selection was done.) We call the resulting subset of k selected objects a "combination of size k."

From Eq. (2.22), there are $k!$ possible orders in which the k objects in the second jar could have been selected. Thus if C_k^n denotes the number of combinations of size k

from a set of size n, then $C_k^n k!$ must be the total number of distinct ordered samples of k objects, which is given by Eq. (2.21). Thus

$$C_k^n k! = n(n - 1) \ldots (n - k + 1), \tag{2.24}$$

and the *number of different combinations of size k from a set of size n, $k \leq n$, is*

$$C_k^n = \frac{n(n - 1) \ldots (n - k + 1)}{k!} = \frac{n!}{k! \, (n - k)!} \triangleq \binom{n}{k}. \tag{2.25}$$

The expression $\binom{n}{k}$ is called a **binomial coefficient** and is read "n choose k."

Note that choosing k objects out of a set of n is equivalent to choosing the $n - k$ objects that are to be left out. It then follows that (also see Problem 2.60):

$$\binom{n}{k} = \binom{n}{n - k}.$$

Example 2.20

Find the number of ways of selecting two objects from $A = \{1, 2, 3, 4, 5\}$ without regard to order.
    Equation (2.25) gives

$$\binom{5}{2} = \frac{5!}{2! \, 3!} = 10.$$

Table 2.1(c) gives the 10 pairs.

Example 2.21

Find the number of distinct permutations of k white balls and $n - k$ black balls.

This problem is equivalent to the following sampling problem: Put n tokens numbered 1 to n in an urn, where each token represents a position in the arrangement of balls; pick a combination of k tokens and put the k white balls in the corresponding positions. Each combination of size k leads to a distinct arrangement (permutation) of k white balls and $n - k$ black balls. Thus the number of distinct permutations of k white balls and $n - k$ black balls is C_k^n.

As a specific example let $n = 4$ and $k = 2$. The number of combinations of size 2 from a set of four distinct objects is

$$\binom{4}{2} = \frac{4!}{2! \, 2!} = \frac{4(3)}{2(1)} = 6.$$

The 6 distinct permutations with 2 whites (zeros) and 2 blacks (ones) are

$$1100 \quad 0110 \quad 0011 \quad 1001 \quad 1010 \quad 0101.$$

Example 2.22   Quality Control

A batch of 50 items contains 10 defective items. Suppose 10 items are selected at random and tested. What is the probability that exactly 5 of the items tested are defective?

The number of ways of selecting 10 items out of a batch of 50 is the number of combinations of size 10 from a set of 50 objects:

$$\binom{50}{10} = \frac{50!}{10! \, 40!}.$$

The number of ways of selecting 5 defective and 5 nondefective items from the batch of 50 is the product $N_1 N_2$, where N_1 is the number of ways of selecting the 5 items from the set of 10 defective items, and N_2 is the number of ways of selecting 5 items from the 40 nondefective items. Thus the probability that exactly 5 tested items are defective is

$$\frac{\binom{10}{5}\binom{40}{5}}{\binom{50}{10}} = \frac{10! \, 40! \, 10! \, 40!}{5! \, 5! \, 35! \, 5! \, 50!} = .016.$$

Example 2.21 shows that sampling without replacement and without ordering is equivalent to partitioning the set of n distinct objects into two sets: B, containing the k items that are picked from the urn, and B^c, containing the $n - k$ left behind. Suppose we partition a set of n distinct objects into $\mathcal{J}$ subsets $B_1, B_2, \ldots, B_{\mathcal{J}}$, where $B_{\mathcal{J}}$ is assigned $k_{\mathcal{J}}$ elements and $k_1 + k_2 + \cdots + k_{\mathcal{J}} = n$.

In Problem 2.61, it is shown that the number of distinct partitions is

$$\frac{n!}{k_1! \, k_2! \ldots k_{\mathcal{J}}!}. \tag{2.26}$$

Equation (2.26) is called the **multinomial coefficient**. The binomial coefficient is the $\mathcal{J} = 2$ case of the multinomial coefficient.

Example 2.23

A six-sided die is tossed 12 times. How many distinct sequences of faces (numbers from the set $\{1, 2, 3, 4, 5, 6\}$) have each number appearing exactly twice? What is the probability of obtaining such a sequence?

The number of distinct sequences in which each face of the die appears exactly twice is the same as the number of partitions of the set $\{1, 2, \ldots, 12\}$ into 6 subsets of size 2, namely

$$\frac{12!}{2! \, 2! \, 2! \, 2! \, 2! \, 2!} = \frac{12!}{2^6} = 7,484,400.$$

From Eq. (2.20) we have that there are 6^{12} possible outcomes in 12 tosses of a die. If we suppose that all of these have equal probabilities, then the probability of obtaining a sequence in which each face appears exactly twice is

$$\frac{12!/2^6}{6^{12}} = \frac{7,484,400}{2,176,782,336} \simeq 3.4(10^{-3}).$$

2.3.5   Sampling with Replacement and without Ordering

Suppose we pick k objects from a set of n distinct objects with replacement and we record the result without regard to order. This can be done by filling out a form which has n columns, one for each distinct object. Each time an object is selected, an "x" is placed in the corresponding column. For example, if we are picking 5 objects from 4 distinct objects, one possible form would look like this:

Object 1	*Object 2*	*Object 3*	*Object 4*
xx /	/	x	/	xx

where the slash symbol ("/") is used to separate the entries for different columns. Note that this form can be summarized by the sequence

$$xx//x/xx$$

where the $n - 1$ /'s indicate the lines between columns, and where nothing appears between consecutive /'s if the corresponding object was not selected. Each different arrangement of 5 x's and 3 /'s leads to a distinct form. If we identify x's with "white balls" and /'s with "black balls," then this problem was considered in Example 2.21, and the number of different arrangements is given by $\binom{8}{3}$.

In the general case the form will involve k x's and $n - 1$ /'s. Thus the *number of different ways of picking k objects from a set of n distinct objects with replacement and without ordering* is given by

$$\binom{n - 1 + k}{k} = \binom{n - 1 + k}{n - 1}.$$

2.4   CONDITIONAL PROBABILITY

Quite often we are interested in determining whether two events, A and B, are related in the sense that knowledge about the occurrence of one, say B, alters the likelihood of occurrence of the other, A. This requires that we find the **conditional probability**, $P[A \mid B]$, of event A given that event B has occurred. The conditional probability is defined by

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]} \qquad \text{for } P[B] > 0. \tag{2.27}$$

Knowledge that event B has occurred implies that the outcome of the experiment is in the set B. In computing $P[A \mid B]$ we can therefore view the experiment as now having the reduced sample space B as shown in Fig. 2.9. The event A occurs in the reduced sample space if and only if the outcome ζ is in $A \cap B$. Equation (2.27) simply renormalizes the probability of events that occur jointly with B. Thus if we let $A = B$, Eq. (2.27) gives $P[B \mid B] = 1$, as required. It is easy to show that $P[A \mid B]$, for fixed B, satisfies the axioms of probability. (See Problem 2.74.)

If we interpret probability as relative frequency, then $P[A \mid B]$ should be the relative frequency of the event $A \cap B$ in experiments where B occurred. Suppose that the experiment is performed n times, and suppose that event B occurs n_B times, and that
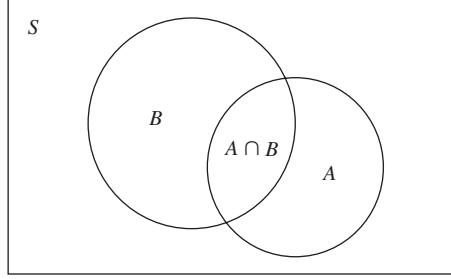
FIGURE 2.9
If *B* is known to have occurred, then *A* can occur only
if $A \cap B$ occurs.

event $A \cap B$ occurs $n_{A \cap B}$ times. The relative frequency of interest is then

$$\frac{n_{A \cap B}}{n_B} = \frac{n_{A \cap B}/n}{n_B/n} \rightarrow \frac{P[A \cap B]}{P[B]},$$

where we have implicitly assumed that $P[B] > 0$. This is in agreement with Eq. (2.27).

Example 2.24

A ball is selected from an urn containing two black balls, numbered 1 and 2, and two white balls, numbered 3 and 4. The number and color of the ball is noted, so the sample space is $\{(1, b), (2, b), (3, w), (4, w)\}$. Assuming that the four outcomes are equally likely, find $P[A \mid B]$ and $P[A \mid C]$, where A, B, and C are the following events:

$A = \{(1, b), (2, b)\}$, "black ball selected,"

$B = \{(2, b), (4, w)\}$, "even-numbered ball selected," and

$C = \{(3, w), (4, w)\}$, "number of ball is greater than 2."

Since $P[A \cap B] = P[(2, b)]$ and $P[A \cap C] = P[\varnothing] = 0$, Eq. (2.24) gives

$$P[A \mid B] = \frac{P[A \cap B]}{P[B]} = \frac{.25}{.5} = .5 = P[A]$$

$$P[A \mid C] = \frac{P[A \cap C]}{P[C]} = \frac{0}{.5} = 0 \neq P[A].$$

In the first case, knowledge of B did not alter the probability of A. In the second case, knowledge of C implied that A had not occurred.

If we multiply both sides of the definition of $P[A \mid B]$ by $P[B]$ we obtain

$$P[A \cap B] = P[A \mid B]P[B]. \tag{2.28a}$$

Similarly we also have that

$$P[A \cap B] = P[B \mid A]P[A]. \tag{2.28b}$$

In the next example we show how this equation is useful in finding probabilities in sequential experiments. The example also introduces a **tree diagram** that facilitates the calculation of probabilities.

Example 2.25

An urn contains two black balls and three white balls. Two balls are selected at random from the urn without replacement and the sequence of colors is noted. Find the probability that both balls are black.

This experiment consists of a sequence of two subexperiments. We can imagine working our way down the tree shown in Fig. 2.10 from the topmost node to one of the bottom nodes: We reach node 1 in the tree if the outcome of the first draw is a black ball; then the next subexperiment consists of selecting a ball from an urn containing one black ball and three white balls. On the other hand, if the outcome of the first draw is white, then we reach node 2 in the tree and the second subexperiment consists of selecting a ball from an urn that contains two black balls and two white balls. Thus if we know which node is reached after the first draw, then we can state the probabilities of the outcome in the next subexperiment.

Let B_1 and B_2 be the events that the outcome is a black ball in the first and second draw, respectively. From Eq. (2.28b) we have

$$P[B_1 \cap B_2] = P[B_2 \,|\, B_1]P[B_1].$$

In terms of the tree diagram in Fig. 2.10, $P[B_1]$ is the probability of reaching node 1 and $P[B_2 \,|\, B_1]$ is the probability of reaching the leftmost bottom node from node 1. Now $P[B_1] = 2/5$ since the first draw is from an urn containing two black balls and three white balls; $P[B_2 \,|\, B_1] = 1/4$ since, given B_1, the second draw is from an urn containing one black ball and three white balls. Thus

$$P[B_1 \cap B_2] = \frac{1}{4}\frac{2}{5} = \frac{1}{10}.$$

In general, the probability of any sequence of colors is obtained by multiplying the probabilities corresponding to the node transitions in the tree in Fig. 2.10.
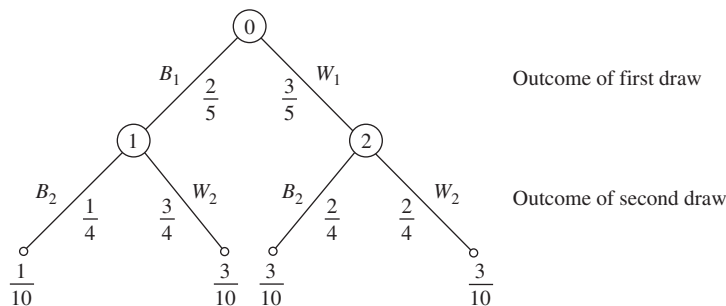
FIGURE 2.10

The paths from the top node to a bottom node correspond to the possible outcomes in the drawing of two balls from an urn without replacement. The probability of a path is the product of the probabilities in the associated transitions.

Example 2.26   Binary Communication System

Many communication systems can be modeled in the following way. First, the user inputs a 0 or a 1 into the system, and a corresponding signal is transmitted. Second, the receiver makes a decision about what was the input to the system, based on the signal it received. Suppose that the user sends 0s with probability $1 - p$ and 1s with probability p, and suppose that the receiver makes random decision errors with probability ε. For $i = 0, 1$, let A_i be the event "input was i," and let B_i be the event "receiver decision was i." Find the probabilities $P[A_i \cap B_j]$ for $i = 0, 1$ and $j = 0, 1$.

The tree diagram for this experiment is shown in Fig. 2.11. We then readily obtain the desired probabilities

$$P[A_0 \cap B_0] = (1 - p)(1 - \varepsilon),$$
$$P[A_0 \cap B_1] = (1 - p)\varepsilon,$$
$$P[A_1 \cap B_0] = p\varepsilon, \text{ and}$$
$$P[A_1 \cap B_1] = p(1 - \varepsilon).$$

Let $B_1, B_2, \ldots, B_n$ be mutually exclusive events whose union equals the sample space S as shown in Fig. 2.12. We refer to these sets as a **partition** of S. Any event A can be represented as the union of mutually exclusive events in the following way:

$$A = A \cap S = A \cap (B_1 \cup B_2 \cup \cdots \cup B_n)$$
$$= (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n).$$

(See Fig. 2.12.) By Corollary 4, the probability of A is

$$P[A] = P[A \cap B_1] + P[A \cap B_2] + \cdots + P[A \cap B_n].$$

By applying Eq. (2.28a) to each of the terms on the right-hand side, we obtain the **theorem on total probability**:

$$P[A] = P[A \mid B_1]P[B_1] + P[A \mid B_2]P[B_2] + \cdots + P[A \mid B_n]P[B_n]. \quad (2.29)$$

This result is particularly useful when the experiments can be viewed as consisting of a sequence of two subexperiments as shown in the tree diagram in Fig. 2.10.
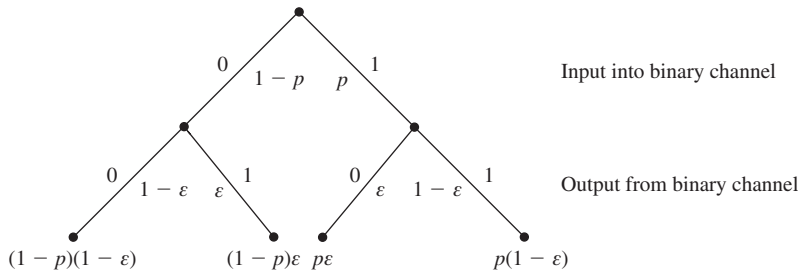
FIGURE 2.11
Probabilities of input-output pairs in a binary transmission system.
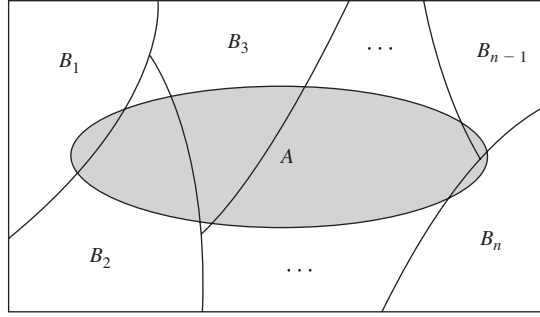
FIGURE 2.12
A partition of S into n disjoint sets.

Example 2.27

In the experiment discussed in Example 2.25, find the probability of the event W_2 that the second ball is white.

The events $B_1 = \{(b, b), (b, w)\}$ and $W_1 = \{(w, b), (w, w)\}$ form a partition of the sample space, so applying Eq. (2.29) we have

$$P[W_2] = P[W_2 | B_1]P[B_1] + P[W_2 | W_1]P[W_1]$$

$$= \frac{3}{4}\frac{2}{5} + \frac{1}{2}\frac{3}{5} = \frac{3}{5}.$$

It is interesting to note that this is the same as the probability of selecting a white ball in the first draw. The result makes sense because we are computing the probability of a white ball in the second draw under the assumption that we have no knowledge of the outcome of the first draw.

Example 2.28

A manufacturing process produces a mix of "good" memory chips and "bad" memory chips. The lifetime of good chips follows the exponential law introduced in Example 2.13, with a rate of failure α. The lifetime of bad chips also follows the exponential law, but the rate of failure is 1000α. Suppose that the fraction of good chips is $1 - p$ and of bad chips, p. Find the probability that a randomly selected chip is still functioning after t seconds.

Let C be the event "chip still functioning after t seconds," and let G be the event "chip is good," and B the event "chip is bad." By the theorem on total probability we have

$$P[C] = P[C|G]P[G] + P[C|B]P[B]$$

$$= P[C|G](1 - p) + P[C|B]p$$

$$= (1 - p)e^{-\alpha t} + pe^{-1000\alpha t},$$

where we used the fact that $P[C|G] = e^{-\alpha t}$ and $P[C|B] = e^{-1000\alpha t}$.

2.4.1 Bayes' Rule

Let $B_1, B_2, \ldots, B_n$ be a partition of a sample space S. Suppose that event A occurs; what is the probability of event B_j? By the definition of conditional probability we have

$$P[B_j|A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A|B_j]P[B_j]}{\sum_{k=1}^{n} P[A|B_k]P[B_k]}, \qquad (2.30)$$

where we used the theorem on total probability to replace $P[A]$. Equation (2.30) is called **Bayes' rule**.

Bayes' rule is often applied in the following situation. We have some random experiment in which the events of interest form a partition. The "a priori probabilities" of these events, $P[B_j]$, are the probabilities of the events before the experiment is performed. Now suppose that the experiment is performed, and we are informed that event A occurred; the "a posteriori probabilities" are the probabilities of the events in the partition, $P[B_j|A]$, given this additional information. The following two examples illustrate this situation.

Example 2.29 Binary Communication System

In the binary communication system in Example 2.26, find which input is more probable given that the receiver has output a 1. Assume that, a priori, the input is equally likely to be 0 or 1.

Let A_k be the event that the input was $k, k = 0, 1$, then A_0 and A_1 are a partition of the sample space of input-output pairs. Let B_1 be the event "receiver output was a 1." The probability of B_1 is

$$P[B_1] = P[B_1|A_0]P[A_0] + P[B_1|A_1]P[A_1]$$

$$= \varepsilon\left(\frac{1}{2}\right) + (1 - \varepsilon)\left(\frac{1}{2}\right) = \frac{1}{2}.$$

Applying Bayes' rule, we obtain the a posteriori probabilities

$$P[A_0|B_1] = \frac{P[B_1|A_0]P[A_0]}{P[B_1]} = \frac{\varepsilon/2}{1/2} = \varepsilon$$

$$P[A_1|B_1] = \frac{P[B_1|A_1]P[A_1]}{P[B_1]} = \frac{(1 - \varepsilon)/2}{1/2} = (1 - \varepsilon).$$

Thus, if ε is less than 1/2, then input 1 is more likely than input 0 when a 1 is observed at the output of the channel.

Example 2.30 Quality Control

Consider the memory chips discussed in Example 2.28. Recall that a fraction p of the chips are bad and tend to fail much more quickly than good chips. Suppose that in order to "weed out" the bad chips, every chip is tested for t seconds prior to leaving the factory. The chips that fail are discarded and the remaining chips are sent out to customers. Find the value of t for which 99% of the chips sent out to customers are good.

Let C be the event "chip still functioning after t seconds," and let G be the event "chip is good," and B be the event "chip is bad." The problem requires that we find the value of t for which

$$P[G|C] = .99.$$

We find $P[G|C]$ by applying Bayes' rule:

$$P[G|C] = \frac{P[C|G]P[G]}{P[C|G]P[G] + P[C|B]P[B]}$$

$$= \frac{(1 - p)e^{-\alpha t}}{(1 - p)e^{-\alpha t} + pe^{-\alpha 1000t}}$$

$$= \frac{1}{1 + \dfrac{pe^{-\alpha 1000t}}{(1 - p)e^{-\alpha t}}} = .99.$$

The above equation can then be solved for t:

$$t = \frac{1}{999\alpha} \ln\left(\frac{99p}{1 - p}\right).$$

For example, if $1/\alpha = 20{,}000$ hours and $p = .10$, then $t = 48$ hours.

2.5   INDEPENDENCE OF EVENTS

If knowledge of the occurrence of an event B does not alter the probability of some other event A, then it would be natural to say that event A is independent of B. In terms of probabilities this situation occurs when

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

The above equation has the problem that the right-hand side is not defined when $P[B] = 0$.

We will define two events A and B to be **independent** if

$$P[A \cap B] = P[A]P[B]. \tag{2.31}$$

Equation (2.31) then implies both

$$P[A|B] = P[A] \tag{2.32a}$$

and

$$P[B|A] = P[B] \tag{2.32b}$$

Note also that Eq. (2.32a) implies Eq. (2.31) when $P[B] \neq 0$ and Eq. (2.32b) implies Eq. (2.31) when $P[A] \neq 0$.

Example 2.31

A ball is selected from an urn containing two black balls, numbered 1 and 2, and two white balls, numbered 3 and 4. Let the events A, B, and C be defined as follows:

$A = \{(1, b), (2, b)\}$, "black ball selected";

$B = \{(2, b), (4, w)\}$, "even-numbered ball selected"; and

$C = \{(3, w), (4, w)\}$, "number of ball is greater than 2."

Are events A and B independent? Are events A and C independent?

First, consider events A and B. The probabilities required by Eq. (2.31) are

$$P[A] = P[B] = \frac{1}{2},$$

and

$$P[A \cap B] = P[\{(2, b)\}] = \frac{1}{4}.$$

Thus

$$P[A \cap B] = \frac{1}{4} = P[A]P[B],$$

and the events A and B are independent. Equation (2.32b) gives more insight into the meaning of independence:

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{P[\{(2, b)\}]}{P[\{(2, b), (4, w)\}]} = \frac{1/4}{1/2} = \frac{1}{2}$$

$$P[A] = \frac{P[A]}{P[S]} = \frac{P[\{(1, b), (2, b)\}]}{P[\{(1, b), (2, b), (3, w), (4, w)\}]} = \frac{1/2}{1}.$$

These two equations imply that $P[A] = P[A|B]$ because *the proportion of outcomes in S that lead to the occurrence of A is equal to the proportion of outcomes in B that lead to A*. Thus knowledge of the occurrence of B does not alter the probability of the occurrence of A.

Events A and C are not independent since $P[A \cap C] = P[\varnothing] = 0$ so

$$P[A|C] = 0 \neq P[A] = .5.$$

In fact, A and C are mutually exclusive since $A \cap C = \varnothing$, so the occurrence of C implies that A has definitely not occurred.

In general if two events have nonzero probability and are mutually exclusive, then they cannot be independent. For suppose they were independent *and* mutually exclusive; then

$$0 = P[A \cap B] = P[A]P[B],$$

which implies that at least one of the events must have zero probability.

Example 2.32

Two numbers x and y are selected at random between zero and one. Let the events A, B, and C be defined as follows:

$$A = \{x > 0.5\}, \qquad B = \{y > 0.5\}, \qquad \text{and } C = \{x > y\}.$$

Are the events A and B independent? Are A and C independent?

Figure 2.13 shows the regions of the unit square that correspond to the above events. Using Eq. (2.32a), we have

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = \frac{1/4}{1/2} = \frac{1}{2} = P[A],$$

so events A and B are independent. Again we have that the "proportion" of outcomes in S leading to A is equal to the "proportion" in B that lead to A.

Using Eq. (2.32b), we have

$$P[A|C] = \frac{P[A \cap C]}{P[C]} = \frac{3/8}{1/2} = \frac{3}{4} \neq \frac{1}{2} = P[A],$$

so events A and C are not independent. Indeed from Fig. 2.13(b) we can see that knowledge of the fact that x is greater than y increases the probability that x is greater than 0.5.

What conditions should three events A, B, and C satisfy in order for them to be independent? First, they should be pairwise independent, that is,

$$P[A \cap B] = P[A]P[B], P[A \cap C] = P[A]P[C], \text{ and } P[B \cap C] = P[B]P[C].$$
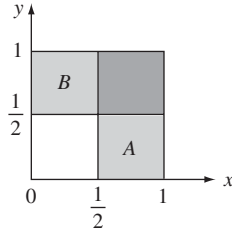
(a) Events A and B are independent.
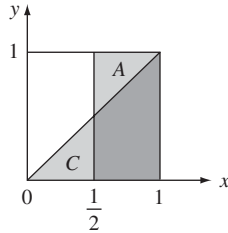
(b) Events A and C are not independent.

FIGURE 2.13
Examples of independent and nonindependent events.

In addition, knowledge of the joint occurrence of any two, say A and B, should not affect the probability of the third, that is,

$$P[C|A \cap B] = P[C].$$

In order for this to hold, we must have

$$P[C|A \cap B] = \frac{P[A \cap B \cap C]}{P[A \cap B]} = P[C].$$

This in turn implies that we must have

$$P[A \cap B \cap C] = P[A \cap B]P[C] = P[A]P[B]P[C],$$

where we have used the fact that A and B are pairwise independent. Thus we conclude that *three events A, B, and C are independent if the probability of the intersection of any pair or triplet of events is equal to the product of the probabilities of the individual events.*

The following example shows that if three events are pairwise independent, it does not necessarily follow that $P[A \cap B \cap C] = P[A]P[B]P[C]$.

Example 2.33

Consider the experiment discussed in Example 2.32 where two numbers are selected at random from the unit interval. Let the events B, D, and F be defined as follows:

$$B = \left\{ y > \frac{1}{2} \right\}, \qquad D = \left\{ x < \frac{1}{2} \right\}$$

$$F = \left\{ x < \frac{1}{2} \text{ and } y < \frac{1}{2} \right\} \cup \left\{ x > \frac{1}{2} \text{ and } y > \frac{1}{2} \right\}.$$

The three events are shown in Fig. 2.14. It can be easily verified that any pair of these events is independent:

$$P[B \cap D] = \frac{1}{4} = P[B]P[D],$$

$$P[B \cap F] = \frac{1}{4} = P[B]P[F], \text{ and}$$

$$P[D \cap F] = \frac{1}{4} = P[D]P[F].$$

However, the three events are not independent, since $B \cap D \cap F = \varnothing$, so

$$P[B \cap D \cap F] = P[\varnothing] = 0 \neq P[B]P[D]P[F] = \frac{1}{8}.$$

In order for a set of n events to be independent, the probability of an event should be unchanged when we are given the joint occurrence of any subset of the other events. This requirement naturally leads to the following definition of independence. *The events $A_1, A_2, \ldots, A_n$ are said to be **independent** if for $k = 2, \ldots, n$,*

$$P[A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}] = P[A_{i_1}]P[A_{i_2}] \ldots P[A_{i_k}], \qquad (2.33)$$

(a) $B = \{y > \frac{1}{2}\}$ (b) $D = \{x < \frac{1}{2}\}$

(c) $F = \{x < \frac{1}{2}$ and $y < \frac{1}{2}\}$ $\{x > \frac{1}{2}$ and $y > \frac{1}{2}\}$
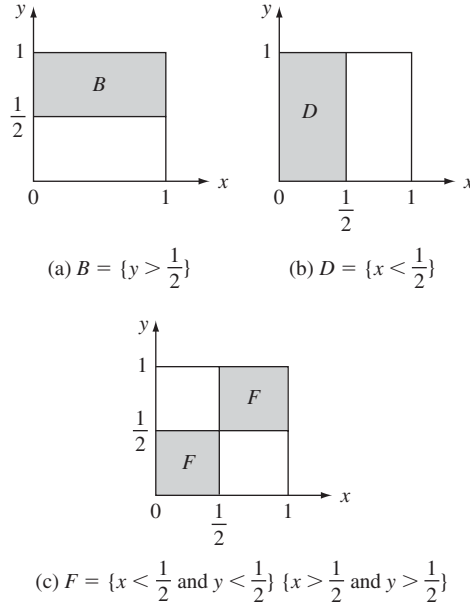
FIGURE 2.14
Events *B, D*, and *F* are pairwise independent, but the
triplet *B, D, F* are not independent events.

where $1 \leq i_1 < i_2 < \cdots < i_k \leq n$. For a set of n events we need to verify that the probabilities of all $2^n - n - 1$ possible intersections factor in the right way.

The above definition of independence appears quite cumbersome because it requires that so many conditions be verified. However, the most common application of the independence concept is in making the assumption that the events of separate experiments are independent. We refer to such experiments as **independent experiments**. For example, it is common to assume that the outcome of a coin toss is independent of the outcomes of all prior and all subsequent coin tosses.

Example 2.34

Suppose a fair coin is tossed three times and we observe the resulting sequence of heads and tails. Find the probability of the elementary events.

The sample space of this experiment is $S = \{$HHH, HHT, HTH, THH, TTH, THT, HTT, TTT$\}$. The assumption that the coin is fair means that the outcomes of a single toss are equiprobable, that is, $P[$H$] = P[$T$] = 1/2$. If we assume that the outcomes of the coin tosses are independent, then

$$P[\{\text{HHH}\}] = P[\{\text{H}\}]P[\{\text{H}\}]P[\{\text{H}\}] = \frac{1}{8},$$

$$P[\{\text{HHT}\}] = P[\{\text{H}\}]P[\{\text{H}\}]P[\{\text{T}\}] = \frac{1}{8},$$

$$P[\{HTH\}] = P[\{H\}]P[\{T\}]P[\{H\}] = \frac{1}{8},$$

$$P[\{THH\}] = P[\{T\}]P[\{H\}]P[\{H\}] = \frac{1}{8},$$

$$P[\{TTH\}] = P[\{T\}]P[\{T\}]P[\{H\}] = \frac{1}{8},$$

$$P[\{THT\}] = P[\{T\}]P[\{H\}]P[\{T\}] = \frac{1}{8},$$

$$P[\{HTT\}] = P[\{H\}]P[\{T\}]P[\{T\}] = \frac{1}{8}, \text{ and}$$

$$P[\{TTT\}] = P[\{T\}]P[\{T\}]P[\{T\}] = \frac{1}{8}.$$

Example 2.35   System Reliability

A system consists of a controller and three peripheral units. The system is said to be "up" if the controller and at least two of the peripherals are functioning. Find the probability that the system is up, assuming that all components fail independently.

Define the following events: A is "controller is functioning" and B_i is "peripheral i is functioning" where $i = 1, 2, 3$. The event F, "two or more peripheral units are functioning," occurs if all three units are functioning or if exactly two units are functioning. Thus

$$F = (B_1 \cap B_2 \cap B_3^c) \cup (B_1 \cap B_2^c \cap B_3)$$

$$\cup (B_1^c \cap B_2 \cap B_3) \cup (B_1 \cap B_2 \cap B_3).$$

Note that the events in the above union are mutually exclusive. Thus

$$P[F] = P[B_1]P[B_2]P[B_3^c] + P[B_1]P[B_2^c]P[B_3]$$

$$+ P[B_1^c]P[B_2]P[B_3] + P[B_1]P[B_2]P[B_3]$$

$$= 3(1 - a)^2 a + (1 - a)^3,$$

where we have assumed that each peripheral fails with probability a, so that $P[B_i] = 1 - a$ and $P[B_i^c] = a$.

The event "system is up" is then $A \cap F$. If we assume that the controller fails with probability p, then

$$P[\text{"system up"}] = P[A \cap F] = P[A]P[F]$$

$$= (1 - p)P[F]$$

$$= (1 - p)\{3(1 - a)^2 a + (1 - a)^3\}.$$

Let $a = 10\%$, then all three peripherals are functioning $(1 - a)^3 = 72.9\%$ of the time and two are functioning and one is "down" $3(1 - a)^2 a = 24.3\%$ of the time. Thus two or more peripherals are functioning 97.2% of the time. Suppose that the controller is not very reliable, say $p = 20\%$, then the system is up only 77.8% of the time, mostly because of controller failures.

Suppose a second identical controller with $p = 20\%$ is added to the system, and that the system is "up" if at least one of the controllers is functioning and if two or more of the peripherals are functioning. In Problem 2.94, you are asked to show that at least one of the controllers is

functioning 96% of the time, and that the system is up 93.3% of the time. This is an increase of 16% over the system with a single controller.

2.6   SEQUENTIAL EXPERIMENTS

Many random experiments can be viewed as sequential experiments that consist of a sequence of simpler subexperiments. These subexperiments may or may not be independent. In this section we discuss methods for obtaining the probabilities of events in sequential experiments.

2.6.1   Sequences of Independent Experiments

Suppose that a random experiment consists of performing experiments $E_1, E_2, \ldots, E_n$. The outcome of this experiment will then be an n-tuple $s = (s_1, \ldots, s_n)$, where s_k is the outcome of the kth subexperiment. The sample space of the sequential experiment is defined as the set that contains the above n-tuples and is denoted by the Cartesian product of the individual sample spaces $S_1 \times S_2 \times \cdots \times S_n$.

    We can usually determine, because of physical considerations, when the subexperiments are independent, in the sense that the outcome of any given subexperiment cannot affect the outcomes of the other subexperiments. Let $A_1, A_2, \ldots, A_n$ be events such that A_k concerns only the outcome of the kth subexperiment. If the subexperiments are independent, then it is reasonable to assume that the above events $A_1, A_2, \ldots, A_n$ are independent. Thus

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1]P[A_2]\ldots P[A_n]. \qquad (2.34)$$

This expression allows us to compute all probabilities of events of the sequential experiment.

Example 2.36

Suppose that 10 numbers are selected at random from the interval $[0, 1]$. Find the probability that the first 5 numbers are less than 1/4 and the last 5 numbers are greater than 1/2. Let $x_1, x_2, \ldots, x_{10}$ be the sequence of 10 numbers, then the events of interest are

$$A_k = \left\{ x_k < \frac{1}{4} \right\} \qquad \text{for } k = 1, \ldots, 5$$

$$A_k = \left\{ x_k > \frac{1}{2} \right\} \qquad \text{for } k = 6, \ldots, 10.$$

If we assume that each selection of a number is independent of the other selections, then

$$P[A_1 \cap A_2 \cap \cdots \cap A_{10}] = P[A_1]P[A_2]\ldots P[A_{10}]$$

$$= \left(\frac{1}{4} \right)^5 \left(\frac{1}{2} \right)^5.$$

    We will now derive several important models for experiments that consist of sequences of independent subexperiments.

2.6.2 The Binomial Probability Law

A **Bernoulli trial** involves performing an experiment once and noting whether a particular event A occurs. The outcome of the Bernoulli trial is said to be a "success" if A occurs and a "failure" otherwise. In this section we are interested in finding the probability of k successes in n independent repetitions of a Bernoulli trial.

We can view the outcome of a single Bernoulli trial as the outcome of a toss of a coin for which the probability of heads (success) is $p = P[A]$. The probability of k successes in n Bernoulli trials is then equal to the probability of k heads in n tosses of the coin.

Example 2.37

Suppose that a coin is tossed three times. If we assume that the *tosses are independent* and the probability of heads is p, then the probability for the sequences of heads and tails is

$$P[\{HHH\}] = P[\{H\}]P[\{H\}]P[\{H\}] = p^3,$$

$$P[\{HHT\}] = P[\{H\}]P[\{H\}]P[\{T\}] = p^2(1 - p),$$

$$P[\{HTH\}] = P[\{H\}]P[\{T\}]P[\{H\}] = p^2(1 - p),$$

$$P[\{THH\}] = P[\{T\}]P[\{H\}]P[\{H\}] = p^2(1 - p),$$

$$P[\{TTH\}] = P[\{T\}]P[\{T\}]P[\{H\}] = p(1 - p)^2,$$

$$P[\{THT\}] = P[\{T\}]P[\{H\}]P[\{T\}] = p(1 - p)^2,$$

$$P[\{HTT\}] = P[\{H\}]P[\{T\}]P[\{T\}] = p(1 - p)^2, \text{ and}$$

$$P[\{TTT\}] = P[\{T\}]P[\{T\}]P[\{T\}] = (1 - p)^3$$

where we used the fact that the tosses are independent. Let k be the number of heads in three trials, then

$$P[k = 0] = P[\{TTT\}] = (1 - p)^3,$$

$$P[k = 1] = P[\{TTH, THT, HTT\}] = 3p(1 - p)^2,$$

$$P[k = 2] = P[\{HHT, HTH, THH\}] = 3p^2(1 - p), \text{ and}$$

$$P[k = 3] = P[\{HHH\}] = p^3.$$

The result in Example 2.37 is the $n = 3$ case of the binomial probability law.

Theorem

Let k be the number of successes in n independent Bernoulli trials, then the probabilities of k are given by the **binomial probability law**:

$$p_n(k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for} \quad k = 0, \ldots, n, \quad (2.35)$$

where $p_n(k)$ is the probability of k successes in n trials, and

$$\binom{n}{k} = \frac{n!}{k! \, (n-k)!} \tag{2.36}$$

is the binomial coefficient.

The term $n!$ in Eq. (2.36) is called n factorial and is defined by $n! = n(n-1)\ldots$ $(2)(1)$. By definition $0!$ is equal to 1.

We now prove the above theorem. Following Example 2.34 we see that each of the sequences with k successes and $n - k$ failures has the same probability, namely $p^k(1-p)^{n-k}$. Let $N_n(k)$ be the number of distinct sequences that have k successes and $n - k$ failures, then

$$p_n(k) = N_n(k)p^k(1-p)^{n-k}. \tag{2.37}$$

The expression $N_n(k)$ is the number of ways of picking k positions out of n for the successes. It can be shown that[5]

$$N_n(k) = \binom{n}{k}. \tag{2.38}$$

The theorem follows by substituting Eq. (2.38) into Eq. (2.37).

Example 2.38

Verify that Eq. (2.35) gives the probabilities found in Example 2.37.
In Example 2.37, let "toss results in heads" correspond to a "success," then

$$p_3(0) = \frac{3!}{0! \, 3!} p^0(1-p)^3 = (1-p)^3,$$

$$p_3(1) = \frac{3!}{1! \, 2!} p^1(1-p)^2 = 3p(1-p)^2,$$

$$p_3(2) = \frac{3!}{2! \, 1!} p^2(1-p)^1 = 3p^2(1-p), \text{ and}$$

$$p_3(3) = \frac{3!}{0! \, 3!} p^3(1-p)^0 = p^3,$$

which are in agreement with our previous results.

You were introduced to the binomial coefficient in an introductory calculus course when the **binomial theorem** was discussed:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{2.39a}$$

[5]See Example 2.21.

If we let $a = b = 1$, then

$$2^n = \sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} N_n(k),$$

which is in agreement with the fact that there are 2^n distinct possible sequences of successes and failures in n trials. If we let $a = p$ and $b = 1 - p$ in Eq. (2.39a), we then obtain

$$1 = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = \sum_{k=0}^{n} p_n(k), \tag{2.39b}$$

which confirms that the probabilities of the binomial probabilities sum to 1.

The term $n!$ grows very quickly with n, so numerical problems are encountered for relatively small values of n if one attempts to compute $p_n(k)$ directly using Eq. (2.35). The following recursive formula avoids the direct evaluation of $n!$ and thus extends the range of n for which $p_n(k)$ can be computed before encountering numerical difficulties:

$$p_n(k + 1) = \frac{(n - k)p}{(k + 1)(1 - p)} p_n(k). \tag{2.40}$$

Later in the book, we present two approximations for the binomial probabilities for the case when n is large.

Example 2.39

Let k be the number of active (nonsilent) speakers in a group of eight noninteracting (i.e., independent) speakers. Suppose that a speaker is active with probability 1/3. Find the probability that the number of active speakers is greater than six.

For $i = 1, \dots, 8$, let A_i denote the event "ith speaker is active." The number of active speakers is then the number of successes in eight Bernoulli trials with $p = 1/3$. Thus the probability that more than six speakers are active is

$$P[k = 7] + P[k = 8] = \binom{8}{7}\left(\frac{1}{3}\right)^7\left(\frac{2}{3}\right) + \binom{8}{8}\left(\frac{1}{3}\right)^8$$

$$= .00244 + .00015 = .00259.$$

Example 2.40   Error Correction Coding

A communication system transmits binary information over a channel that introduces random bit errors with probability $\varepsilon = 10^{-3}$. The transmitter transmits each information bit three times, and a decoder takes a majority vote of the received bits to decide on what the transmitted bit was. Find the probability that the receiver will make an incorrect decision.

The receiver can correct a single error, but it will make the wrong decision if the channel introduces two or more errors. If we view each transmission as a Bernoulli trial in which a "success" corresponds to the introduction of an error, then the probability of two or more errors in three Bernoulli trials is

$$P[k \geq 2] = \binom{3}{2}(.001)^2(.999) + \binom{3}{3}(.001)^3 \simeq 3(10^{-6}).$$

2.6.3   The Multinomial Probability Law

The binomial probability law can be generalized to the case where we note the occurrence of more than one event. Let $B_1, B_2, \ldots, B_M$ be a partition of the sample space S of some random experiment and let $P[B_j] = p_j$. The events are mutually exclusive, so

$$p_1 + p_2 + \cdots + p_M = 1.$$

Suppose that n independent repetitions of the experiment are performed. Let k_j be the number of times event B_j occurs, then the vector $(k_1, k_2, \ldots, k_M)$ specifies the number of times each of the events B_j occurs. The probability of the vector $(k_1, \ldots, k_M)$ satisfies the **multinomial probability law**:

$$P[(k_1, k_2, \ldots, k_M)] = \frac{n!}{k_1!\, k_2! \ldots k_M!} p_1^{k_1} p_2^{k_2} \ldots p_M^{k_M}, \tag{2.41}$$

where $k_1 + k_2 + \cdots + k_M = n$. The binomial probability law is the $M = 2$ case of the multinomial probability law. The derivation of the multinomial probabilities is identical to that of the binomial probabilities. We only need to note that the number of different sequences with $k_1, k_2, \ldots, k_M$ instances of the events $B_1, B_2, \ldots, B_M$ is given by the multinomial coefficient in Eq. (2.26).

Example 2.41

A dart is thrown nine times at a target consisting of three areas. Each throw has a probability of .2, .3, and .5 of landing in areas 1, 2, and 3, respectively. Find the probability that the dart lands exactly three times in each of the areas.

This experiment consists of nine independent repetitions of a subexperiment that has three possible outcomes. The probability for the number of occurrences of each outcome is given by the multinomial probabilities with parameters $n = 9$ and $p_1 = .2$, $p_2 = .3$, and $p_3 = .5$:

$$P[(3, 3, 3)] = \frac{9!}{3!\, 3!\, 3!} (.2)^3 (.3)^3 (.5)^3 = .04536.$$

Example 2.42

Suppose we pick 10 telephone numbers at random from a telephone book and note the last digit in each of the numbers. What is the probability that we obtain each of the integers from 0 to 9 only once?

The probabilities for the number of occurrences of the integers is given by the multinomial probabilities with parameters $M = 10$, $n = 10$, and $p_j = 1/10$ if we assume that the 10 integers in the range 0 to 9 are equiprobable. The probability of obtaining each integer once in 10 draws is then

$$\frac{10!}{1!\, 1! \ldots 1!} (.1)^{10} \simeq 3.6(10^{-4}).$$

2.6.4   The Geometric Probability Law

Consider a sequential experiment in which we repeat independent Bernoulli trials until the occurrence of the first success. Let the outcome of this experiment be m, the number of trials carried out until the occurrence of the first success. The sample space

for this experiment is the set of positive integers. The probability, $p(m)$, that m trials are required is found by noting that this can only happen if the first $m - 1$ trials result in failures and the mth trial in success.[6] The probability of this event is

$$p(m) = P[A_1^c A_2^c \ldots A_{m-1}^c A_m] = (1 - p)^{m-1} p \qquad m = 1, 2, \ldots, \qquad (2.42\text{a})$$

where A_i is the event "success in ith trial." The probability assignment specified by Eq. (2.42a) is called the **geometric probability law**.

The probabilities in Eq. (2.42a) sum to 1:

$$\sum_{m=1}^{\infty} p(m) = p \sum_{m=1}^{\infty} q^{m-1} = p \frac{1}{1 - q} = 1, \qquad (2.42\text{b})$$

where $q = 1 - p$, and where we have used the formula for the summation of a geometric series. The probability that more than K trials are required before a success occurs has a simple form:

$$P[\{m > K\}] = p \sum_{m=K+1}^{\infty} q^{m-1} = pq^K \sum_{j=0}^{\infty} q^j$$

$$= pq^K \frac{1}{1 - q}$$

$$= q^K. \qquad (2.43)$$

Example 2.43   Error Control by Retransmission

Computer A sends a message to computer B over an unreliable radio link. The message is encoded so that B can detect when errors have been introduced into the message during transmission. If B detects an error, it requests A to retransmit it. If the probability of a message transmission error is $q = .1$, what is the probability that a message needs to be transmitted more than two times?

Each transmission of a message is a Bernoulli trial with probability of success $p = 1 - q$. The Bernoulli trials are repeated until the first success (error-free transmission). The probability that more than two transmissions are required is given by Eq. (2.43):

$$P[m > 2] = q^2 = 10^{-2}.$$

2.6.5   Sequences of Dependent Experiments

In this section we consider a sequence or "chain" of subexperiments in which the outcome of a given subexperiment determines which subexperiment is performed next. We first give a simple example of such an experiment and show how diagrams can be used to specify the sample space.

Example 2.44

A sequential experiment involves repeatedly drawing a ball from one of two urns, noting the number on the ball, and replacing the ball in its urn. Urn 0 contains a ball with the number 1 and two balls with the number 0, and urn 1 contains five balls with the number 1 and one ball

---

[6]See Example 2.11 in Section 2.2 for a relative frequency interpretation of how the geometric probability law comes about.

with the number 0. The urn from which the first draw is made is selected at random by flipping a fair coin. Urn 0 is used if the outcome is heads and urn 1 if the outcome is tails. Thereafter the urn used in a subexperiment corresponds to the number on the ball selected in the previous subexperiment.

The sample space of this experiment consists of sequences of 0s and 1s. Each possible sequence corresponds to a path through the "trellis" diagram shown in Fig. 2.15(a). The nodes in the diagram denote the urn used in the nth subexperiment, and the labels in the branches denote the outcome of a subexperiment. Thus the path 0011 corresponds to the sequence: The coin toss was heads so the first draw was from urn 0; the outcome of the first draw was 0, so the second draw was from urn 0; the outcome of the second draw was 1, so the third draw was from urn 1; and the outcome from the third draw was 1, so the fourth draw is from urn 1.

Now suppose that we want to compute the probability of a particular sequence of outcomes, say s_0, s_1, s_2. Denote this probability by $P[\{s_0\} \cap \{s_1\} \cap \{s_2\}]$. Let $A = \{s_2\}$ and $B = \{s_0\} \cap \{s_1\}$, then since $P[A \cap B] = P[A|B]P[B]$ we have

$$P[\{s_0\} \cap \{s_1\} \cap \{s_2\}] = P[\{s_2\}|\{s_0\} \cap \{s_1\}]P[\{s_0\} \cap \{s_1\}]$$
$$= P[\{s_2\}|\{s_0\} \cap \{s_1\}]P[\{s_1\}|\{s_0\}]P[\{s_0\}]. \qquad (2.44)$$

Now note that in the above urn example the probability $P[\{s_n\}|\{s_0\} \cap \cdots \cap \{s_{n-1}\}]$ depends only on $\{s_{n-1}\}$ since the most recent outcome determines which subexperiment is performed:

$$P[\{s_n\}|\{s_0\} \cap \cdots \cap \{s_{n-1}\}] = P[\{s_n\}|\{s_{n-1}\}]. \qquad (2.45)$$
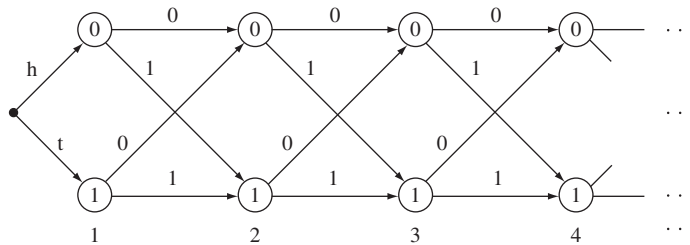
(a) Each sequence of outcomes corresponds
to a path through this trellis diagram.
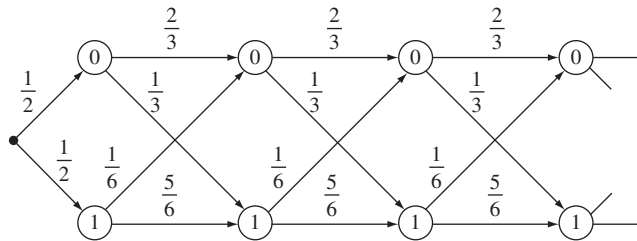
(b) The probability of a sequence of outcomes is the
product of the probabilities along the associated path.

FIGURE 2.15
Trellis diagram for a Markov chain.

Therefore for the sequence of interest we have that

$$P[\{s_0\} \cap \{s_1\} \cap \{s_2\}] = P[\{s_2\}|\{s_1\}]P[\{s_1\}|\{s_0\}]P[\{s_0\}]. \qquad (2.46)$$

Sequential experiments that satisfy Eq. (2.45) are called **Markov chains**. For these experiments, the probability of a sequence $s_0, s_1, \ldots, s_n$ is given by

$$P[s_0, s_1, \ldots, s_n] = P[s_n|s_{n-1}]P[s_{n-1}|s_{n-2}] \ldots P[s_1|s_0]P[s_0] \qquad (2.47)$$

where we have simplified notation by omitting braces. Thus the probability of the sequence $s_0, \ldots, s_n$ is given by the product of the probability of the first outcome s_0 and the probabilities of all subsequent transitions, s_0 to s_1, s_1 to s_2, and so on. Chapter 11 deals with Markov chains.

Example 2.45

Find the probability of the sequence 0011 for the urn experiment introduced in Example 2.44.
   Recall that urn 0 contains two balls with label 0 and one ball with label 1, and that urn 1 contains five balls with label 1 and one ball with label 0. We can readily compute the probabilities of sequences of outcomes by labeling the branches in the trellis diagram with the probability of the corresponding transition as shown in Fig. 2.15(b). Thus the probability of the sequence 0011 is given by

$$P[0011] = P[1|1]P[1|0]P[0|0]P[0],$$

where the transition probabilities are given by

$$P[1|0] = \frac{1}{3} \quad \text{and} \quad P[0|0] = \frac{2}{3}$$

$$P[1|1] = \frac{5}{6} \quad \text{and} \quad P[0|1] = \frac{1}{6},$$

and the initial probabilities are given by

$$P(0) = \frac{1}{2} = P[1].$$

If we substitute these values into the expression for $P[0011]$, we obtain

$$P[0011] = \left(\frac{5}{6}\right)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = \frac{5}{54}.$$

The two-urn experiment in Examples 2.44 and 2.45 is the simplest example of the Markov chain models that are discussed in Chapter 11. The two-urn experiment discussed here is used to model situations in which there are only two outcomes, and in which the outcomes tend to occur in bursts. For example, the two-urn model has been used to model the "bursty" behavior of the voice packets generated by a single speaker where bursts of active packets are separated by relatively long periods of silence. The model has also been used for the sequence of black and white dots that result from scanning a black and white image line by line.

This section introduces the basic method for generating sequences of "random" numbers using a computer. Any computer simulation of a system that involves randomness must include a method for generating sequences of random numbers. These random numbers must satisfy long-term average properties of the processes they are simulating. In this section we focus on the problem of generating random numbers that are "uniformly distributed" in the interval [0, 1]. In the next chapter we will show how these random numbers can be used to generate numbers with arbitrary probability laws.

The first problem we must confront in generating a random number in the interval [0, 1] is the fact that there are an uncountably infinite number of points in the interval, but the computer is limited to representing numbers with finite precision only. We must therefore be content with generating equiprobable numbers from some finite set, say $\{0, 1, \ldots, M - 1\}$ or $\{1, 2, \ldots, M\}$. By dividing these numbers by M, we obtain numbers in the unit interval. These numbers can be made increasingly dense in the unit interval by making M very large.

The next step involves finding a mechanism for generating random numbers. The direct approach involves performing random experiments. For example, we can generate integers in the range 0 to $2^m - 1$ by flipping a fair coin m times and replacing the sequence of heads and tails by 0s and 1s to obtain the binary representation of an integer. Another example would involve drawing a ball from an urn containing balls numbered 1 to M. Computer simulations involve the generation of long sequences of random numbers. If we were to use the above mechanisms to generate random numbers, we would have to perform the experiments a large number of times and store the outcomes in computer storage for access by the simulation program. It is clear that this approach is cumbersome and quickly becomes impractical.

2.7.1   Pseudo-Random Number Generation

The preferred approach for the computer generation of random numbers involves the use of recursive formulas that can be implemented easily and quickly. These **pseudo-random number generators** produce a sequence of numbers that appear to be random but that in fact repeat after a very long period. The currently preferred pseudo-random number generator is the so-called Mersenne Twister, which is based on a matrix linear recurrence over a binary field. This algorithm can yield sequences with an extremely long period of $2^{19937} - 1$. The Mersenne Twister generates 32-bit integers, so $M = 2^{32} - 1$ in terms of our previous discussion. We obtain a sequence of numbers in the unit interval by dividing the 32-bit integers by 2^{32}. The sequence of such numbers should be equally distributed over unit cubes of very high dimensionality. The Mersenne Twister has been shown to meet this condition up to 632-dimensionality. In addition, the algorithm is fast and efficient in terms of storage.

Software implementations of the Mersenne Twister are widely available and incorporated into numerical packages such as MATLAB® and Octave.[7] Both MATLAB and Octave provide a means to generate random numbers from the unit interval using the

---

[7]MATLAB® and Octave are interactive computer programs for numerical computations involving matrices. MATLAB® is a commercial product sold by The Mathworks, Inc. Octave is a free, open-source program that is mostly compatible with MATLAB in terms of computation. Long [9] provides an introduction to Octave.

rand command. The rand (n, m) operator returns an *n* row by *m* column matrix with elements that are random numbers from the interval $[0, 1)$. This operator is the starting point for generating all types of random numbers.

Example 2.46   Generation of Numbers from the Unit Interval

First, generate 6 numbers from the unit interval. Next, generate 10,000 numbers from the unit interval. Plot the histogram and empirical distribution function for the sequence of 10,000 numbers.

　　　The following command results in the generation of six numbers from the unit interval.

```
>rand(1,6)
ans =
Columns 1 through 6:
0.642667 0.147811 0.317465 0.512824 0.710823 0.406724
```

　　　The following set of commands will generate 10000 numbers and produce the histogram shown in Fig. 2.16.

```
>X-rand(10000,1);          % Return result in a 10,000-element column vector X.

>K=0.005:0.01;0.995;       % Produce column vector K consisting of the mid points
                           % for 100 bins of width 0.01 in the unit interval.

>Hist(X,K)                 % Produce the desired histogram in Fig 2.16.

>plot(K,empirical_cdf(K,X))  % Plot the proportion of elements in the array X less
                             % than or equal to k, where k is an element of K.
```

The empirical cdf is shown in Fig. 2.17. It is evident that the array of random numbers is uniformly distributed in the unit interval.
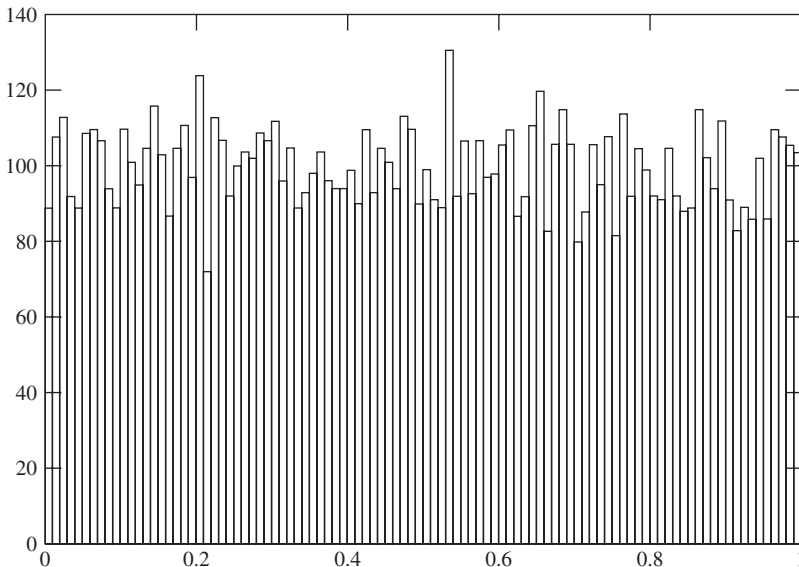
FIGURE 2.16
Histogram resulting from experiment to generate 10,000 numbers in the unit interval.
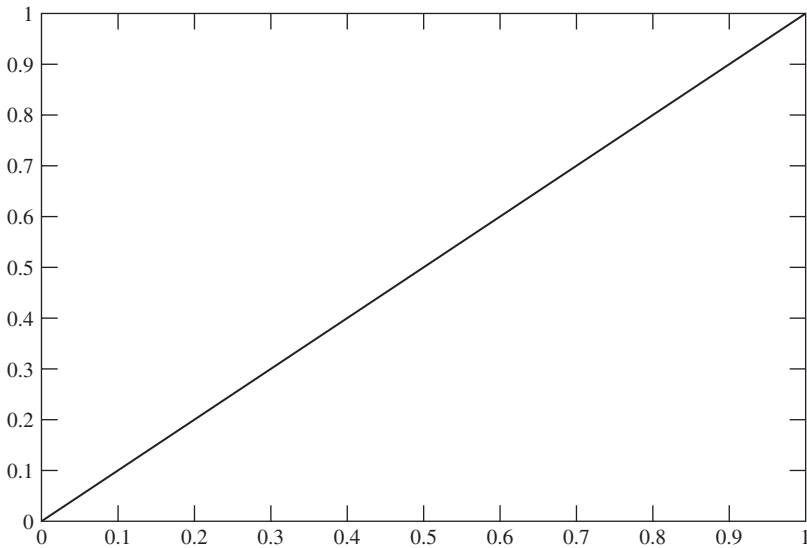
FIGURE 2.17
Empirical cdf of experiment that generates 10,000 numbers.

2.7.2   Simulation of Random Experiments

MATLAB® and Octave provide functions that are very useful in carrying out numerical evaluation of probabilities involving the most common distributions. Functions are also provided for the generation of random numbers with specific probability distributions. In this section we consider Bernoulli trials and binomial distributions. In Chapter 3 we consider experiments with discrete sample spaces.

Example 2.47   Bernoulli Trials and Binomial Probabilities

First, generate the outcomes of eight Bernoulli trials. Next, generate the outcomes of 100 repetitions of a random experiment that counts the number of successes in 16 Bernoulli trials with probability of success $1/2$. Plot the histogram of the outcomes in the 100 experiments and compare to the binomial probabilities with $n = 16$ and $p = 1/2$.

The following command will generate the outcomes of eight Bernoulli trials, as shown by the answer that follows.

```
>X=rand(1,8)<0.5;                        % Generate 1 row of Bernoulli trials with p = 0.5
X =
 0 1 1 0 0 0 1 1
```

If the number produced by rand for a given Bernoulli trial is less than $p = 0.5$, then the outcome of the Bernoulli trial is 1.

Next we show the set of commands to generate the outcomes of 100 repetitions of random experiments where each involves 16 Bernoulli trials.

```
>X=rand(100,16)<0.5;        % Generate 100 rows of 16 Bernoulli trials with
                            % p = 0.5.

>Y=sum(X,2);                % Add the results of each row to obtain the number of
                            % successes in each experiment. Y contains the 100
                            % outcomes.

>K=0:16;

>Z=empirical_pdf(K,Y));     % Find the relative frequencies of the outcomes in Y.

>Bar(K,Z)                   % Produce a bar graph of the relative frequencies.

>hold on                    % Retains the graph for next command.

>stem(K,binomial_pdf(K,16,0.5))   % Plot the binomial probabilities along
                                  % with the corresponding relative frequencies.
```

Figure 2.18 shows that there is good agreement between the relative frequencies and the binomial probabilities.

---

*2.8   FINE POINTS: EVENT CLASSES[8]

If the sample space S is discrete, then the event class can consist of all subsets of S. There are situations where we may wish or are compelled to let the event class $\mathcal{F}$ be a smaller class of subsets of S. In these situations, only the subsets that belong to this class are considered events. In this section we explain how these situations arise.

Let $\mathcal{C}$ be the class of events of interest in a random experiment. It is reasonable to expect that any set operation on events in $\mathcal{C}$ will produce a set that is also an event in $\mathcal{C}$. We can then ask any question regarding events of the random experiment, express it using set operations, and obtain an event that is in $\mathcal{C}$. Mathematically, we require that $\mathcal{C}$ be a field.

A collection of sets $\mathcal{F}$ is called a **field** if it satisfies the following conditions:

  (i) $\varnothing \in \mathcal{F}$ (2.48a)

 (ii) if $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$ (2.48b)

(iii) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.   (2.48c)

Using DeMorgan's rule we can show that (ii) and (iii) imply that if $A \in \mathcal{F}$ and $B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$. Conditions (ii) and (iii) then imply that any finite union or intersection of events in $\mathcal{F}$ will result in an event that is also in $\mathcal{F}$.

Example 2.48

Let $S = \{T, H\}$. Find the field generated by set operations on the class consisting of elementary events of $S: \mathcal{C} = \{\{H\}, \{T\}\}$.

---

[8]The "Fine Points" sections elaborate on concepts and distinctions that are not required in an introductory course. The material in these sections is not necessarily more mathematical, but rather is not usually covered in a first course in probability.

7. W. Feller, *An Introduction to Probability Theory and Its Applications*, 3d ed., Wiley, New York, 1968.
8. A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis*, Dover Publications, New York, 1970.
9. P. J. G. Long, "Introduction to Octave," University of Cambridge, September 2005, available online.
10. A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, New York, 2000.

PROBLEMS

Section 2.1: Specifying Random Experiments

2.1. The (loose) minute hand in a clock is spun hard and the hour at which the hand comes to rest is noted.

    (a) What is the sample space?

    (b) Find the sets corresponding to the events: A = "hand is in first 4 hours"; B = "hand is between 2nd and 8th hours inclusive"; and D = "hand is in an odd hour."

    (c) Find the events: $A \cap B \cap D$, $A^c \cap B$, $A \cup (B \cap D^c)$, $(A \cup B) \cap D^c$.

2.2. A die is tossed twice and the number of dots facing up in each toss is counted and noted in the order of occurrence.

    (a) Find the sample space.

    (b) Find the set A corresponding to the event "number of dots in first toss is not less than number of dots in second toss."

    (c) Find the set B corresponding to the event "number of dots in first toss is 6."

    (d) Does A imply B or does B imply A?

    (e) Find $A \cap B^c$ and describe this event in words.

    (f) Let C correspond to the event "number of dots in dice differs by 2." Find $A \cap C$.

2.3. Two dice are tossed and the magnitude of the difference in the number of dots facing up in the two dice is noted.

    (a) Find the sample space.

    (b) Find the set A corresponding to the event "magnitude of difference is 3."

    (c) Express each of the elementary events in this experiment as the union of elementary events from Problem 2.2.

2.4. A binary communication system transmits a signal X that is either a $+2$ voltage signal or a -2 voltage signal. A malicious channel reduces the magnitude of the received signal by the number of heads it counts in two tosses of a coin. Let Y be the resulting signal.

    (a) Find the sample space.

    (b) Find the set of outcomes corresponding to the event "transmitted signal was definitely $+2$."

    (c) Describe in words the event corresponding to the outcome $Y = 0$.

2.5. A desk drawer contains six pens, four of which are dry.

    (a) The pens are selected at random one by one until a good pen is found. The sequence of test results is noted. What is the sample space?

(b) Suppose that only the number, and not the sequence, of pens tested in part a is noted. Specify the sample space.

(c) Suppose that the pens are selected one by one and tested until both good pens have been identified, and the sequence of test results is noted. What is the sample space?

(d) Specify the sample space in part c if only the number of pens tested is noted.

2.6. Three friends (Al, Bob, and Chris) put their names in a hat and each draws a name from the hat. (Assume Al picks first, then Bob, then Chris.)

(a) Find the sample space.

(b) Find the sets A, B, and C that correspond to the events "Al draws his name," "Bob draws his name," and "Chris draws his name."

(c) Find the set corresponding to the event, "no one draws his own name."

(d) Find the set corresponding to the event, "everyone draws his own name."

(e) Find the set corresponding to the event, "one or more draws his own name."

2.7. Let M be the number of message transmissions in Experiment E_6.

(a) What is the set A corresponding to the event "M is even"?

(b) What is the set B corresponding to the event "M is a multiple of 3"?

(c) What is the set C corresponding to the event "6 or fewer transmissions are required"?

(d) Find the sets $A \cap B$, $A - B$, $A \cap B \cap C$ and describe the corresponding events in words.

2.8. A number U is selected at random from the unit interval. Let the events A and B be: $A =$ "U differs from 1/2 by more than 1/4" and $B =$ "$1 - U$ is less than 1/2." Find the events $A \cap B$, $A^c \cap B$, $A \cup B$.

2.9. The sample space of an experiment is the real line. Let the events A and B correspond to the following subsets of the real line: $A = (-\infty, r]$ and $B = (-\infty, s]$, where $r \leq s$. Find an expression for the event $C = (r, s]$ in terms of A and B. Show that $B = A \cup C$ and $A \cap C = \varnothing$.

2.10. Use Venn diagrams to verify the set identities given in Eqs. (2.2) and (2.3). You will need to use different colors or different shadings to denote the various regions clearly.

2.11. Show that:

(a) If event A implies B, and B implies C, then A implies C.

(b) If event A implies B, then B^c implies A^c.

2.12. Show that if $A \cup B = A$ and $A \cap B = A$ then $A = B$.

2.13. Let A and B be events. Find an expression for the event "exactly one of the events A and B occurs." Draw a Venn diagram for this event.

2.14. Let A, B, and C be events. Find expressions for the following events:

(a) Exactly one of the three events occurs.

(b) Exactly two of the events occur.

(c) One or more of the events occur.

(d) Two or more of the events occur.

(e) None of the events occur.

2.15. Figure P2.1 shows three systems of three components, C_1, C_2, and C_3. Figure P2.1(a) is a "series" system in which the system is functioning only if all three components are functioning. Figure 2.1(b) is a "parallel" system in which the system is functioning as long as at least one of the three components is functioning. Figure 2.1(c) is a "two-out-of-three"

system in which the system is functioning as long as at least two components are functioning. Let A_k be the event "component k is functioning." For each of the three system configurations, express the event "system is functioning" in terms of the events A_k.

(a) Series system   (b) Parallel system   (c) Two-out-of-three system
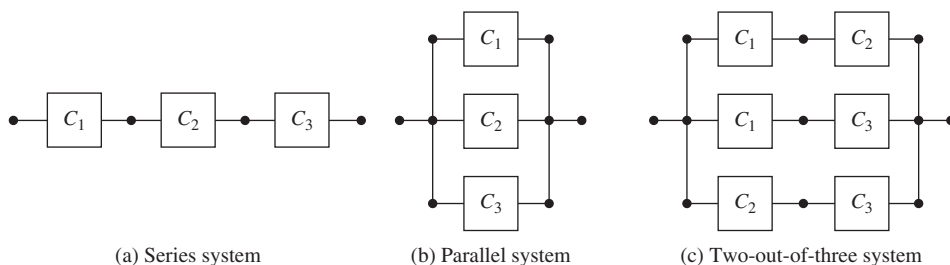
FIGURE P2.1

2.16. A system has two key subsystems. The system is "up" if both of its subsystems are functioning. Triple redundant systems are configured to provide high reliability. The overall system is operational as long as one of three systems is "up." Let A_{jk} correspond to the event "unit k in system j is functioning," for $j = 1, 2, 3$ and $k = 1, 2$.

(a) Write an expression for the event "overall system is up."

(b) Explain why the above problem is equivalent to the problem of having a connection in the network of switches shown in Fig. P2.2.
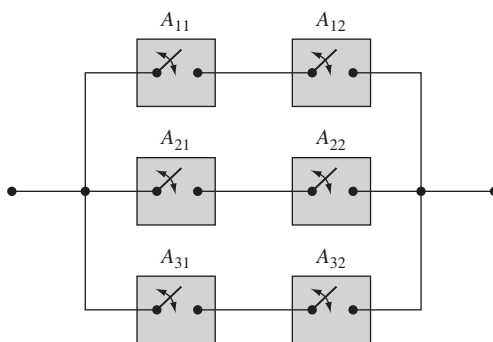
FIGURE P2.2

2.17. In a specified 6-AM-to-6-AM 24-hour period, a student wakes up at time t_1 and goes to sleep at some later time t_2.

(a) Find the sample space and sketch it on the x-y plane if the outcome of this experiment consists of the pair (t_1, t_2).

(b) Specify the set A and sketch the region on the plane corresponding to the event "student is asleep at noon."

(c) Specify the set B and sketch the region on the plane corresponding to the event "student sleeps through breakfast (7–9 AM)."

(d) Sketch the region corresponding to $A \cap B$ and describe the corresponding event in words.

2.18. A road crosses a railroad track at the top of a steep hill. The train cannot stop for oncoming cars and cars, cannot see the train until it is too late. Suppose a train begins crossing the road at time t_1 and that the car begins crossing the track at time t_2, where $0 < t_1 < T$ and $0 < t_2 < T$.

(a) Find the sample space of this experiment.

(b) Suppose that it takes the train d_1 seconds to cross the road and it takes the car d_2 seconds to cross the track. Find the set that corresponds to a collision taking place.

(c) Find the set that corresponds to a collision is missed by 1 second or less.

2.19. A random experiment has sample space $S = \{-1, 0, +1\}$.

(a) Find all the subsets of S.

(b) The outcome of a random experiment consists of pairs of outcomes from S where the elements of the pair cannot be equal. Find the sample space S' of this experiment. How many subsets does S' have?

2.20. (a) A coin is tossed twice and the sequence of heads and tails is noted. Let S be the sample space of this experiment. Find all subsets of S.

(b) A coin is tossed twice and the number of heads is noted. Let S? be the sample space of this experiment. Find all subsets of S'.

(c) Consider parts a and b if the coin is tossed 10 times. How many subsets do S and S' have? How many bits are needed to assign a binary number to each possible subset?

Section 2.2: The Axioms of Probability

2.21. A die is tossed and the number of dots facing up is noted.

(a) Find the probability of the elementary events under the assumption that all faces of the die are equally likely to be facing up after a toss.

(b) Find the probability of the events: $A = \{\text{more than 3 dots}\}$; $B = \{\text{odd number of dots}\}$.

(c) Find the probability of $A \cup B$, $A \cap B$, A^c.

2.22. In Problem 2.2, a die is tossed twice and the number of dots facing up in each toss is counted and noted in the order of occurrence.

(a) Find the probabilities of the elementary events.

(b) Find the probabilities of events A, B, C, $A \cap B^c$, and $A \cap C$ defined in Problem 2.2.

2.23. A random experiment has sample space $S = \{a, b, c, d\}$. Suppose that $P[\{c, d\}] = 3/8$, $P[\{b, c\}] = 6/8$, and $P[\{d\}] = 1/8$, $P[\{c, d\}] = 3/8$. Use the axioms of probability to find the probabilities of the elementary events.

2.24. Find the probabilities of the following events in terms of $P[A]$, $P[B]$, and $P[A \cap B]$:

(a) A occurs and B does not occur; B occurs and A does not occur.

(b) Exactly one of A or B occurs.

(c) Neither A nor B occur.

2.25. Let the events A and B have $P[A] = x$, $P[B] = y$, and $P[A \cup B] = z$. Use Venn diagrams to find $P[A \cap B]$, $P[A^c \cap B^c]$, $P[A^c \cup B^c]$, $P[A \cap B^c]$, $P[A^c \cup B]$.

2.26. Show that
$$P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[A \cap B] - P[A \cap C] - P[B \cap C]$$
$$+ P[A \cap B \cap C].$$

2.27. Use the argument from Problem 2.26 to prove Corollary 6 by induction.

2.28. A hexadecimal character consists of a group of three bits. Let A_i be the event "ith bit in a character is a 1."

    (a) Find the probabilities for the following events: $A_1, A_1 \cap A_3, A_1 \cap A_2 \cap A_3$ and $A_1 \cup A_2 \cup A_3$. Assume that the values of bits are determined by tosses of a fair coin.

    (b) Repeat part a if the coin is biased.

2.29. Let M be the number of message transmissions in Problem 2.7. Find the probabilities of the events $A, B, C, C^c, A \cap B, A - B, A \cap B \cap C$. Assume the probability of successful transmission is 1/2.

2.30. Use Corollary 7 to prove the following:

    (a) $P[A \cup B \cup C] \le P[A] + P[B] + P[C]$.

    (b) $P\left[\bigcup_{k=1}^{n} A_k\right] \le \sum_{k=1}^{n} P[A_k]$.

    (c) $P\left[\bigcap_{k=1}^{n} A_k\right] \ge 1 - \sum_{k=1}^{n} P[A_k^c]$.

The second expression is called the **union bound**.

2.31. Let p be the probability that a single character appears incorrectly in this book. Use the union bound for the probability of there being any errors in a page with n characters.

2.32. A die is tossed and the number of dots facing up is noted.

    (a) Find the probability of the elementary events if faces with an even number of dots are twice as likely to come up as faces with an odd number.

    (b) Repeat parts b and c of Problem 2.21.

2.33. Consider Problem 2.1 where the minute hand in a clock is spun. Suppose that we now note the *minute* at which the hand comes to rest.

    (a) Suppose that the minute hand is very loose so the hand is equally likely to come to rest anywhere in the clock. What are the probabilities of the elementary events?

    (b) Now suppose that the minute hand is somewhat sticky and so the hand is 1/2 as likely to land in the second minute than in the first, 1/3 as likely to land in the third minute as in the first, and so on. What are the probabilities of the elementary events?

    (c) Now suppose that the minute hand is very sticky and so the hand is 1/2 as likely to land in the second minute than in the first, 1/2 as likely to land in the third minute as in the second, and so on. What are the probabilities of the elementary events?

    (d) Compare the probabilities that the hand lands in the last minute in parts a, b, and c.

2.34. A number x is selected at random in the interval $[-1, 2]$. Let the events $A = \{x < 0\}$, $B = \{|x - 0.5| < 0.5\}$, and $C = \{x > 0.75\}$.

    (a) Find the probabilities of A, B, $A \cap B$, and $A \cap C$.

    (b) Find the probabilities of $A \cup B$, $A \cup C$, and $A \cup B \cup C$, first, by directly evaluating the sets and then their probabilities, and second, by using the appropriate axioms or corollaries.

2.35. A number x is selected at random in the interval $[-1, 2]$. Numbers from the subinterval $[0, 2]$ occur half as frequently as those from $[-1, 0)$.

    (a) Find the probability assignment for an interval completely within $[-1, 0)$; completely within $[0, 2]$; and partly in each of the above intervals.

    (b) Repeat Problem 2.34 with this probability assignment.

2.36. The lifetime of a device behaves according to the probability law $P[(t, \infty)] = 1/t$ for $t > 1$. Let A be the event "lifetime is greater than 4," and B the event "lifetime is greater than 8."

   (a) Find the probability of $A \cap B$, and $A \cup B$.

   (b) Find the probability of the event "lifetime is greater than 6 but less than or equal to 12."

2.37. Consider an experiment for which the sample space is the real line. A probability law assigns probabilities to subsets of the form $(-\infty, r]$.

   (a) Show that we must have $P[(-\infty, r]] \leq P[(-\infty, s]]$ when $r < s$.

   (b) Find an expression for $P[(r, s]]$ in terms of $P[(-\infty, r]]$ and $P[(-\infty, s]]$

   (c) Find an expression for $P[(s, \infty)]$.

2.38. Two numbers (x, y) are selected at random from the interval $[0, 1]$.

   (a) Find the probability that the pair of numbers are inside the unit circle.

   (b) Find the probability that $y > 2x$.

*Section 2.3: Computing Probabilities Using Counting Methods

2.39. The combination to a lock is given by three numbers from the set $\{0, 1, \ldots, 59\}$. Find the number of combinations possible.

2.40. How many seven-digit telephone numbers are possible if the first number is not allowed to be 0 or 1?

2.41. A pair of dice is tossed, a coin is flipped twice, and a card is selected at random from a deck of 52 distinct cards. Find the number of possible outcomes.

2.42. A lock has two buttons: a "0" button and a "1" button. To open a door you need to push the buttons according to a preset 8-bit sequence. How many sequences are there? Suppose you press an arbitrary 8-bit sequence; what is the probability that the door opens? If the first try does not succeed in opening the door, you try another number; what is the probability of success?

2.43. A Web site requires that users create a password with the following specifications:
   - Length of 8 to 10 characters
   - Includes at least one special character $\{!, @, \#, \$, \%, \hat{}, \&, *, (,), +, =, \{, \}, |, <, >, \backslash, —, -, [,], /, ?\}$
   - No spaces
   - May contain numbers (0–9), lower and upper case letters (a–z, A–Z)
   - Is case-sensitive.

   How many passwords are there? How long would it take to try all passwords if a password can be tested in 1 microsecond?

2.44. A multiple choice test has 10 questions with 3 choices each. How many ways are there to answer the test? What is the probability that two papers have the same answers?

2.45. A student has five different t-shirts and three pairs of jeans ("brand new," "broken in," and "perfect").

   (a) How many days can the student dress without repeating the combination of jeans and t-shirt?

   (b) How many days can the student dress without repeating the combination of jeans and t-shirt and without wearing the same t-shirt on two consecutive days?

2.46. Ordering a "deluxe" pizza means you have four choices from 15 available toppings. How many combinations are possible if toppings can be repeated? If they cannot be repeated? Assume that the order in which the toppings are selected does not matter.

2.47. A lecture room has 60 seats. In how many ways can 45 students occupy the seats in the room?

2.48. List all possible permutations of two distinct objects; three distinct objects; four distinct objects. Verify that the number is $n!$.

2.49. A toddler pulls three volumes of an encyclopedia from a bookshelf and, after being scolded, places them back in random order. What is the probability that the books are in the correct order?

2.50. Five balls are placed at random in five buckets. What is the probability that each bucket has a ball?

2.51. List all possible combinations of two objects from two distinct objects; three distinct objects; four distinct objects. Verify that the number is given by the binomial coefficient.

2.52. A dinner party is attended by four men and four women. How many unique ways can the eight people sit around the table? How many unique ways can the people sit around the table with men and women alternating seats?

2.53. A hot dog vendor provides onions, relish, mustard, ketchup, Dijon ketchup, and hot peppers for your hot dog. How many variations of hot dogs are possible using one condiment? Two condiments? None, some, or all of the condiments?

2.54. A lot of 100 items contains k defective items. M items are chosen at random and tested.

   (a) What is the probability that m are found defective? This is called the *hypergeometric distribution*.

   (b) A lot is accepted if 1 or fewer of the M items are defective. What is the probability that the lot is accepted?

2.55. A park has N raccoons of which eight were previously captured and tagged. Suppose that 20 raccoons are captured. Find the probability that four of these are found to be tagged. Denote this probability, which depends on N, by $p(N)$. Find the value of N that maximizes this probability. *Hint:* Compare the ratio $p(N)/p(N-1)$ to unity.

2.56. A lot of 50 items has 40 good items and 10 bad items.

   (a) Suppose we test five samples from the lot, with replacement. Let X be the number of defective items in the sample. Find $P[X = k]$.

   (b) Suppose we test five samples from the lot, without replacement. Let Y be the number of defective items in the sample. Find $P[Y = k]$.

2.57. How many distinct permutations are there of four red balls, two white balls, and three black balls?

2.58. A hockey team has 6 forwards, 4 defensemen, and 2 goalies. At any time, 3 forwards, 2 defensemen, and 1 goalie can be on the ice. How many combinations of players can a coach put on the ice?

2.59. Find the probability that in a class of 28 students exactly four were born in each of the seven days of the week.

2.60. Show that

$$\binom{n}{k} = \binom{n}{n-k}$$

2.61. In this problem we derive the multinomial coefficient. Suppose we partition a set of n distinct objects into J subsets $B_1, B_2, \ldots, B_J$ of size $k_1, \ldots, k_J$, respectively, where $k_i \geq 0$, and $k_1 + k_2 + \ldots + k_J = n$.

   (a) Let N_i denote the number of possible outcomes when the ith subset is selected. Show that

$$N_1 = \binom{n}{k_1}, N_2 = \binom{n - k_1}{k_2}, \ldots, N_{J-1} = \binom{n - k_1 - \cdots - k_{J-2}}{k_{J-1}}.$$

(b) Show that the number of partitions is then:

$$N_1 N_2 \ldots N_{J-1} = \frac{n!}{k_1! \, k_2! \ldots k_J!}.$$

Section 2.4: Conditional Probability

2.62. A die is tossed twice and the number of dots facing up is counted and noted in the order of occurrence. Let A be the event "number of dots in first toss is not less than number of dots in second toss," and let B be the event "number of dots in first toss is 6." Find $P[A|B]$ and $P[B|A]$.

2.63. Use conditional probabilities and tree diagrams to find the probabilities for the elementary events in the random experiments defined in parts a to d of Problem 2.5.

2.64. In Problem 2.6 (name in hat), find $P[B \cap C|A]$ and $P[C|A \cap B]$.

2.65. In Problem 2.29 (message transmissions), find $P[B|A]$ and $P[A|B]$.

2.66. In Problem 2.8 (unit interval), find $P[B|A]$ and $P[A|B]$.

2.67. In Problem 2.36 (device lifetime), find $P[B|A]$ and $P[A|B]$.

2.68. In Problem 2.33, let $A = \{\text{hand rests in last 10 minutes}\}$ and $B = \{\text{hand rests in last 5 minutes}\}$. Find $P[B|A]$ for parts a, b, and c.

2.69. A number x is selected at random in the interval $[-1, 2]$. Let the events $A = \{x < 0\}$, $B = \{|x - 0.5| < 0.5\}$, and $C = \{x > 0.75\}$. Find $P[A|B]$, $P[B|C]$, $P[A|C^c]$, $P[B|C^c]$.

2.70. In Problem 2.36, let A be the event "lifetime is greater than t," and B the event "lifetime is greater than $2t$." Find $P[B|A]$. Does the answer depend on t? Comment.

2.71. Find the probability that two or more students in a class of 20 students have the same birthday. *Hint:* Use Corollary 1. How big should the class be so that the probability that two or more students have the same birthday is 1/2?

2.72. A cryptographic hash takes a message as input and produces a fixed-length string as output, called the digital fingerprint. A brute force attack involves computing the hash for a large number of messages until a pair of distinct messages with the same hash is found. Find the number of attempts required so that the probability of obtaining a match is 1/2. How many attempts are required to find a matching pair if the digital fingerprint is 64 bits long? 128 bits long?

2.73. **(a)** Find $P[A|B]$ if $A \cap B = \varnothing$; if $A \subset B$; if $A \supset B$.
   (b) Show that if $P[A|B] > P[A]$, then $P[B|A] > P[B]$.

2.74. Show that $P[A|B]$ satisfies the axioms of probability.
   (i) $0 \le P[A|B] \le 1$
   (ii) $P[S|B] = 1$
   (iii) If $A \cap C = \varnothing$, then $P[A \cup C|B] = P[A|B] + P[C|B]$.

2.75. Show that $P[A \cap B \cap C] = P[A|B \cap C]P[B|C]P[C]$.

2.76. In each lot of 100 items, two items are tested, and the lot is rejected if either of the tested items is found defective.
   (a) Find the probability that a lot with k defective items is accepted.
   (b) Suppose that when the production process malfunctions, 50 out of 100 items are defective. In order to identify when the process is malfunctioning, how many items should be tested so that the probability that one or more items are found defective is at least 99%?

2.77. A nonsymmetric binary communications channel is shown in Fig. P2.3. Assume the input is "0" with probability p and "1" with probability $1 - p$.

(a) Find the probability that the output is 0.

(b) Find the probability that the input was 0 given that the output is 1. Find the probability that the input is 1 given that the output is 1. Which input is more probable?

Input   Output

0 • ———— $1 - \varepsilon_1$ ————— • 0

          ε_1

     ε_2

          $1 - \varepsilon_2$

1 • ——————————————— • 1
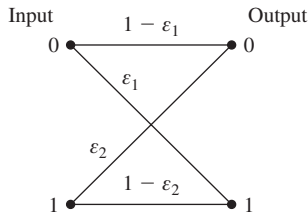
FIGURE P2.3

2.78. The transmitter in Problem 2.4 is equally likely to send $X = +2$ as $X = -2$. The malicious channel counts the number of heads in two tosses of a fair coin to decide by how much to reduce the magnitude of the input to produce the output Y.

(a) Use a tree diagram to find the set of possible input-output pairs.

(b) Find the probabilities of the input-output pairs.

(c) Find the probabilities of the output values.

(d) Find the probability that the input was $X = +2$ given that $Y = k$.

2.79. One of two coins is selected at random and tossed three times. The first coin comes up heads with probability p_1 and the second coin with probability $p_2 = 2/3 > p_1 = 1/3$.

(a) What is the probability that the number of heads is k?

(b) Find the probability that coin 1 was tossed given that k heads were observed, for $k = 0, 1, 2, 3$.

(c) In part b, which coin is more probable when k heads have been observed?

(d) Generalize the solution in part b to the case where the selected coin is tossed m times. In particular, find a threshold value T such that when $k > T$ heads are observed, coin 1 is more probable, and when $k < T$ are observed, coin 2 is more probable.

(e) Suppose that $p_2 = 1$ (that is, coin 2 is two-headed) and $0 < p_1 < 1$. What is the probability that we do not determine with certainty whether the coin is 1 or 2?

2.80. A computer manufacturer uses chips from three sources. Chips from sources A, B, and C are defective with probabilities .005, .001, and .010, respectively. If a randomly selected chip is found to be defective, find the probability that the manufacturer was A; that the manufacturer was C. Assume that the proportions of chips from A, B, and C are 0.5, 0.1, and 0.4, respectively.

2.81. A ternary communication system is shown in Fig. P2.4. Suppose that input symbols 0, 1, and 2 occur with probability 1/3 respectively.

(a) Find the probabilities of the output symbols.

(b) Suppose that a 1 is observed at the output. What is the probability that the input was 0? 1? 2?
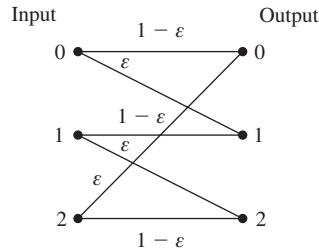
FIGURE P2.4

Section 2.5: Independence of Events

2.82. Let $S = \{1, 2, 3, 4\}$ and $A = \{1, 2\}$, $B = \{1, 3\}$, $C = \{1, 4\}$. Assume the outcomes are equiprobable. Are A, B, and C independent events?

2.83. Let U be selected at random from the unit interval. Let $A = \{0 < U < 1/2\}$, $B = \{1/4 < U < 3/4\}$, and $C = \{1/2 < U < 1\}$. Are any of these events independent?

2.84. Alice and Mary practice free throws at the basketball court after school. Alice makes free throws with probability p_a and Mary makes them with probability p_m. Find the probability of the following outcomes when Alice and Mary each take one shot: Alice scores a basket; Either Alice or Mary scores a basket; both score; both miss.

2.85. Show that if A and B are independent events, then the pairs A and B^c, A^c and B, and A^c and B^c are also independent.

2.86. Show that events A and B are independent if $P[A|B] = P[A|B^c]$.

2.87. Let A, B, and C be events with probabilities $P[A]$, $P[B]$, and $P[C]$.
   (a) Find $P[A \cup B]$ if A and B are independent.
   (b) Find $P[A \cup B]$ if A and B are mutually exclusive.
   (c) Find $P[A \cup B \cup C]$ if A, B, and C are independent.
   (d) Find $P[A \cup B \cup C]$ if A, B, and C are pairwise mutually exclusive.

2.88. An experiment consists of picking one of two urns at random and then selecting a ball from the urn and noting its color (black or white). Let A be the event "urn 1 is selected" and B the event "a black ball is observed." Under what conditions are A and B independent?

2.89. Find the probabilities in Problem 2.14 assuming that events A, B, and C are independent.

2.90. Find the probabilities that the three types of systems are "up" in Problem 2.15. Assume that all units in the system fail independently and that a type k unit fails with probability p_k.

2.91. Find the probabilities that the system is "up" in Problem 2.16. Assume that all units in the system fail independently and that a type k unit fails with probability p_k.

2.92. A random experiment is repeated a large number of times and the occurrence of events A and B is noted. How would you test whether events A and B are independent?

2.93. Consider a very long sequence of hexadecimal characters. How would you test whether the relative frequencies of the four bits in the hex characters are consistent with independent tosses of coin?

2.94. Compute the probability of the system in Example 2.35 being "up" when a second controller is added to the system.

2.95. In the binary communication system in Example 2.26, find the value of ε for which the input of the channel is independent of the output of the channel. Can such a channel be used to transmit information?

2.96. In the ternary communication system in Problem 2.81, is there a choice of ε for which the input of the channel is independent of the output of the channel?

Section 2.6: Sequential Experiments

2.97. A block of 100 bits is transmitted over a binary communication channel with probability of bit error $p = 10^{-2}$.
  - **(a)** If the block has 1 or fewer errors then the receiver accepts the block. Find the probability that the block is accepted.
  - **(b)** If the block has more than 1 error, then the block is retransmitted. Find the probability that M retransmissions are required.

2.98. A fraction p of items from a certain production line is defective.
  - **(a)** What is the probability that there is more than one defective item in a batch of n items?
  - **(b)** During normal production $p = 10^{-3}$ but when production malfunctions $p = 10^{-1}$. Find the size of a batch that should be tested so that if any items are found defective we are 99% sure that there is a production malfunction.

2.99. A student needs eight chips of a certain type to build a circuit. It is known that 5% of these chips are defective. How many chips should he buy for there to be a greater than 90% probability of having enough chips for the circuit?

2.100. Each of n terminals broadcasts a message in a given time slot with probability p.
  - **(a)** Find the probability that exactly one terminal transmits so the message is received by all terminals without collision.
  - **(b)** Find the value of p that maximizes the probability of successful transmission in part a.
  - **(c)** Find the asymptotic value of the probability of successful transmission as n becomes large.

2.101. A system contains eight chips. The lifetime of each chip has a Weibull probability law: with parameters λ and $k = 2$: $P[(t, \infty)] = e^{-(\lambda t)^k}$ for $t \geq 0$. Find the probability that at least two chips are functioning after $2/\lambda$ seconds.

2.102. A machine makes errors in a certain operation with probability p. There are two types of errors. The fraction of errors that are type 1 is α, and type 2 is $1 - \alpha$.
  - **(a)** What is the probability of k errors in n operations?
  - **(b)** What is the probability of k_1 type 1 errors in n operations?
  - **(c)** What is the probability of k_2 type 2 errors in n operations?
  - **(d)** What is the joint probability of k_1 and k_2 type 1 and 2 errors, respectively, in n operations?

2.103. Three types of packets arrive at a router port. Ten percent of the packets are "expedited forwarding (EF)," 30 percent are "assured forwarding (AF)," and 60 percent are "best effort (BE)."
  - **(a)** Find the probability that k of N packets are not expedited forwarding.
  - **(b)** Suppose that packets arrive one at a time. Find the probability that k packets are received before an expedited forwarding packet arrives.
  - **(c)** Find the probability that out of 20 packets, 4 are EF packets, 6 are AF packets, and 10 are BE.

2.104. A run-length coder segments a binary information sequence into strings that consist of either a "run" of k "zeros" punctuated by a "one", for $k = 0, \ldots, m - 1$, or a string of m "zeros." The $m = 3$ case is:

String	Run-length k
1	0
01	1
001	2
000	3

Suppose that the information is produced by a sequence of Bernoulli trials with $P[\text{"one"}] = P[\text{success}] = p$.

(a) Find the probability of run-length k in the $m = 3$ case.

(b) Find the probability of run-length k for general m.

2.105. The amount of time cars are parked in a parking lot follows a geometric probability law with $p = 1/2$. The charge for parking in the lot is $1 for each half-hour or less.

(a) Find the probability that a car pays k dollars.

(b) Suppose that there is a maximum charge of $6. Find the probability that a car pays k dollars.

2.106. A biased coin is tossed repeatedly until heads has come up three times. Find the probability that k tosses are required. *Hint:* Show that $\{\text{"}k \text{ tosses are required"}\} = A \cap B$, where $A = \{\text{"}k\text{th toss is heads"}\}$ and $B = \{\text{"2 heads occurs in } k - 1 \text{ tosses"}\}$.

2.107. An urn initially contains two black balls and two white balls. The following experiment is repeated indefinitely: A ball is drawn from the urn; if the color of the ball is the same as the majority of balls remaining in the urn, then the ball is put back in the urn. Otherwise the ball is left out.

(a) Draw the trellis diagram for this experiment and label the branches by the transition probabilities.

(b) Find the probabilities for all sequences of outcomes of length 2 and length 3.

(c) Find the probability that the urn contains no black balls after three draws; no white balls after three draws.

(d) Find the probability that the urn contains two black balls after n trials; two white balls after n trials.

2.108. In Example 2.45, let $p_0(n)$ and $p_1(n)$ be the probabilities that urn 0 or urn 1 is used in the nth subexperiment.

(a) Find $p_0(1)$ and $p_1(1)$.

(b) Express $p_0(n + 1)$ and $p_1(n + 1)$ in terms of $p_0(n)$ and $p_1(n)$.

(c) Evaluate $p_0(n)$ and $p_1(n)$ for $n = 2, 3, 4$.

(d) Find the solution to the recursion in part b with the initial conditions given in part a.

(e) What are the urn probabilities as n approaches infinity?

*Section 2.7: Synthesizing Randomness: Number Generators

2.109. An urn experiment is to be used to simulate a random experiment with sample space $S = \{1, 2, 3, 4, 5\}$ and probabilities $p_1 = 1/3$, $p_2 = 1/5$, $p_3 = 1/4$, $p_4 = 1/7$, and $p_5 = 1 - (p_1 + p_2 + p_3 + p_4)$. How many balls should the urn contain? Generalize

the result to show that an urn experiment can be used to simulate any random experiment with finite sample space and with probabilities given by rational numbers.

2.110. Suppose we are interested in using tosses of a fair coin to simulate a random experiment in which there are six equally likely outcomes, where $S = \{0, 1, 2, 3, 4, 5\}$. The following version of the "rejection method" is proposed:

  1. Toss a fair coin three times and obtain a binary number by identifying heads with zero and tails with one.

  2. If the outcome of the coin tosses in step 1 is the binary representation for a number in S, output the number. Otherwise, return to step 1.

  (a) Find the probability that a number is produced in step 2.

  (b) Show that the numbers that are produced in step 2 are equiprobable.

  (c) Generalize the above algorithm to show how coin tossing can be used to simulate any random urn experiment.

2.111. Use the `rand` function in Octave to generate 1000 pairs of numbers in the unit square. Plot an *x-y* scattergram to confirm that the resulting points are uniformly distributed in the unit square.

2.112. Apply the rejection method introduced above to generate points that are uniformly distributed in the $x > y$ portion of the unit square. Use the `rand` function to generate a pair of numbers in the unit square. If $x > y$, accept the number. If not, select another pair. Plot an *x-y* scattergram for the pair of accepted numbers and confirm that the resulting points are uniformly distributed in the $x > y$ region of the unit square.

2.113. The *sample mean-squared value* of the numerical outcomes $X(1), X(2), \ldots X(n)$ of a series of n repetitions of an experiment is defined by

$$\langle X^2 \rangle_n = \frac{1}{n} \sum_{j=1}^{n} X^2(j).$$

  (a) What would you expect this expression to converge to as the number of repetitions n becomes very large?

  (b) Find a recursion formula for $\langle X^2 \rangle_n$ similar to the one found in Problem 1.9.

2.114. The *sample variance* is defined as the mean-squared value of the variation of the samples about the sample mean

$$\langle V^2 \rangle_n = \frac{1}{n} \sum_{j=1}^{n} \{X(j) - \langle X \rangle_n\}^2.$$

Note that the $\langle X \rangle_n$ also depends on the sample values. (It is customary to replace the n in the denominator with $n - 1$ for technical reasons that will be discussed in Chapter 8. For now we will use the above definition.)

  (a) Show that the sample variance satisfies the following expression:

$$\langle V^2 \rangle_n = \langle X^2 \rangle_n - \langle X \rangle_n^2.$$

  (b) Show that the sample variance satisfies the following recursion formula:

$$\langle V^2 \rangle_n = \left(1 - \frac{1}{n}\right)\langle V^2 \rangle_{n-1} + \frac{1}{n}\left(1 - \frac{1}{n}\right)(X(n) - \langle X \rangle_{n-1})^2,$$

with $\langle V^2 \rangle_0 = 0$.

2.115. Suppose you have a program to generate a sequence of numbers U_n that is uniformly distributed in $[0, 1]$. Let $Y_n = \alpha U_n + \beta$.

    (a) Find α and β so that Y_n is uniformly distributed in the interval $[a, b]$.

    (b) Let $a = -5$ and $b = 15$. Use Octave to generate Y_n and to compute the sample mean and sample variance in 1000 repetitions. Compare the sample mean and sample variance to $(a + b)/2$ and $(b - a)^2/12$, respectively.

2.116. Use Octave to simulate 100 repetitions of the random experiment where a coin is tossed 16 times and the number of heads is counted.

    (a) Confirm that your results are similar to those in Figure 2.18.

    (b) Rerun the experiment with $p = 0.25$ and $p = 0.75$. Are the results as expected?

*Section 2.8: Fine Points: Event Classes

2.117. In Example 2.49, Homer maps the outcomes from Lisa's sample space $S_L = \{r, g, t\}$ into a smaller sample space $S_H = \{R, G\}: f(r) = R, f(g) = G$, and $f(t) = G$.

Define the inverse image events as follows:

$$f^{-1}(\{R\}) = A_1 = \{r\} \quad \text{and} \quad f^{-1}(\{G\}) = A_2 = \{g, t\}.$$

Let A and B be events in Homer's sample space.

    (a) Show that $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.

    (b) Show that $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.

    (c) Show that $f^{-1}(A^c) = f^{-1}(A)^c$.

    (d) Show that the results in parts a, b, and c hold for a general mapping f from a sample space S to a set S'.

2.118. Let f be a mapping from a sample space S to a finite set $S' = \{y_1, y_2, \dots, y_n\}$.

    (a) Show that the set of inverse images $A_k = f^{-1}(\{y_k\})$ forms a partition of S.

    (b) Show that any event B of S' can be related to a union of A_k's.

2.119. Let A be any subset of S. Show that the class of sets $\{\varnothing, A, A^c, S\}$ is a field.

*Section 2.9: Fine Points: Probabilities of Sequences of Events

2.120. Find the countable union of the following sequences of events:

    (a) $A_n = [a + 1/n, b - 1/n]$.

    (b) $B_n = (-n, b - 1/n]$.

    (c) $C_n = [a + 1/n, b)$.

2.121. Find the countable intersection of the following sequences of events:

    (a) $A_n = (a - 1/n, b + 1/n)$.

    (b) $B_n = [a, b + 1/n)$.

    (c) $C_n = (a - 1/n, b]$.

2.122. **(a)** Show that the Borel field can be generated from the complements and countable intersections and unions of open sets (a, b).

    (b) Suggest other classes of sets that can generate the Borel field.

2.123. Find expressions for the probabilities of the events in Problem 2.120.

2.124. Find expressions for the probabilities of the events in Problem 2.121.

Problems Requiring Cumulative Knowledge

2.125. Compare the binomial probability law and the hypergeometric law introduced in Problem 2.54 as follows.

  (a) Suppose a lot has 20 items of which five are defective. A batch of ten items is tested without replacement. Find the probability that k are found defective for $k = 0, \ldots, 10$. Compare this to the binomial probabilities with $n = 10$ and $p = 5/20 = .25$.

  (b) Repeat but with a lot of 1000 items of which 250 are defective. A batch of ten items is tested without replacement. Find the probability that k are found defective for $k = 0, \ldots, 10$. Compare this to the binomial probabilities with $n = 10$ and $p = 5/20 = .25$.

2.126. Suppose that in Example 2.43, computer A sends each message to computer B simultaneously over two unreliable radio links. Computer B can detect when errors have occurred in either link. Let the probability of message transmission error in link 1 and link 2 be q_1 and q_2 respectively. Computer B requests retransmissions until it receives an error-free message on either link.

  (a) Find the probability that more than k transmissions are required.

  (b) Find the probability that in the last transmission, the message on link 2 is received free of errors.

2.127. In order for a circuit board to work, seven identical chips must be in working order. To improve reliability, an additional chip is included in the board, and the design allows it to replace any of the seven other chips when they fail.

  (a) Find the probability p_b that the board is working in terms of the probability p that an individual chip is working.

  (b) Suppose that n circuit boards are operated in parallel, and that we require a 99.9% probability that at least one board is working. How many boards are needed?

2.128. Consider a well-shuffled deck of cards consisting of 52 distinct cards, of which four are aces and four are kings.

  (a) Find the probability of obtaining an ace in the first draw.

  (b) Draw a card from the deck and look at it. What is the probability of obtaining an ace in the second draw? Does the answer change if you had not observed the first draw?

  (c) Suppose we draw seven cards from the deck. What is the probability that the seven cards include three aces? What is the probability that the seven cards include two kings? What is the probability that the seven cards include three aces and/or two kings?

  (d) Suppose that the entire deck of cards is distributed equally among four players. What is the probability that each player gets an ace?

Discrete Random Variables

In most random experiments we are interested in a numerical attribute of the outcome of the experiment. A random variable is defined as a function that assigns a numerical value to the outcome of the experiment. In this chapter we introduce the concept of a random variable and methods for calculating probabilities of events involving a random variable. We focus on the simplest case, that of discrete random variables, and introduce the probability mass function. We define the expected value of a random variable and relate it to our intuitive notion of an average. We also introduce the conditional probability mass function for the case where we are given partial information about the random variable. These concepts and their extension in Chapter 4 provide us with the tools to evaluate the probabilities and averages of interest in the design of systems involving randomness.

Throughout the chapter we introduce important random variables and discuss typical applications where they arise. We also present methods for generating random variables. These methods are used in computer simulation models that predict the behavior and performance of complex modern systems.

3.1 THE NOTION OF A RANDOM VARIABLE

The outcome of a random experiment need not be a number. However, we are usually interested not in the outcome itself, but rather in some measurement or numerical attribute of the outcome. For example, in n tosses of a coin, we may be interested in the total number of heads and not in the specific order in which heads and tails occur. In a randomly selected Web document, we may be interested only in the length of the document. In each of these examples, *a measurement assigns a numerical value to the outcome of the random experiment.* Since the outcomes are random, the results of the measurements will also be random. Hence it makes sense to talk about the probabilities of the resulting numerical values. The concept of a random variable formalizes this notion.

A **random variable** X is a function that assigns a real number, $X(\zeta)$, to each outcome ζ in the sample space of a random experiment. Recall that a function is simply a rule for assigning a numerical value to each element of a set, as shown pictorially in
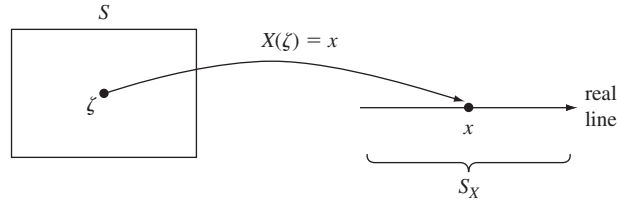
FIGURE 3.1
A random variable assigns a number $X(\zeta)$ to each outcome ζ in the
sample space S of a random experiment.

Fig. 3.1. The specification of a measurement on the outcome of a random experiment
defines a function on the sample space, and hence a random variable. The sample space
S is the *domain* of the random variable, and the set S_X of all values taken on by X is the
range of the random variable. Thus S_X is a subset of the set of all real numbers. We will
use the following notation: capital letters denote random variables, e.g., X or Y, and
lower case letters denote possible values of the random variables, e.g., x or y.

Example 3.1   Coin Tosses

A coin is tossed three times and the sequence of heads and tails is noted. The sample space for this
experiment is $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. Let X be the number of
heads in the three tosses. X assigns each outcome ζ in S a number from the set $S_X = \{0, 1, 2, 3\}$.
The table below lists the eight outcomes of S and the corresponding values of X.

ζ:	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(\zeta)$:	3	2	2	2	1	1	1	0

X is then a random variable taking on values in the set $S_X = \{0, 1, 2, 3\}$.

Example 3.2   A Betting Game

A player pays \$1.50 to play the following game: A coin is tossed three times and the number of
heads X is counted. The player receives \$1 if $X = 2$ and \$8 if $X = 3$, but nothing otherwise. Let
Y be the reward to the player. Y is a function of the random variable X and its outcomes can be
related back to the sample space of the underlying random experiment as follows:

ζ:	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(\zeta)$:	3	2	2	2	1	1	1	0
$Y(\zeta)$:	8	1	1	1	0	0	0	0

Y is then a random variable taking on values in the set $S_Y = \{0, 1, 8\}$.

The above example shows that a function of a random variable produces another random variable.

For random variables, the function or rule that assigns values to each outcome is fixed and deterministic, as, for example, in the rule "count the total number of dots facing up in the toss of two dice." The randomness in the experiment is complete as soon as the toss is done. The process of counting the dots facing up is deterministic. Therefore the distribution of the values of a random variable X is determined by the probabilities of the outcomes ζ in the random experiment. In other words, the randomness in the observed values of X is *induced* by the underlying random experiment, and we should therefore be able to compute the probabilities of the observed values of X in terms of the probabilities of the underlying outcomes.

Example 3.3   Coin Tosses and Betting

Let X be the number of heads in three independent tosses of a fair coin. Find the probability of the event $\{X = 2\}$. Find the probability that the player in Example 3.2 wins \$8.

Note that $X(\zeta) = 2$ if and only if ζ is in $\{HHT, HTH, THH\}$. Therefore

$$P[X = 2] = P[\{HHT, HTH, HHT\}]$$
$$= P[\{HHT\}] + P[\{HTH\}] + P[\{HHT\}]$$
$$= 3/8.$$

The event $\{Y = 8\}$ occurs if and only if the outcome ζ is HHH, therefore

$$P[Y = 8] = P[\{HHH\}] = 1/8.$$

Example 3.3 illustrates a general technique for finding the probabilities of events involving the random variable X. Let the underlying random experiment have sample space S and event class $\mathcal{F}$. To find the probability of a subset B of R, e.g., $B = \{x_k\}$, we need to find the outcomes in S that are mapped to B, that is,

$$A = \{\zeta : X(\zeta) \in B\} \tag{3.1}$$

as shown in Fig. 3.2. If event A occurs then $X(\zeta) \in B$, so event B occurs. Conversely, if event B occurs, then the value $X(\zeta)$ implies that ζ is in A, so event A occurs. Thus the probability that X is in B is given by:

$$P[X \in B] = P[A] = P[\{\zeta : X(\zeta) \in B\}]. \tag{3.2}$$
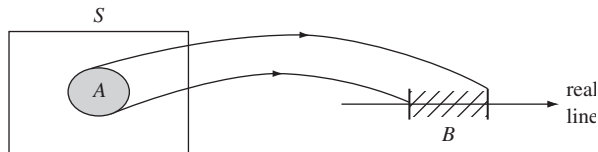
FIGURE 3.2
$P[X \text{ in } B] = P[\zeta \text{ in } A]$

We refer to A and B as **equivalent events**.

In some random experiments the outcome ζ is already the numerical value we are interested in. In such cases we simply let $X(\zeta) = \zeta$, that is, the identity function, to obtain a random variable.

*3.1.1   Fine Point: Formal Definition of a Random Variable

In going from Eq. (3.1) to Eq. (3.2) we actually need to check that the event A is in $\mathcal{F}$, because only events in $\mathcal{F}$ have probabilities assigned to them. The formal definition of a random variable in Chapter 4 will explicitly state this requirement.

If the event class $\mathcal{F}$ consists of all subsets of S, then the set A will always be in $\mathcal{F}$, and any function from S to R will be a random variable. However, if the event class $\mathcal{F}$ does not consist of all subsets of S, then some functions from S to R may not be random variables, as illustrated by the following example.

Example 3.4   A Function That Is Not a Random Variable

This example shows why the definition of a random variable requires that we check that the set A is in $\mathcal{F}$. An urn contains three balls. One ball is electronically coded with a label 00. Another ball is coded with 01, and the third ball has a 10 label. The sample space for this experiment is $S = \{00, 01, 10\}$. Let the event class $\mathcal{F}$ consist of all unions, intersections, and complements of the events $A_1 = \{00, 10\}$ and $A_2 = \{01\}$. In this event class, the outcome 00 cannot be distinguished from the outcome 10. For example, this could result from a faulty label reader that cannot distinguish between 00 and 10. The event class has four events $\mathcal{F} = \{\varnothing, \{00, 10\}, \{01\}, \{00, 01, 10\}\}$. Let the probability assignment for the events in $\mathcal{F}$ be $P[\{00, 10\}] = 2/3$ and $P[\{01\}] = 1/3$.

Consider the following function X from S to R: $X(00) = 0$, $X(01) = 1$, $X(10) = 2$. To find the probability of $\{X = 0\}$, we need the probability of $\{\zeta : X(\zeta) = 0\} = \{00\}$. However, $\{00\}$ is not in the class $\mathcal{F}$, and so X is not a random variable because we cannot determine the probability that $X = 0$.

3.2   DISCRETE RANDOM VARIABLES AND PROBABILITY MASS FUNCTION

A **discrete random variable** X is defined as a random variable that assumes values from a countable set, that is, $S_X = \{x_1, x_2, x_3, \dots\}$. A discrete random variable is said to be **finite** if its range is finite, that is, $S_X = \{x_1, x_2, \dots, x_n\}$. We are interested in finding the probabilities of events involving a discrete random variable X. Since the sample space S_X is discrete, we only need to obtain the probabilities for the events $A_k = \{\zeta : X(\zeta) = x_k\}$ in the underlying random experiment. The probabilities of all events involving X can be found from the probabilities of the A_k's.

The **probability mass function (pmf) of a discrete random variable** X is defined as:

$$p_X(x) = P[X = x] = P[\{\zeta : X(\zeta) = x\}] \quad \text{for } x \text{ a real number.} \qquad (3.3)$$

Note that $p_X(x)$ is a function of x over the real line, and that $p_X(x)$ can be nonzero only at the values $x_1, x_2, x_3, \dots$. For x_k in S_X, we have $p_X(x_k) = P[A_k]$.
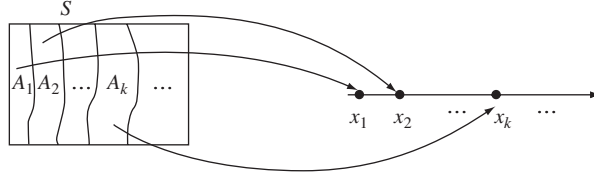
FIGURE 3.3
Partition of sample space S associated with a discrete random variable.

The events $A_1, A_2, \ldots$ form a partition of S as illustrated in Fig. 3.3. To see this, we first show that the events are disjoint. Let $j \neq k$, then

$$A_j \cap A_k = \{\zeta: X(\zeta) = x_j \text{ and } X(\zeta) = x_k\} = \varnothing$$

since each ζ is mapped into one and only one value in S_X. Next we show that S is the union of the A_k's. Every ζ in S is mapped into some x_k so that every ζ belongs to an event A_k in the partition. Therefore:

$$S = A_1 \cup A_2 \cup \ldots.$$

All events involving the random variable X can be expressed as the union of events A_k's. For example, suppose we are interested in the event X in $B = \{x_2, x_5\}$, then

$$P[X \text{ in } B] = P[\{\zeta: X(\zeta) = x_2\} \cup \{\zeta: X(\zeta) = x_5\}]$$

$$= P[A_2 \cup A_5] = P[A_2] + P[A_5]$$

$$= p_X(2) + p_X(5).$$

The pmf $p_X(x)$ satisfies three properties that provide all the information required to calculate probabilities for events involving the discrete random variable X:

(i) $p_X(x) \geq 0$ for all x (3.4a)

(ii) $\displaystyle\sum_{x \in S_X} p_X(x) = \sum_{\text{all } k} p_X(x_k) = \sum_{\text{all } k} P[A_k] = 1$ (3.4b)

(iii) $P[X \text{ in } B] = \displaystyle\sum_{x \in B} p_X(x)$ where $B \subset S_X$.   (3.4c)

Property (i) is true because the pmf values are defined as a probability, $p_X(x) = P[X = x]$. Property (ii) follows because the events $A_k = \{X = x_k\}$ form a partition of S. Note that the summations in Eqs. (3.4b) and (3.4c) will have a finite or infinite number of terms depending on whether the random variable is finite or not. Next consider property (iii). Any event B involving X is the union of elementary events, so by Axiom III' we have:

$$P[X \text{ in } B] = P\left[\bigcup_{x \in B} \{\zeta: X(\zeta) = x\}\right] = \sum_{x \in B} P[X = x] = \sum_{x \in B} p_X(x).$$

The pmf of X gives us the probabilities for all the elementary events from S_X. The probability of any subset of S_X is obtained from the sum of the corresponding elementary events. In fact we have everything required to specify a probability law for the outcomes in S_X. If we are only interested in events concerning X, then we can forget about the underlying random experiment and its associated probability law and just work with S_X and the pmf of X.

Example 3.5   Coin Tosses and Binomial Random Variable

Let X be the number of heads in three independent tosses of a coin. Find the pmf of X.
    Proceeding as in Example 3.3, we find:

$$p_0 = P[X = 0] = P[\{TTT\}] = (1 - p)^3,$$

$$p_1 = P[X = 1] = P[\{HTT\}] + P[\{THT\}] + P[\{TTH\}] = 3(1 - p)^2 p,$$

$$p_2 = P[X = 2] = P[\{HHT\}] + P[\{HTH\}] + P[\{THH\}] = 3(1 - p)p^2,$$

$$p_3 = P[X = 3] = P[\{HHH\}] = p^3.$$

Note that $p_X(0) + p_X(1) + p_X(2) + p_X(3) = 1$.

Example 3.6   A Betting Game

A player receives \$1 if the number of heads in three coin tosses is 2, \$8 if the number is 3, but nothing otherwise. Find the pmf of the reward Y.

$$p_Y(0) = P[\zeta \in \{TTT, TTH, THT, HTT\}] = 4/8 = 1/2$$

$$p_Y(1) = P[\zeta \in \{THH, HTH, HHT\}] = 3/8$$

$$p_Y(8) = P[\zeta \in \{HHH\}] = 1/8.$$

Note that $p_Y(0) + p_Y(1) + p_Y(8) = 1$.

Figures 3.4(a) and (b) show the graph of $p_X(x)$ versus x for the random variables in Examples 3.5 and 3.6, respectively. In general, the graph of the pmf of a discrete random variable has vertical arrows of height $p_X(x_k)$ at the values x_k in S_X. We may view the total probability as one unit of mass and $p_X(x)$ as the amount of probability mass that is placed at each of the discrete points $x_1, x_2, \ldots$. The relative values of pmf at different points give an indication of the relative likelihoods of occurrence.

Example 3.7   Random Number Generator

A random number generator produces an integer number X that is equally likely to be any element in the set $S_X = \{0, 1, 2, \ldots, M - 1\}$. Find the pmf of X.
    For each k in S_X, we have $p_X(k) = 1/M$. Note that

$$p_X(0) + p_X(1) + \ldots + p_X(M - 1) = 1.$$

We call X the **uniform random variable** in the set $\{0, 1, \ldots, M - 1\}$.
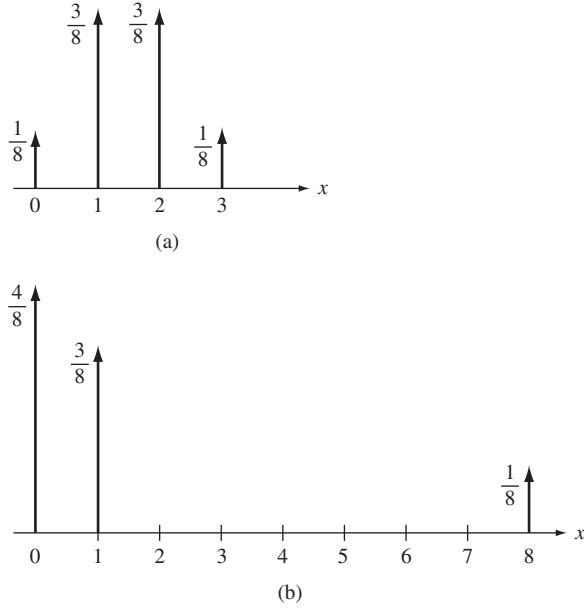
FIGURE 3.4
(a) Graph of pmf in three coin tosses; (b) Graph of pmf in betting game.

Example 3.8   Bernoulli Random Variable

Let A be an event of interest in some random experiment, e.g., a device is not defective. We say that a "success" occurs if A occurs when we perform the experiment. The Bernoulli random variable I_A is equal to 1 if A occurs and zero otherwise, and is given by the *indicator function for A*:

$$I_A(\zeta) = \begin{cases} 0 & \text{if } \zeta \text{ not in } A \\ 1 & \text{if } \zeta \text{ in } A. \end{cases} \tag{3.5a}$$

Find the pmf of I_A.

$I_A(\zeta)$ is a finite discrete random variable with values from $S_I = \{0, 1\}$, with pmf:

$$p_I(0) = P[\{\zeta : \zeta \in A^c\}] = 1 - p$$

$$p_I(1) = P[\{\zeta : \zeta \in A\}] = p. \tag{3.5b}$$

We call I_A the **Bernoulli random variable**. Note that $p_I(1) + p_I(2) = 1$.

Example 3.9   Message Transmissions

Let X be the number of times a message needs to be transmitted until it arrives correctly at its destination. Find the pmf of X. Find the probability that X is an even number.

X is a discrete random variable taking on values from $S_X = \{1, 2, 3, \dots\}$. The event $\{X = k\}$ occurs if the underlying experiment finds $k - 1$ consecutive erroneous transmissions

("failures") followed by a error-free one ("success"):

$$p_X(k) = P[X = k] = P[00\ldots01] = (1 - p)^{k-1}p = q^{k-1}p \quad k = 1, 2, \ldots. \quad (3.6)$$

We call X the **geometric random variable**, and we say that X is geometrically distributed. In Eq. (2.42b), we saw that the sum of the geometric probabilities is 1.

$$P[X \text{ is even}] = \sum_{k=1}^{\infty} p_X(2k) = p\sum_{k=1}^{\infty} q^{2k-1} = p\frac{1}{1 - q^2} = \frac{1}{1 + q}.$$

Example 3.10   Transmission Errors

A binary communications channel introduces a bit error in a transmission with probability p. Let X be the number of errors in n independent transmissions. Find the pmf of X. Find the probability of one or fewer errors.

X takes on values in the set $S_X = \{0, 1, \ldots, n\}$. Each transmission results in a "0" if there is no error and a "1" if there is an error, $P["1"] = p$ and $P["0"] = 1 - p$. The probability of k errors in n bit transmissions is given by the probability of an error pattern that has k 1's and $n - k$ 0's:

$$p_X(k) = P[X = k] = \binom{n}{k}p^k(1 - p)^{n-k} \quad k = 0, 1, \ldots, n. \quad (3.7)$$

We call X the **binomial random variable**, with parameters n and p. In Eq. (2.39b), we saw that the sum of the binomial probabilities is 1.

$$P[X \leq 1] = \binom{n}{0}p^0(1 - p)^{n-0} + \binom{n}{1}p^1(1 - p)^{n-1} = (1 - p)^n + np(1 - p)^{n-1}.$$

Finally, let's consider the relationship between relative frequencies and the pmf $p_X(x_k)$. Suppose we perform n independent repetitions to obtain n observations of the discrete random variable X. Let $N_k(n)$ be the number of times the event $X = x_k$ occurs and let $f_k(n) = N_k(n)/n$ be the corresponding relative frequency. As n becomes large we expect that $f_k(n) \to p_X(x_k)$. Therefore the graph of relative frequencies should approach the graph of the pmf. Figure 3.5(a) shows the graph of relative
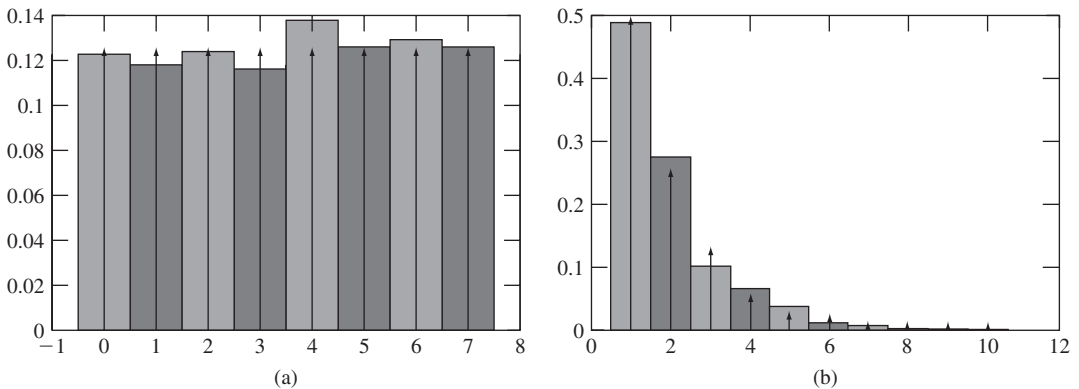
FIGURE 3.5
(a) Relative frequencies and corresponding uniform pmf; (b) Relative frequencies and corresponding geometric pmf.

frequencies for 1000 repetitions of an experiment that generates a uniform random variable from the set $\{0, 1, \ldots, 7\}$ and the corresponding pmf. Figure 3.5(b) shows the graph of relative frequencies and pmf for a geometric random variable with $p = 1/2$ and $n = 1000$ repetitions. In both cases we see that the graph of relative frequencies approaches that of the pmf.

3.3   EXPECTED VALUE AND MOMENTS OF DISCRETE RANDOM VARIABLE

In order to completely describe the behavior of a discrete random variable, an entire function, namely $p_X(x)$, must be given. In some situations we are interested in a few parameters that summarize the information provided by the pmf. For example, Fig. 3.6 shows the results of many repetitions of an experiment that produces two random variables. The random variable Y varies about the value 0, whereas the random variable X varies around the value 5. It is also clear that X is more spread out than Y. In this section we introduce parameters that quantify these properties.

The **expected value** or **mean** of a **discrete random variable** X is defined by

$$m_X = E[X] = \sum_{x \in S_X} x p_X(x) = \sum_k x_k p_X(x_k). \tag{3.8}$$

The expected value $E[X]$ is defined if the above sum converges absolutely, that is,

$$E[|X|] = \sum_k |x_k| p_X(x_k) < \infty. \tag{3.9}$$

There are random variables for which Eq. (3.9) does not converge. In such cases, we say that the expected value does not exist.
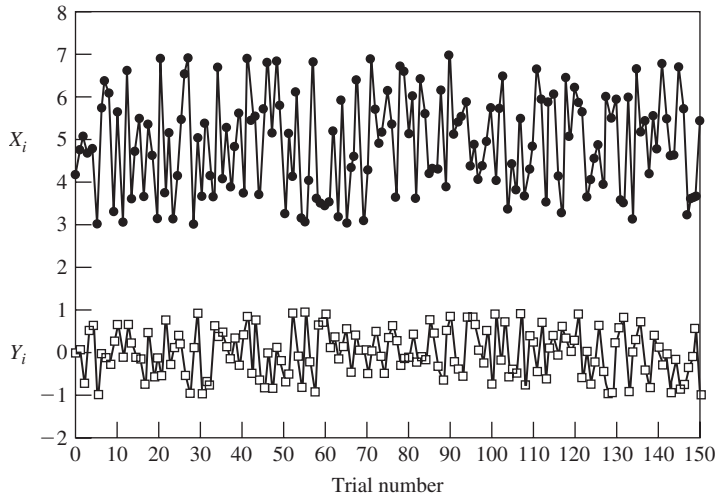
FIGURE 3.6
The graphs show 150 repetitions of the experiments yielding X and Y. It is clear that X is centered about the value 5 while Y is centered about 0. It is also clear that X is more spread out than Y.

If we view $p_X(x)$ as the distribution of mass on the points $x_1, x_2, \ldots$ in the real line, then $E[X]$ represents the center of mass of this distribution. For example, in Fig. 3.5(a), we can see that the pmf of a discrete random variable that is uniformly distributed in $\{0, \ldots, 7\}$ has a center of mass at 3.5.

Example 3.11   Mean of Bernoulli Random Variable

Find the expected value of the Bernoulli random variable I_A.

From Example 3.8, we have

$$E[I_A] = 0p_I(0) + 1p_I(1) = p.$$

where p is the probability of success in the Bernoulli trial.

Example 3.12   Three Coin Tosses and Binomial Random Variable

Let X be the number of heads in three tosses of a fair coin. Find $E[X]$.

Equation (3.8) and the pmf of X that was found in Example 3.5 gives:

$$E[X] = \sum_{k=0}^{3} k p_X(k) = 0\left(\frac{1}{8}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) = 1.5.$$

Note that the above is the $n = 3$, $p = 1/2$ case of a binomial random variable, which we will see has $E[X] = np$.

Example 3.13   Mean of a Uniform Discrete Random Variable

Let X be the random number generator in Example 3.7. Find $E[X]$.

From Example 3.5 we have $p_X(j) = 1/M$ for $j = 0, \ldots, M - 1$, so

$$E[X] = \sum_{k=0}^{M-1} k\frac{1}{M} = \frac{1}{M}\{0 + 1 + 2 + \cdots + M - 1\} = \frac{(M-1)M}{2M} = \frac{(M-1)}{2}$$

where we used the fact that $1 + 2 + \cdots + L = (L + 1)L/2$. Note that for $M = 8$, $E[X] = 3.5$, which is consistent with our observation of the center of mass in Fig. 3.5(a).

The use of the term "expected value" does not mean that we expect to observe $E[X]$ when we perform the experiment that generates X. For example, the expected value of a Bernoulli trial is p, but its outcomes are always either 0 or 1.

$E[X]$ corresponds to the "average of X" in a large number of observations of X. Suppose we perform n independent repetitions of the experiment that generates X, and we record the observed values as $x(1), x(2), \ldots, x(n)$, where $x(j)$ is the observation in the jth experiment. Let $N_k(n)$ be the number of times x_k is observed, and let $f_k(n) = N_k(n)/n$ be the corresponding relative frequency. The arithmetic average, or **sample mean**, of the observations, is:

$$\langle X \rangle_n = \frac{x(1) + x(2) + \cdots + x(n)}{n} = \frac{x_1 N_1(n) + x_2 N_2(n) + \cdots + x_k N_k(n) + \cdots}{n}$$

$$= x_1 f_1(n) + x_2 f_2(n) + \cdots + x_k f_k(n) + \cdots$$

$$= \sum_k x_k f_k(n). \tag{3.10}$$

The first numerator adds the observations in the order in which they occur, and the second numerator counts how many times each x_k occurs and then computes the total. As n becomes large, we expect relative frequencies to approach the probabilities $p_X(x_k)$:

$$\lim_{n \to \infty} f_k(n) = p_X(x_k) \quad \text{for all } k. \tag{3.11}$$

Equation (3.10) then implies that:

$$\langle X \rangle_n = \sum_k x_k f_k(n) \to \sum_k x_k p_X(x_k) = E[X]. \tag{3.12}$$

Thus we expect the sample mean to converge to E[X] as n becomes large.

Example 3.14   A Betting Game

A player at a fair pays $1.50 to toss a coin three times. The player receives $1 if the number of heads is 2, $8 if the number is 3, but nothing otherwise. Find the expected value of the reward Y. What is the expected value of the gain?

The expected reward is:

$$E[Y] = 0p_Y(0) + 1p_Y(1)) + 8p_Y(8) = 0\left(\frac{4}{8}\right) + 1\left(\frac{3}{8}\right) + 8\left(\frac{1}{8}\right) = \left(\frac{11}{8}\right).$$

The expected gain is:

$$E[Y - 1.5] = \frac{11}{8} - \frac{12}{8} = -\frac{1}{8}.$$

Players lose 12.5 cents on average per game, so the house makes a nice profit over the long run. In Example 3.18 we will see that some engineering designs also "bet" that users will behave a certain way.

Example 3.15   Mean of a Geometric Random Variable

Let X be the number of bytes in a message, and suppose that X has a geometric distribution with parameter p. Find the mean of X.

X can take on arbitrarily large values since $S_X = \{1, 2, \ldots\}$. The expected value is:

$$E[X] = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty} kq^{k-1}.$$

This expression is readily evaluated by differentiating the series

$$\frac{1}{1 - x} = \sum_{k=0}^{\infty} x^k \tag{3.13}$$

to obtain

$$\frac{1}{(1 - x)^2} = \sum_{k=0}^{\infty} kx^{k-1}. \tag{3.14}$$

Letting $x = q$, we obtain

$$E[X] = p\frac{1}{(1 - q)^2} = \frac{1}{p}. \tag{3.15}$$

We see that X has a finite expected value as long as $p > 0$.

For certain random variables large values occur sufficiently frequently that the expected value does not exist, as illustrated by the following example.

Example 3.16   St. Petersburg Paradox

A fair coin is tossed repeatedly until a tail comes up. If X tosses are needed, then the casino pays the gambler $Y = 2^X$ dollars. How much should the gambler be willing to pay to play this game?

If the gambler plays this game a large number of times, then the payoff should be the expected value of $Y = 2^X$. If the coin is fair, $P[X = k] = (1/2)^k$ and $P[Y = 2^k] = (1/2)^k$, so:

$$E[Y] = \sum_{k=1}^{\infty} 2^k p_Y(2^k) = \sum_{k=1}^{\infty} 2^k \left(\frac{1}{2}\right)^k = 1 + 1 + \cdots = \infty.$$

This game does indeed appear to offer the gambler a sweet deal, and so the gambler should be willing to pay any amount to play the game! The paradox is that a sane person would not pay a lot to play this game. Problem 3.34 discusses ways to resolve the paradox.

Random variables with unbounded expected value are not uncommon and appear in models where outcomes that have extremely large values are not that rare. Examples include the sizes of files in Web transfers, frequencies of words in large bodies of text, and various financial and economic problems.

3.3.1   Expected Value of Functions of a Random Variable

Let X be a discrete random variable, and let $Z = g(X)$. Since X is discrete, $Z = g(X)$ will assume a countable set of values of the form $g(x_k)$ where $x_k \in S_X$. Denote the set of values assumed by $g(X)$ by $\{z_1, z_2, \ldots\}$. One way to find the expected value of Z is to use Eq. (3.8), which requires that we first find the pmf of Z. Another way is to use the following result:

$$E[Z] = E[g(X)] = \sum_k g(x_k) p_X(x_k). \tag{3.16}$$

To show Eq. (3.16) group the terms x_k that are mapped to each value z_j:

$$\sum_k g(x_k) p_X(x_k) = \sum_j z_j \left\{ \sum_{x_k : g(x_k) = z_j} p_X(x_k) \right\} = \sum_j z_j p_Z(z_j) = E[Z].$$

The sum inside the braces is the probability of all terms x_k for which $g(x_k) = z_j$, which is the probability that $Z = z_j$, that is, $p_Z(z_j)$.

Example 3.17   Square-Law Device

Let X be a noise voltage that is uniformly distributed in $S_X = \{-3, -1, +1, +3\}$ with $p_X(k) = 1/4$ for k in S_X. Find $E[Z]$ where $Z = X^2$.

Using the first approach we find the pmf of Z:

$$p_Z(9) = P[X \in \{-3, +3\}] = p_X(-3) + p_X(3) = 1/2$$

$$p_Z(1) = p_X(-1) + p_X(1) = 1/2$$

and so

$$E[Z] = 1\left(\frac{1}{2}\right) + 9\left(\frac{1}{2}\right) = 5.$$

The second approach gives:

$$E[Z] = E[X^2] = \sum_k k^2 p_X(k) = \frac{1}{4}\{(-3)^2 + (-1)^2 + 1^2 + 3^2\} = \frac{20}{4} = 5.$$

Equation 3.16 implies several very useful results. Let Z be the function

$$Z = ag(X) + bh(X) + c$$

where $a, b,$ and c are real numbers, then:

$$E[Z] = aE[g(X)] + bE[h(X)] + c. \tag{3.17a}$$

From Eq. (3.16) we have:

$$E[Z] = E[ag(X) + bh(X) + c] = \sum_k (ag(x_k) + bh(x_k) + c)p_X(x_k)$$

$$= a\sum_k g(x_k)p_X(x_k) + b\sum_k h(x_k)p_X(x_k) + c\sum_k p_X(x_k)$$

$$= aE[g(X)] + bE[h(X)] + c.$$

Equation (3.17a), by setting $a, b,$ and/or c to 0 or 1, implies the following expressions:

$$E[g(X) + h(X)] = E[g(X)] + E[h(X)]. \tag{3.17b}$$

$$E[aX] = aE[X]. \tag{3.17c}$$

$$E[X + c] = E[X] + c. \tag{3.17d}$$

$$E[c] = c. \tag{3.17e}$$

Example 3.18   Square-Law Device

The noise voltage X in the previous example is amplified and shifted to obtain $Y = 2X + 10$, and then squared to produce $Z = Y^2 = (2X + 10)^2$. Find $E[Z]$.

$$E[Z] = E[(2X + 10)^2] = E[4X^2 + 40X + 100]$$

$$= 4E[X^2] + 40E[X] + 100 = 4(5) + 40(0) + 100 = 120.$$

Example 3.19   Voice Packet Multiplexer

Let X be the number of voice packets containing active speech produced by $n = 48$ independent speakers in a 10-millisecond period as discussed in Section 1.4. X is a binomial random variable with parameter n and probability $p = 1/3$. Suppose a packet multiplexer transmits up to $M = 20$ active packets every 10 ms, and any excess active packets are discarded. Let Z be the number of packets discarded. Find $E[Z]$.

The number of packets discarded every 10 ms is the following function of X:

$$Z = (X - M)^+ \triangleq \begin{cases} 0 & \text{if } X \le M \\ X - M & \text{if } X > M. \end{cases}$$

$$E[Z] = \sum_{k=20}^{48} (k - 20)\binom{48}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{48-k} = 0.182.$$

Every 10 ms $E[X] = np = 16$ active packets are produced on average, so the fraction of active packets discarded is $0.182/16 = 1.1\%$, which users will tolerate. This example shows that engineered systems also play "betting" games where favorable statistics are exploited to use resources efficiently. In this example, the multiplexer transmits 20 packets per period instead of 48 for a reduction of $28/48 = 58\%$.

3.3.2   Variance of a Random Variable

The expected value $E[X]$, by itself, provides us with limited information about X. For example, if we know that $E[X] = 0$, then it could be that X is zero all the time. However, it is also possible that X can take on extremely large positive and negative values. We are therefore interested not only in the mean of a random variable, but also in the extent of the random variable's variation about its mean. Let the deviation of the random variable X about its mean be $X - E[X]$, which can take on positive and negative values. Since we are interested in the magnitude of the variations only, it is convenient to work with the square of the deviation, which is always positive, $D(X) = (X - E[X])^2$. The expected value is a constant, so we will denote it by $m_X = E[X]$. The **variance of the random variable** X is defined as the expected value of D:

$$\sigma_X^2 = \text{VAR}[X] = E[(X - m_X)^2]$$

$$= \sum_{x \in S_X} (x - m_X)^2 p_X(x) = \sum_{k=1}^{\infty}(x_k - m_X)^2 p_X(x_k). \qquad (3.18)$$

The **standard deviation of the random variable** X is defined by:

$$\sigma_X = \text{STD}[X] = \text{VAR}[X]^{1/2}. \qquad (3.19)$$

By taking the square root of the variance we obtain a quantity with the same units as X.

An alternative expression for the variance can be obtained as follows:

$$\text{VAR}[X] = E[(X - m_X)^2] = E[X^2 - 2m_X X + m_X^2]$$

$$= E[X^2] - 2m_X E[X] + m_X^2$$

$$= E[X^2] - m_X^2. \qquad (3.20)$$

$E[X^2]$ is called the **second moment of** X. The ***n*th moment of** X is defined as $E[X^n]$.

Equations (3.17c), (3.17d), and (3.17e) imply the following useful expressions for the variance. Let $Y = X + c$, then

$$\text{VAR}[X + c] = E[(X + c - (E[X] + c)])^2]$$

$$= E[(X - E[X])^2] = \text{VAR}[X]. \qquad (3.21)$$

Adding a constant to a random variable does not affect the variance. Let $Z = cX$, then:

$$\text{VAR}[cX] = E[(cX - cE[X])^2] = E[c^2(X - E[X])^2] = c^2\,\text{VAR}[X]. \quad (3.22)$$

Scaling a random variable by c scales the variance by c^2 and the standard deviation by $|c|$.

Now let $X = c$, a random variable that is equal to a constant with probability 1, then

$$\text{VAR}[X] = E[(X - c)^2] = E[0] = 0. \quad (3.23)$$

A constant random variable has zero variance.

Example 3.20   Three Coin Tosses

Let X be the number of heads in three tosses of a fair coin. Find VAR[X].

$$E[X^2] = 0\left(\frac{1}{8}\right) + 1^2\left(\frac{3}{8}\right) + 2^2\left(\frac{3}{8}\right) + 3^2\left(\frac{1}{8}\right) = 3 \quad \text{and}$$

$$\text{VAR}[X] = E[X^2] - m_X^2 = 3 - 1.5^2 = 0.75.$$

Recall that this is an $n = 3$, $p = 1/2$ binomial random variable. We see later that variance for the binomial random variable is npq.

Example 3.21   Variance of Bernoulli Random Variable

Find the variance of the Bernoulli random variable I_A.

$$E[I_A^2] = 0p_I(0) + 1^2p_I(1) = p \quad \text{and so}$$

$$\text{VAR}[I_A] = p - p^2 = p(1 - p) = pq. \quad (3.24)$$

Example 3.22   Variance of Geometric Random Variable

Find the variance of the geometric random variable.

Differentiate the term $(1 - x^2)^{-1}$ in Eq. (3.14) to obtain

$$\frac{2}{(1 - x)^3} = \sum_{k=0}^{\infty} k(k - 1)x^{k-2}.$$

Let $x = q$ and multiply both sides by pq to obtain:

$$\frac{2pq}{(1 - q)^3} = pq\sum_{k=0}^{\infty} k(k - 1)q^{k-2}$$

$$= \sum_{k=0}^{\infty} k(k - 1)pq^{k-1} = E[X^2] - E[X].$$

So the second moment is

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

and the variance is

$$\mathrm{VAR}[X] = E[X^2] - E[X]^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

3.4 CONDITIONAL PROBABILITY MASS FUNCTION

In many situations we have partial information about a random variable X or about the outcome of its underlying random experiment. We are interested in how this information changes the probability of events involving the random variable. The conditional probability mass function addresses this question for discrete random variables.

3.4.1 Conditional Probability Mass Function

Let X be a discrete random variable with pmf $p_X(x)$, and let C be an event that has nonzero probability, $P[C] > 0$. See Fig. 3.7. The **conditional probability mass function** of X is defined by the conditional probability:

$$p_X(x|C) = P[X = x|C] \qquad \text{for } x \text{ a real number.} \tag{3.25}$$

Applying the definition of conditional probability we have:

$$p_X(x|C) = \frac{P[\{X = x\} \cap C]}{P[C]}. \tag{3.26}$$

The above expression has a nice intuitive interpretation: The conditional probability of the event $\{X = x_k\}$ is given by the probabilities of outcomes ζ for which *both $X(\zeta) = x_k$ and ζ are in C*, normalized by $P[C]$.

The conditional pmf satisfies Eqs. (3.4a) – (3.4c). Consider Eq. (3.4b). The set of events $A_k = \{X = x_k\}$ is a partition of S, so

$$C = \bigcup_k (A_k \cap C), \text{ and}$$

$$\sum_{x_k \in S_X} p_X(x_k|C) = \sum_{\text{all } k} p_X(x_k|C) = \sum_{\text{all } k} \frac{P[\{X = x_k\} \cap C]}{P[C]}$$

$$= \frac{1}{P[C]} \sum_{\text{all } k} P[A_k \cap C] = \frac{P[C]}{P[C]} = 1.$$
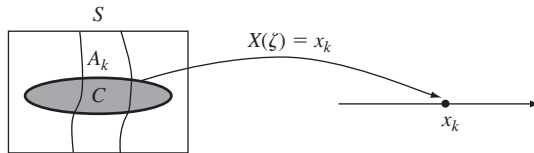
FIGURE 3.7
Conditional pmf of X given event C.

Similarly we can show that:

$$P[X \text{ in } B|C] = \sum_{x \in B} p_X(x|C) \text{ where } B \subset S_X.$$

Example 3.23   A Random Clock

The minute hand in a clock is spun and the outcome ζ is the minute where the hand comes to rest. Let X be the hour where the hand comes to rest. Find the pmf of X. Find the conditional pmf of X given $B = \{\text{first 4 hours}\}$; given $D = \{1 < \zeta \le 11\}$.

We assume that the hand is equally likely to rest at any of the minutes in the range $S = \{1, 2, \ldots, 60\}$, so $P[\zeta = k] = 1/60$ for k in S. X takes on values from $S_X = \{1, 2, \ldots, 12\}$ and it is easy to show that $p_X(j) = 1/12$ for j in S_X. Since $B = \{1, 2, 3, 4\}$:

$$p_X(j|B) = \frac{P[\{X = j\} \cap B]}{P[B]} = \frac{P[X \in \{j\} \cap \{1, 2, 3, 4\}]}{P[X \in \{1, 2, 3, 4\}]}$$

$$= \begin{cases} \dfrac{P[X = j]}{1/3} = \dfrac{1}{4} & \text{if } j \in \{1, 2, 3, 4\} \\ 0 & \text{otherwise.} \end{cases}$$

The event B above involves X only. The event D, however, is stated in terms of the outcomes in the underlying experiment (i.e., minutes not hours), so the probability of the intersection has to be expressed accordingly:

$$p_X(j|D) = \frac{P[\{X = j\} \cap D]}{P[D]} = \frac{P[\zeta : X(\zeta) = j \text{ and } \zeta \in \{2, \ldots, 11\}]}{P[\zeta \in \{2, \ldots, 11\}]}$$

$$= \begin{cases} \dfrac{P[\zeta \in \{2, 3, 4, 5\}]}{10/60} = \dfrac{4}{10} & \text{for } j = 1 \\ \dfrac{P[\zeta \in \{6, 7, 8, 9, 10\}]}{10/60} = \dfrac{5}{10} & \text{for } j = 2 \\ \dfrac{P[\zeta \in \{11\}]}{10/60} = \dfrac{1}{10} & \text{for } j = 3. \end{cases}$$

Most of the time the event C is defined in terms of X, for example $C = \{X > 10\}$ or $C = \{a \le X \le b\}$. For x_k in S_X, we have the following general result:

$$p_X(x_k|C) = \begin{cases} \dfrac{p_X(x_k)}{P[C]} & \text{if } x_k \in C \\ 0 & \text{if } x_k \notin C. \end{cases} \tag{3.27}$$

The above expression is determined entirely by the pmf of X.

Example 3.24   Residual Waiting Times

Let X be the time required to transmit a message, where X is a uniform random variable with $S_X = \{1, 2, \ldots, L\}$. Suppose that a message has already been transmitting for m time units, find the probability that the remaining transmission time is j time units.

We are given $C = \{X > m\}$, so for $m + 1 \le m + j \le L$:

$$p_X(m + j | X > m) = \frac{P[X = m + j]}{P[X > m]}$$

$$= \frac{\dfrac{1}{L}}{\dfrac{L - m}{L}} = \frac{1}{L - m} \qquad \text{for } m + 1 \le m + j \le L. \qquad (3.28)$$

X is equally likely to be any of the remaining $L - m$ possible values. As m increases, $1/(L - m)$ increases implying that the end of the message transmission becomes increasingly likely.

Many random experiments have natural ways of partitioning the sample space S into the union of disjoint events $B_1, B_2, \ldots, B_n$. Let $p_X(x | B_i)$ be the conditional pmf of X given event B_i. The theorem on total probability allows us to find the pmf of X in terms of the conditional pmf's:

$$p_X(x) = \sum_{i=1}^{n} p_X(x | B_i) P[B_i]. \qquad (3.29)$$

Example 3.25   Device Lifetimes

A production line yields two types of devices. Type 1 devices occur with probability α and work for a relatively short time that is geometrically distributed with parameter r. Type 2 devices work much longer, occur with probability $1 - \alpha$, and have a lifetime that is geometrically distributed with parameter s. Let X be the lifetime of an arbitrary device. Find the pmf of X.

The random experiment that generates X involves selecting a device type and then observing its lifetime. We can partition the sets of outcomes in this experiment into event B_1, consisting of those outcomes in which the device is type 1, and B_2, consisting of those outcomes in which the device is type 2. The conditional pmf's of X given the device type are:

$$p_{X|B_1}(k) = (1 - r)^{k-1} r \qquad \text{for } k = 1, 2, \ldots$$

and

$$p_{X|B_2}(k) = (1 - s)^{k-1} s \qquad \text{for } k = 1, 2, \ldots.$$

We obtain the pmf of X from Eq. (3.29):

$$p_X(k) = p_X(k | B_1) P[B_1] + p_X(k | B_2) P[B_2]$$

$$= (1 - r)^{k-1} r \alpha + (1 - s)^{k-1} s (1 - \alpha) \qquad \text{for } k = 1, 2, \ldots.$$

3.4.2   Conditional Expected Value

Let X be a discrete random variable, and suppose that we know that event B has occurred. The **conditional expected value of X given B** is defined as:

$$m_{X|B} = E[X | B] = \sum_{x \in S_X} x p_X(x | B) = \sum_k x_k p_X(x_k | B) \qquad (3.30)$$

where we apply the absolute convergence requirement on the summation. The **conditional variance of X given B** is defined as:

$$\text{VAR}[X|B] = E[(X - m_{X|B})^2|B] = \sum_{k=1}^{\infty} (x_k - m_{X|B})^2 p_X(x_k|B)$$

$$= E[X^2|B] - m_{X|B}^2.$$

Note that the variation is measured with respect to $m_{X|B}$, not m_X.

Let $B_1, B_2,..., B_n$ be the partition of S, and let $p_X(x|B_i)$ be the conditional pmf of X given event B_i. $E[X]$ can be calculated from the conditional expected values $E[X|B]$:

$$E[X] = \sum_{i=1}^{n} E[X|B_i]P[B_i]. \tag{3.31a}$$

By the theorem on total probability we have:

$$E[X] = \sum_{k} k p_X(x_k) = \sum_{k} k \left\{ \sum_{i=1}^{n} p_X(x_k|B_i)P[B_i] \right\}$$

$$= \sum_{i=1}^{n} \left\{ \sum_{k} k p_X(x_k|B_i) \right\} P[B_i] = \sum_{i=1}^{n} E[X|B_i]P[B_i],$$

where we first express $p_X(x_k)$ in terms of the conditional pmf's, and we then change the order of summation. Using the same approach we can also show

$$E[g(X)] = \sum_{i=1}^{n} E[g(X)|B_i]P[B_i]. \tag{3.31b}$$

Example 3.26 Device Lifetimes

Find the mean and variance for the devices in Example 3.25.

The conditional mean and second moment of each device type is that of a geometric random variable with the corresponding parameter:

$$m_{X|B_1} = 1/r \quad E[X^2|B_1] = (1 + r)/r^2$$

$$m_{X|B_2} = 1/s \quad E[X^2|B_2] = (1 + s)/s^2.$$

The mean and the second moment of X are then:

$$m_X = m_{X|B_1}\alpha + m_{X|B_2}(1 - \alpha) = \alpha/r + (1 - \alpha)/s$$

$$E[X^2] = E[X^2|B_1]\alpha + E[X^2|B_2](1 - \alpha) = \alpha(1 + r)/r^2 + (1 - \alpha)(1 + s)/s^2.$$

Finally, the variance of X is:

$$\text{VAR}[X] = E[X^2] - m_X^2 = \frac{\alpha(1 + r)}{r^2} + \frac{(1 - \alpha)(1 + s)}{s^2} - \left(\frac{\alpha}{r} + \frac{(1 - \alpha)}{s} \right)^2.$$

Note that we do *not* use the conditional variances to find VAR[Y] because Eq. (3.31b) does not apply to conditional variances. (See Problem 3.40.) However, the equation does apply to the conditional second moments.

3.5   IMPORTANT DISCRETE RANDOM VARIABLES

Certain random variables arise in many diverse, unrelated applications. The pervasiveness of these random variables is due to the fact that they model fundamental mechanisms that underlie random behavior. In this section we present the most important of the discrete random variables and discuss how they arise and how they are interrelated. Table 3.1 summarizes the basic properties of the discrete random variables discussed in this section. By the end of this chapter, most of these properties presented in the table will have been introduced.

TABLE 3.1   Discrete random variables

Bernoulli Random Variable

$S_X = \{0, 1\}$

$p_0 = q = 1 - p \qquad p_1 = p \qquad 0 \le p \le 1$

$E[X] = p \quad \text{VAR}[X] = p(1 - p) \qquad G_X(z) = (q + pz)$

Remarks: The Bernoulli random variable is the value of the indicator function I_A for some event A; $X = 1$ if A occurs and 0 otherwise.

Binomial Random Variable

$S_X = \{0, 1, \ldots, n\}$

$p_k = \binom{n}{k} p^k (1 - p)^{n-k} \qquad k = 0, 1, \ldots, n$

$E[X] = np \quad \text{VAR}[X] = np(1 - p) \qquad G_X(z) = (q + pz)^n$

Remarks: X is the number of successes in n Bernoulli trials and hence the sum of n independent, identically distributed Bernoulli random variables.

Geometric Random Variable

First Version: $S_X = \{0, 1, 2, \ldots\}$

$p_k = p(1 - p)^k \qquad k = 0, 1, \ldots$

$E[X] = \dfrac{1 - p}{p} \qquad \text{VAR}[X] = \dfrac{1 - p}{p^2} \qquad G_X(z) = \dfrac{p}{1 - qz}$

Remarks: X is the number of failures before the first success in a sequence of independent Bernoulli trials. The geometric random variable is the only discrete random variable with the memoryless property.

Second Version: $S_{X'} = \{1, 2, \ldots\}$

$p_k = p(1 - p)^{k-1} \qquad k = 1, 2, \ldots$

$E[X'] = \dfrac{1}{p} \qquad \text{VAR}[X'] = \dfrac{1 - p}{p^2} \qquad G_{X'}(z) = \dfrac{pz}{1 - qz}$

Remarks: $X' = X + 1$ is the number of trials until the first success in a sequence of independent Bernoulli trials.

<div align="right">(Continued)</div>

TABLE 3.1 *Continued*

Negative Binomial Random Variable

$S_X = \{r, r+1, \dots\}$ where r is a positive integer

$$p_k = \binom{k-1}{r-1} p^r (1-p)^{k-r} \qquad k = r, r+1, \dots$$

$$E[X] = \frac{r}{p} \qquad \text{VAR}[X] = \frac{r(1-p)}{p^2} \qquad G_X(z) = \left(\frac{pz}{1-qz}\right)^r$$

Remarks: X is the number of trials until the *r*th success in a sequence of independent Bernoulli trials.

Poisson Random Variable

$S_X = \{0, 1, 2, \dots\}$

$$p_k = \frac{\alpha^k}{k!} e^{-\alpha} \qquad k = 0, 1, \dots \quad \text{and } \alpha > 0$$

$$E[X] = \alpha \qquad \text{VAR}[X] = \alpha \qquad G_X(z) = e^{\alpha(z-1)}$$

Remarks: X is the number of events that occur in one time unit when the time between events is exponentially distributed with mean $1/\alpha$.

Uniform Random Variable

$S_X = \{1, 2, \dots, L\}$

$$p_k = \frac{1}{L} \qquad k = 1, 2, \dots, L$$

$$E[X] = \frac{L+1}{2} \qquad \text{VAR}[X] = \frac{L^2-1}{12} \qquad G_X(z) = \frac{z}{L} \frac{1-z^L}{1-z}$$

Remarks: The uniform random variable occurs whenever outcomes are equally likely. It plays a key role in the generation of random numbers.

Zipf Random Variable

$S_X = \{1, 2, \dots, L\}$ where L is a positive integer

$$p_k = \frac{1}{c_L} \frac{1}{k} \qquad k = 1, 2, \dots, L \text{ where } c_L \text{ is given by Eq. (3.45)}$$

$$E[X] = \frac{L}{c_L} \qquad \text{VAR}[X] = \frac{L(L+1)}{2c_L} - \frac{L^2}{c_L^2}$$

Remarks: The Zipf random variable has the property that a few outcomes occur frequently but most outcomes occur rarely.

Discrete random variables arise mostly in applications where counting is involved. We begin with the Bernoulli random variable as a model for a single coin toss. By counting the outcomes of multiple coin tosses we obtain the binomial, geometric, and Poisson random variables.

3.5.1   The Bernoulli Random Variable

Let A be an event related to the outcomes of some random experiment. The Bernoulli random variable I_A (defined in Example 3.8) equals one if the event A occurs, and zero otherwise. I_A is a discrete random variable since it assigns a number to each outcome of S. It is a discrete random variable with range $= \{0, 1\}$, and its pmf is

$$p_I(0) = 1 - p \quad \text{and} \quad p_I(1) = p, \tag{3.32}$$

where $P[A] = p$.

In Example 3.11 we found the mean of I_A:

$$m_I = E[I_A] = p.$$

The sample mean in n independent Bernoulli trials is simply the relative frequency of successes and converges to p as n increases:

$$\langle I_A \rangle_n = \frac{0 N_0(n) + 1 N_1(n)}{n} = f_1(n) \to p.$$

In Example 3.21 we found the variance of I_A:

$$\sigma_I^2 = \text{VAR}[I_A] = p(1 - p) = pq.$$

The variance is quadratic in p, with value zero at $p = 0$ and $p = 1$ and maximum at $p = 1/2$. This agrees with intuition since values of p close to 0 or to 1 imply a preponderance of successes or failures and hence less variability in the observed values. The maximum variability occurs when $p = 1/2$ which corresponds to the case that is most difficult to predict.

Every Bernoulli trial, regardless of the event A, is equivalent to the tossing of a biased coin with probability of heads p. In this sense, coin tossing can be viewed as representative of a fundamental mechanism for generating randomness, and the Bernoulli random variable is the model associated with it.

3.5.2   The Binomial Random Variable

Suppose that a random experiment is repeated n independent times. Let X be the number of times a certain event A occurs in these n trials. X is then a random variable with range $S_X = \{0, 1, \ldots, n\}$. For example, X could be the number of heads in n tosses of a coin. If we let I_j be the indicator function for the event A in the jth trial, then

$$X = I_1 + I_2 + \ldots + I_n,$$

that is, X is the sum of the Bernoulli random variables associated with each of the n independent trials.

In Section 2.6, we found that X has probabilities that depend on n and p:

$$P[X = k] = p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad \text{for } k = 0, \ldots, n. \tag{3.33}$$

X is called the **binomial random variable**. Figure 3.8 shows the pdf of X for $n = 24$ and $p = .2$ and $p = .5$. Note that $P[X = k]$ is maximum at $k_{\max} = [(n + 1)p]$, where $[x]$
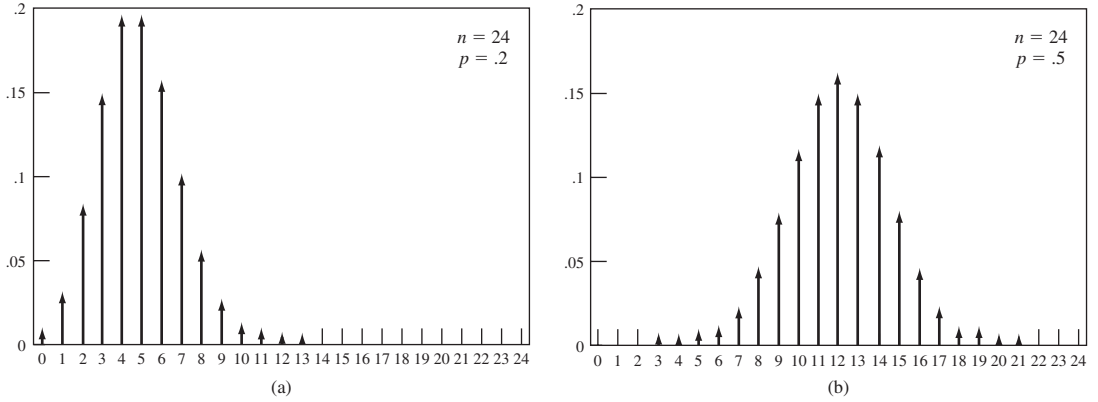
FIGURE 3.8
Probability mass functions of binomial random variable (a) $p = 0.2$; (b) $p = 0.5$.

denotes the largest integer that is smaller than or equal to x. When $(n + 1)p$ is an integer, then the maximum is achieved at k_{max} and $k_{max} - 1$. (See Problem 3.50.)

The factorial terms grow large very quickly and cause overflow problems in the calculation of $\binom{n}{k}$. We can use Eq. (2.40) for the ratio of successive terms in the pmf allows us to calculate $p_X(k + 1)$ in terms of $p_X(k)$ and delays the onset of overflows:

$$\frac{p_X(k + 1)}{p_X(k)} = \frac{n - k}{k + 1} \frac{p}{1 - p} \qquad \text{where } p_X(0) = (1 - p)^n. \qquad (3.34)$$

The binomial random variable arises in applications where there are two types of objects (i.e., heads/tails, correct/erroneous bits, good/defective items, active/silent speakers), and we are interested in the number of type 1 objects in a randomly selected batch of size n, where the type of each object is independent of the types of the other objects in the batch. Examples involving the binomial random variable were given in Section 2.6.

Example 3.27   Mean of a Binomial Random Variable

The expected value of X is:

$$E[X] = \sum_{k=0}^{n} k p_X(k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1 - p)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{k!(n - k)!} p^k (1 - p)^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n - 1)!}{(k - 1)!(n - k)!} p^{k-1} (1 - p)^{n-k}$$

$$= np \sum_{j=0}^{n-1} \frac{(n - 1)!}{j!(n - 1 - j)!} p^j (1 - p)^{n-1-j} = np, \qquad (3.35)$$

where the first line uses the fact that the $k = 0$ term in the sum is zero, the second line cancels out the k and factors np outside the summation, and the last line uses the fact that the summation is equal to one since it adds all the terms in a binomial pmf with parameters $n - 1$ and p.

The expected value $E[X] = np$ agrees with our intuition since we expect a fraction p of the outcomes to result in success.

Example 3.28   Variance of a Binomial Random Variable

To find $E[X^2]$ below, we remove the $k = 0$ term and then let $k' = k - 1$:

$$E[X^2] = \sum_{k=0}^{n} k^2 \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k}$$

$$= np \sum_{k'=0}^{n-1} (k'+1) \binom{n-1}{k'} p^{k'} (1-p)^{n-1-k}$$

$$= np \left\{ \sum_{k'=0}^{n-1} k' \binom{n-1}{k'} p^{k'} (1-p)^{n-1-k} + \sum_{k'=0}^{n-1} 1 \binom{n-1}{k'} p^{k'} (1-p)^{n-1-k'} \right\}$$

$$= np\{(n-1)p + 1\} = np(np + q).$$

In the third line we see that the first sum is the mean of a binomial random variable with parameters $(n-1)$ and p, and hence equal to $(n-1)p$. The second sum is the sum of the binomial probabilities and hence equal to 1.

We obtain the variance as follows:

$$\sigma_X^2 = E[X^2] - E[X]^2 = np(np + q) - (np)^2 = npq = np(1-p).$$

We see that the variance of the binomial is n times the variance of a Bernoulli random variable. We observe that values of p close to 0 or to 1 imply smaller variance, and that the maximum variability is when $p = 1/2$.

Example 3.29   Redundant Systems

A system uses triple redundancy for reliability: Three microprocessors are installed and the system is designed so that it operates as long as one microprocessor is still functional. Suppose that the probability that a microprocessor is still active after t seconds is $p = e^{-\lambda t}$. Find the probability that the system is still operating after t seconds.

Let X be the number of microprocessors that are functional at time t. X is a binomial random variable with parameter $n = 3$ and p. Therefore:

$$P[X \geq 1] = 1 - P[X = 0] = 1 - (1 - e^{-\lambda t})^3.$$

3.5.3   The Geometric Random Variable

The geometric random variable arises when we count the number M of independent Bernoulli trials until the first occurrence of a success. M is called the geometric random variable and it takes on values from the set $\{1, 2, \dots\}$. In Section 2.6, we found that the pmf of M is given by

$$P[M = k] = p_M(k) = (1-p)^{k-1} p \quad k = 1, 2, \dots, \tag{3.36}$$

where $p = P[A]$ is the probability of "success" in each Bernoulli trial. Figure 3.5(b) shows the geometric pmf for $p = 1/2$. Note that $P[M = k]$ decays geometrically with k, and that the ratio of consecutive terms is $p_M(k+1)/p_M(k) = (1-p) = q$. As p increases, the pmf decays more rapidly.

The probability that $M \leq k$ can be written in closed form:

$$P[M \leq k] = \sum_{j=1}^{k} pq^{j-1} = p \sum_{j'=0}^{k-1} q^{j'} = p\frac{1-q^k}{1-q} = 1 - q^k. \tag{3.37}$$

Sometimes we are interested in $M' = M - 1$, the number of *failures* before a success occurs. We also refer to M' as a geometric random variable. Its pmf is:

$$P[M' = k] = P[M = k + 1] = (1 - p)^k p \quad k = 0, 1, 2, \dots. \tag{3.38}$$

In Examples 3.15 and 3.22, we found the mean and variance of the geometric random variable:

$$m_M = E[M] = 1/p \qquad \text{VAR}[M] = \frac{1-p}{p^2}.$$

We see that the mean and variance increase as p, the success probability, decreases.

The geometric random variable is the only discrete random variable that satisfies the memoryless property:

$$P[M \geq k + j | M > j] = P[M \geq k] \quad \text{for all } j, k > 1.$$

(See Problems 3.54 and 3.55.) The above expression states that if a success has not occurred in the first j trials, then the probability of having to perform at least k more trials is the same as the probability of initially having to perform at least k trials. Thus, each time a failure occurs, the system "forgets" and begins anew as if it were performing the first trial.

The geometric random variable arises in applications where one is interested in the time (i.e., number of trials) that elapses between the occurrence of events in a sequence of independent experiments, as in Examples 2.11 and 2.43. Examples where the modified geometric random variable M' arises are: number of customers awaiting service in a queueing system; number of white dots between successive black dots in a scan of a black-and-white document.

3.5.4   The Poisson Random Variable

In many applications, we are interested in counting the number of occurrences of an event in a certain time period or in a certain region in space. The Poisson random variable arises in situations where the events occur "completely at random" in time or space. For example, the Poisson random variable arises in counts of emissions from radioactive substances, in counts of demands for telephone connections, and in counts of defects in a semiconductor chip.

The pmf for the **Poisson random variable** is given by

$$P[N = k] = p_N(k) = \frac{\alpha^k}{k!}e^{-\alpha} \qquad \text{for } k = 0, 1, 2, \dots, \tag{3.39}$$

where α is the average number of event occurrences in a specified time interval or region in space. Figure 3.9 shows the Poisson pmf for several values of α. For $\alpha < 1$, $P[N = k]$ is maximum at $k = 0$; for $\alpha > 1$, $P[N = k]$ is maximum at $[\alpha]$; if α is a positive integer, the $P[N = k]$ is maximum at $k = \alpha$ and at $k = \alpha - 1$.

(a)

(b)

(c)
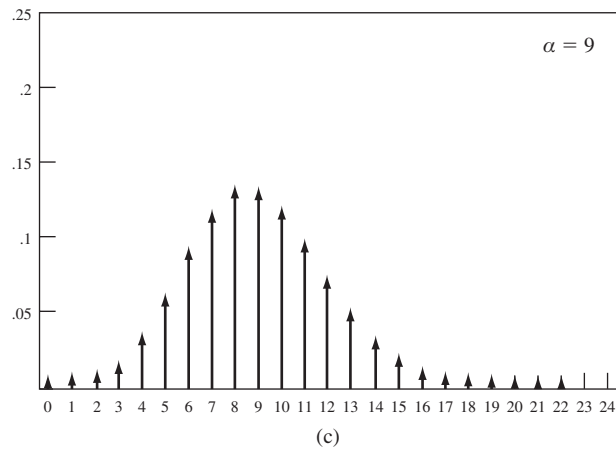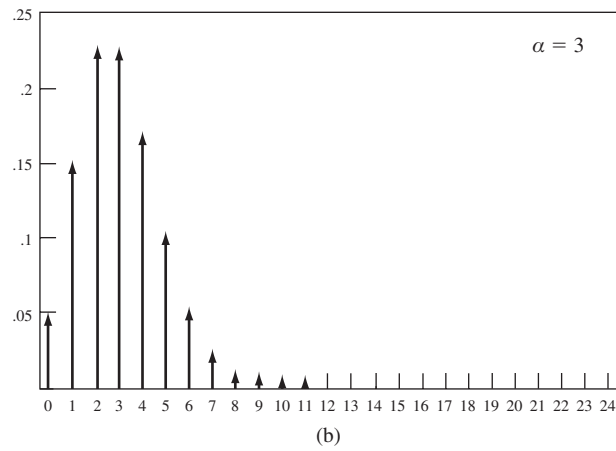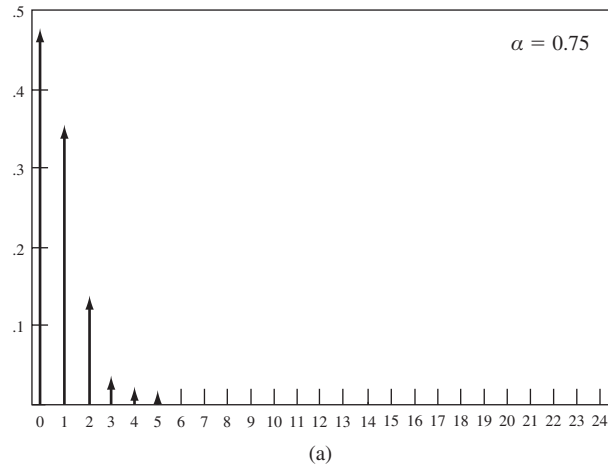
FIGURE 3.9
Probability mass functions of Poisson random variable (a) $\alpha = 0.75$;
(b) $\alpha = 3$; (c) $\alpha = 9$.

The pmf of the Poisson random variable sums to one, since

$$\sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{-\alpha} = e^{-\alpha} \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = e^{-\alpha} e^{\alpha} = 1,$$

where we used the fact that the second summation is the infinite series expansion for e^{α}.

It is easy to show that the mean and variance of a Poisson random variable is given by:

$$E[N] = \alpha \quad \text{and} \quad \sigma_N^2 = \text{VAR}[N] = \alpha.$$

Example 3.30   Queries at a Call Center

The number N of queries arriving in t seconds at a call center is a Poisson random variable with $\alpha = \lambda t$ where λ is the average arrival rate in queries/second. Assume that the arrival rate is four queries per minute. Find the probability of the following events: (a) more than 4 queries in 10 seconds; (b) fewer than 5 queries in 2 minutes.

The arrival rate in queries/second is $\lambda = 4$ queries/60 sec $= 1/15$ queries/sec. In part a, the time interval is 10 seconds, so we have a Poisson random variable with $\alpha = (1/15$ queries/sec$) *$ 10 seconds $= 10/15$ queries. The probability of interest is evaluated numerically:

$$P[N > 4] = 1 - P[N \leq 4] = 1 - \sum_{k=0}^{4} \frac{(2/3)^k}{k!} e^{-2/3} = 6.33(10^{-4}).$$

In part b, the time interval of interest is $t = 120$ seconds, so $\alpha = 1/15 * 120$ seconds $= 8$. The probability of interest is:

$$P[N \leq 5] = \sum_{k=0}^{5} \frac{(8)^k}{k!} e^{-8} = 0.10.$$

Example 3.31   Arrivals at a Packet Multiplexer

The number N of packet arrivals in t seconds at a multiplexer is a Poisson random variable with $\alpha = \lambda t$ where λ is the average arrival rate in packets/second. Find the probability that there are no packet arrivals in t seconds.

$$P[N = 0] = \frac{\alpha^0}{0!} e^{-\lambda t} = e^{-\lambda t}.$$

This equation has an interesting interpretation. Let Z be the time until the first packet arrival. Suppose we ask, "What is the probability that $X > t$, that is, the next arrival occurs t or more seconds later?" Note that $\{N = 0\}$ implies $\{Z > t\}$ and vice versa, so $P[Z > t] = e^{-\lambda t}$. The probability of no arrival decreases exponentially with t.

Note that we can also show that

$$P[N(t) \geq n] = 1 - P[N(t) < n] = 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

One of the applications of the Poisson probabilities in Eq. (3.39) is to approximate the binomial probabilities in the case where p is very small and n is very large,

that is, where the event A of interest is very rare but the number of Bernoulli trials is very large. We show that *if* $\alpha = np$ is fixed, then as n becomes *large*:

$$p_k = \binom{n}{k}p^k(1 - p)^{n-k} \simeq \frac{\alpha^k}{k!}e^{-\alpha} \qquad \text{for } k = 0, 1,\dots. \tag{3.40}$$

Equation (3.40) is obtained by taking the limit $n \to \infty$ in the expression for p_k, while keeping $\alpha = np$ fixed. First, consider the probability that no events occur in n trials:

$$p_0 = (1 - p)^n = \left(1 - \frac{\alpha}{n}\right)^n \to e^{-\alpha} \quad \text{as } n \to \infty, \tag{3.41}$$

where the limit in the last expression is a well known result from calculus. Consider the ratio of successive binomial probabilities:

$$\frac{p_{k+1}}{p_k} = \frac{(n - k)p}{(k + 1)q} = \frac{(1 - k/n)\alpha}{(k + 1)(1 - \alpha/n)}$$

$$\to \frac{\alpha}{k + 1} \quad \text{as } n \to \infty.$$

Thus the limiting probabilities satisfy

$$p_{k+1} = \frac{\alpha}{k + 1}p_k = \left(\frac{\alpha}{k + 1}\right)\left(\frac{\alpha}{k}\right)\cdots\left(\frac{\alpha}{1}\right)p_0 = \frac{\alpha^k}{k!}e^{-\alpha}. \tag{3.42}$$

Thus the Poisson pmf can be used to approximate the binomial pmf for large n and small p, using $\alpha = np$.

Example 3.32   Errors in Optical Transmission

An optical communication system transmits information at a rate of 10^9 bits/second. The probability of a bit error in the optical communication system is 10^{-9}. Find the probability of five or more errors in 1 second.

   Each bit transmission corresponds to a Bernoulli trial with a "success" corresponding to a bit error in transmission. The probability of k errors in $n = 10^9$ transmissions (1 second) is then given by the binomial probability with $n = 10^9$ and $p = 10^{-9}$. The Poisson approximation uses $\alpha = np = 10^9(10^{-9}) = 1$. Thus

$$P[N \geq 5] = 1 - P[N < 5] = 1 - \sum_{k=0}^{4}\frac{\alpha^k}{k!}e^{-\alpha}$$

$$= 1 - e^{-1}\left\{1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!}\right\} = .00366.$$

   The Poisson random variable appears in numerous physical situations because many models are very large in scale and involve very rare events. For example, the Poisson pmf gives an accurate prediction for the relative frequencies of the number of particles emitted by a radioactive mass during a fixed time period. This correspondence can be explained as follows. A radioactive mass is composed of a large number of atoms, say n. In a fixed time interval each atom has a very small probability p of disintegrating and emitting a radioactive particle. If atoms disintegrate independently of
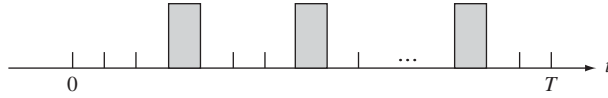
FIGURE 3.10
Event occurrences in *n* subintervals of [0, *T*].

other atoms, then the number of emissions in a time interval can be viewed as the number of successes in n trials. For example, one microgram of radium contains about $n = 10^{16}$ atoms, and the probability that a single atom will disintegrate during a one-millisecond time interval is $p = 10^{-15}$ [Rozanov, p. 58]. Thus it is an understatement to say that the conditions for the approximation in Eq. (3.40) hold: n is so large and p so small that one could argue that the limit $n \to \infty$ has been carried out and that the number of emissions is *exactly* a Poisson random variable.

The Poisson random variable also comes up in situations where we can imagine a sequence of Bernoulli trials taking place in time or space. Suppose we count the number of event occurrences in a T-second interval. Divide the time interval into a very large number, n, of subintervals as shown in Fig. 3.10. A pulse in a subinterval indicates the occurrence of an event. Each subinterval can be viewed as one in a sequence of independent Bernoulli trials if the following conditions hold: (1) At most one event can occur in a subinterval, that is, the probability of more than one event occurrence is negligible; (2) the outcomes in different subintervals are independent; and (3) the probability of an event occurrence in a subinterval is $p = \alpha/n$, where α is the average number of events observed in a 1-second interval. The number N of events in 1 second is a binomial random variable with parameters n and $p = \alpha/n$. Thus as $n \to \infty$, N becomes a Poisson random variable with parameter α. In Chapter 9 we will revisit this result when we discuss the Poisson random process.

3.5.5 The Uniform Random Variable

The discrete uniform random variable Y takes on values in a set of consecutive integers $S_Y = \{j + 1, \ldots, j + L\}$ with equal probability:

$$p_Y(k) = \frac{1}{L} \quad \text{for} \quad k \in \{j + 1, \ldots, j + L\}. \tag{3.43}$$

This humble random variable occurs whenever outcomes are equally likely, e.g., toss of a fair coin or a fair die, spinning of an arrow in a wheel divided into equal segments, selection of numbers from an urn. It is easy to show that the mean and variance are:

$$E[Y] = j + \frac{L + 1}{2} \quad \text{and} \quad \text{VAR}[Y] = \frac{L^2 - 1}{12}.$$

Example 3.33 Discrete Uniform Random Variable in Unit Interval

Let X be a uniform random variable in $S_X = \{0, 1, \ldots, L - 1\}$. We define the discrete uniform random variable in the unit interval by

$$U = \frac{X}{L} \quad \text{so} \quad S_U = \left\{ 0, \frac{1}{L}, \frac{2}{L}, \frac{3}{L}, \ldots, 1 - \frac{1}{L} \right\}.$$

U has pmf:

$$p_U\left(\frac{k}{L}\right) = \frac{1}{L} \qquad \text{for } k = 0, 2, \ldots, L - 1.$$

The pmf of U puts equal probability mass $1/L$ on equally spaced points $x_k = k/L$ in the unit interval. The probability of a subinterval of the unit interval is equal to the number of points in the subinterval multiplied by $1/L$. As L becomes very large, this probability is essentially the length of the subinterval.

3.5.6   The Zipf Random Variable

The Zipf random variable is named for George Zipf who observed that the frequency of words in a large body of text is proportional to their rank. Suppose that words are ranked from most frequent, to next most frequent, and so on. Let X be the rank of a word, then $S_X = \{1, 2, \ldots, L\}$ where L is the number of distinct words. The pmf of X is:

$$p_X(k) = \frac{1}{c_L}\frac{1}{k} \qquad \text{for } k = 1, 2, \ldots, L. \tag{3.44}$$

where c_L is a normalization constant. The second word has $1/2$ the frequency of occurrence as the first, the third word has $1/3$ the frequency of the first, and so on. The normalization constant c_L is given by the sum:

$$c_L = \sum_{j=1}^{L}\frac{1}{j} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{L} \tag{3.45}$$

The constant c_L occurs frequently in calculus and is called the Lth harmonic mean and increases approximately as $\ln L$. For example, for $L = 100$, $c_L = 5.187378$ and $c_L - \ln(L) = 0.582207$. It can be shown that as $L \to \infty$, $c_L - \ln L \to 0.57721 \ldots$.

The mean of X is given by:

$$E[X] = \sum_{j=1}^{L} j p_X(j) = \sum_{j=1}^{L} j \frac{1}{c_L j} = \frac{L}{c_L}. \tag{3.46}$$

The second moment and variance of X are:

$$E[X^2] = \sum_{j=1}^{L} j^2 \frac{1}{c_L j} = \frac{1}{c_L}\sum_{j=1}^{L} j = \frac{L(L+1)}{2c_L}$$

and

$$\text{VAR}[X] = \frac{L(L+1)}{2c_L} - \frac{L^2}{c_L^2}. \tag{3.47}$$

The Zipf and related random variables have gained prominence with the growth of the Internet where they have been found in a variety of measurement studies involving Web page sizes, Web access behavior, and Web page interconnectivity. These random variables had previously been found extensively in studies on the distribution of wealth and, not surprisingly, are now found in Internet video rentals and book sales.
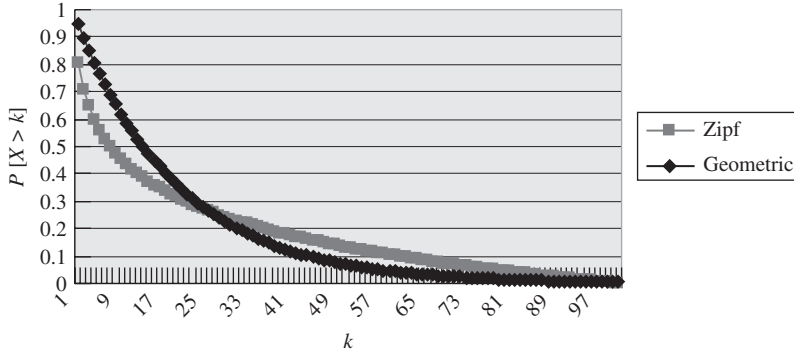
FIGURE 3.11
Zipf distribution and its long tail.
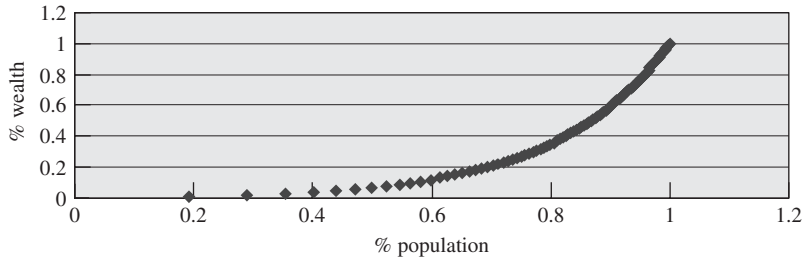
FIGURE 3.12
Lorenz curve for Zipf random variable with $L = 100$.

Example 3.34 Rare Events and Long Tails

The Zipf random variable X has the property that a few outcomes (words) occur frequently but most outcomes occur rarely. Find the probability of words with rank higher than m.

$$P[X > m] = 1 - P[X \leq m] = 1 - \frac{1}{c_L} \sum_{j=1}^{m} \frac{1}{j} = 1 - \frac{c_m}{c_L} \qquad \text{for } m \leq L. \qquad (3.48)$$

We call $P[X > m]$ the probability of the tail of the distribution of X. Figure 3.11 shows the $P[X > m]$ with $L = 100$ which has $E[X] = 100/c_{100} = 19.28$. Figure 3.12 also shows $P[Y > m]$ for a geometric random variable with the same mean, that is, $1/p = 19.28$. It can be seen that $P[Y > m]$ for the geometric random variable drops off much more quickly than $P[X > m]$. The Zipf distribution is said to have a "long tail" because rare events are more likely to occur than in traditional probability models.

Example 3.35 80/20 Rule and the Lorenz Curve

Let X correspond to a level of wealth and $p_X(k)$ be the proportion of a population that has wealth k. Suppose that X is a Zipf random variable. Thus $p_X(1)$ is the proportion of the population with wealth 1, $p_X(2)$ the proportion with wealth 2, and so on. The long tail of the Zipf distribution suggests that very rich individuals are not very rare. We frequently hear statements such as "20% of the population owns 80% of the wealth." The *Lorenz curve* plots the proportion

of wealth owned by the poorest fraction x of the population, as the x varies from 0 to 1. Find the Lorenz curve for $L = 100$.

For k in $\{1, 2, \ldots, L\}$, the fraction of the population with wealth k or less is:

$$F_k = P[X \leq k] = \frac{1}{c_L} \sum_{j=1}^{k} \frac{1}{j} = \frac{c_k}{c_L}. \tag{3.49}$$

The proportion of wealth owned by the population that has wealth k or less is:

$$W_k = \frac{\sum_{j=1}^{k} j p_X(j)}{\sum_{i=1}^{L} i p_X(i)} = \frac{\frac{1}{c_L} \sum_{j=1}^{k} j \frac{1}{j}}{\frac{1}{c_L} \sum_{i=1}^{L} i \frac{1}{i}} = \frac{k}{L}. \tag{3.50}$$

The denominator in the above expression is the total wealth of the entire population. The Lorenz curve consists of the plot of points (F_k, W_k) which is shown in Fig. 3.12 for $L = 100$. In the graph the 70% poorest proportion of the population own only 20% of the total wealth, or conversely, the 30% wealthiest fraction of the population owns 80% of the wealth. See Problem 3.75 for a discussion of what the Lorenz curve should look like in the cases of extreme fairness and extreme unfairness.

The explosive growth in the Internet has led to systems of huge scale. For probability models this growth has implied random variables that can attain very large values. Measurement studies have revealed many instances of random variables with long tail distributions.

If we try to let L approach infinity in Eq. (3.45), c_L grows without bound since the series does not converge. However, if we make the pmf proportional to $(1/k)^\alpha$ then the series converges as long as $\alpha > 1$. We define the **Zipf or zeta random variable** with range $\{1, 2, 3, \ldots\}$ to have pmf:

$$p_Z(k) = \frac{1}{z_\alpha} \frac{1}{k^\alpha} \qquad \text{for } k = 1, 2, \ldots, \tag{3.51}$$

where z_α is a normalization constant given by the zeta function which is defined by:

$$z_\alpha = \sum_{j=1}^{\infty} \frac{1}{j^\alpha} = 1 + \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \ldots \qquad \text{for } \alpha > 1. \tag{3.52}$$

The convergence of the above series is discussed in standard calculus books.

The mean of Z is given by:

$$E[Z] = \sum_{j=1}^{L} j p_Z(j) = \sum_{j=1}^{L} j \frac{1}{z_\alpha j^\alpha} = \frac{1}{z_\alpha} \sum_{j=1}^{L} \frac{1}{j^{\alpha-1}} = \frac{z_{\alpha-1}}{z_\alpha} \qquad \text{for } \alpha > 2,$$

where the sum of the sequence $1/j^{\alpha-1}$ converges only if $\alpha - 1 > 1$, that is, $\alpha > 2$. We can similarly show that the second moment (and hence the variance) exists only if $\alpha > 3$.

3.6 GENERATION OF DISCRETE RANDOM VARIABLES

Suppose we wish to generate the outcomes of a random experiment that has sample space $S = \{a_1, a_2, \ldots, a_n\}$ with probability of elementary events $p_j = P[\{a_j\}]$. We divide the unit interval into n subintervals. The jth subinterval has length p_j and
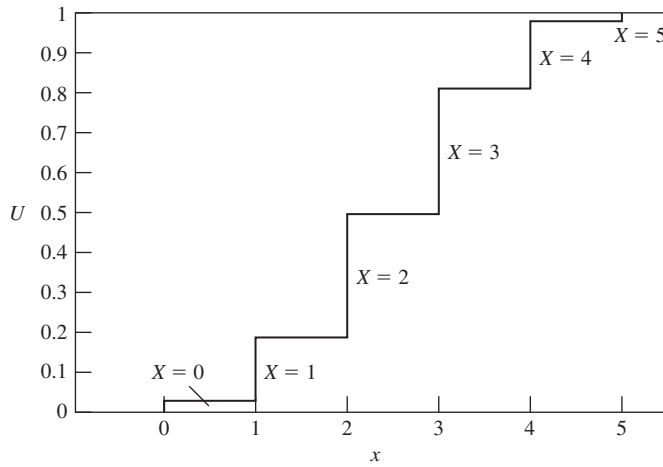
FIGURE 3.13
Generating a binomial random variable with $n = 5$, $p = 1/2$.

corresponds to outcome a_j. Each trial of the experiment first uses `rand` to obtain a number U in the unit interval. The outcome of the experiment is a_j if U is in the jth subinterval. Figure 3.13 shows the portioning of the unit interval according to the pmf of an $n = 5$, $p = 0.5$ binomial random variable.

The Octave function `discrete_rnd` implements the above method and can be used to generate random numbers with desired probabilities. Functions to generate random numbers with common distributions are also available. For example, `poisson_rnd (lambda, r, c)` can be used to generate an array of Poisson-distributed random numbers with rate lambda.

Example 3.36   Generation of Tosses of a Die

Use `discrete_rnd` to generate 20 samples of a toss of a die.

```
> V=1:6;                          % Define S_X = {1, 2, 3, 4, 5, 6}.

> P=[1/6, 1/6, 1/6, 1/6, 1/6, 1/6];   % Set all the pmf values for X to 1/6.

> discrete_rnd (20, V, P)         % Generate 20 samples from S_X with pmf P.

ans =

    6  2  2  6  5  2  6  1  3  6  3  1  6  3  4  2  5  3  4  1
```

Example 3.37   Generation of Poisson Random Variable

Use the built-in function to generate 20 samples of a Poisson random variable with $\alpha = 2$.

```
> Poisson_rnd (2,1,20)            % Generate a 1 × 20 array of samples of a Poisson
                                  % random variable with α = 2.

ans =
    4  3  0  2  3  2  1  2  1  4  0  1  2  2  3  4  0  1  3
```

The problems at the end of the chapter elaborate on the rich set of experiments that can be simulated using these basic capabilities of MATLAB or Octave. In the remainder of this book, we will use Octave in examples because it is freely available.

SUMMARY

- A random variable is a function that assigns a real number to each outcome of a random experiment. A random variable is defined if the outcome of a random experiment is a number, or if a numerical attribute of an outcome is of interest.
- The notion of an equivalent event enables us to derive the probabilities of events involving a random variable in terms of the probabilities of events involving the underlying outcomes.
- A random variable is discrete if it assumes values from some countable set. The probability mass function is sufficient to calculate the probability of all events involving a discrete random variable.
- The probability of events involving discrete random variable X can be expressed as the sum of the probability mass function $p_X(x)$.
- If X is a random variable, then $Y = g(X)$ is also a random variable.
- The mean, variance, and moments of a discrete random variable summarize some of the information about the random variable X. These parameters are useful in practice because they are easier to measure and estimate than the pmf.
- The conditional pmf allows us to calculate the probability of events given partial information about the random variable X.
- There are a number of methods for generating discrete random variables with prescribed pmf's in terms of a random variable that is uniformly distributed in the unit interval.

CHECKLIST OF IMPORTANT TERMS

Discrete random variable

Equivalent event

Expected value of X

Function of a random variable

nth moment of X

Probability mass function

Random variable

Standard deviation of X

Variance of X

ANNOTATED REFERENCES

Reference [1] is the standard reference for electrical engineers for the material on random variables. Reference [2] discusses some of the finer points regarding the concepts of a random variable at a level accessible to students of this course. Reference [3] is a classic text, rich in detailed examples. Reference [4] presents detailed discussions of the various methods for generating random numbers with specified distributions. Reference [5] is entirely focused on discrete random variables.

1. A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed., McGraw-Hill, New York, 2002.
2. K. L. Chung, *Elementary Probability Theory*, Springer-Verlag, New York, 1974.
3. W. Feller, *An Introduction to Probability Theory and Its Applications*, Wiley, New York, 1968.

4. A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, New York, 2000.
5. N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate Discrete Distributions*, Wiley, New York, 2005.
6. Y. A. Rozanov, *Probability Theory: A Concise Course*, Dover Publications, New York, 1969.

PROBLEMS

Section 3.1: The Notion of a Random Variable

3.1. Let X be the maximum of the number of heads obtained when Carlos and Michael each flip a fair coin twice.

   (a) Describe the underlying space S of this random experiment and specify the probabilities of its elementary events.

   (b) Show the mapping from S to S_X, the range of X.

   (c) Find the probabilities for the various values of X.

3.2. A die is tossed and the random variable X is defined as the number of full pairs of dots in the face showing up.

   (a) Describe the underlying space S of this random experiment and specify the probabilities of its elementary events.

   (b) Show the mapping from S to S_X, the range of X.

   (c) Find the probabilities for the various values of X.

   (d) Repeat parts a, b, and c, if Y is the number of full or partial pairs of dots in the face showing up.

   (e) Explain why $P[X = 0]$ and $P[Y = 0]$ are not equal.

3.3. The loose minute hand of a clock is spun hard. The coordinates (x, y) of the point where the tip of the hand comes to rest is noted. Z is defined as the sgn function of the product of x and y, where sgn(t) is 1 if $t > 0$, 0 if $t = 0$, and -1 if $t < 0$.

   (a) Describe the underlying space S of this random experiment and specify the probabilities of its events.

   (b) Show the mapping from S to S_X, the range of X.

   (c) Find the probabilities for the various values of X.

3.4. A data source generates hexadecimal characters. Let X be the integer value corresponding to a hex character. Suppose that the four binary digits in the character are independent and each is equally likely to be 0 or 1.

   (a) Describe the underlying space S of this random experiment and specify the probabilities of its elementary events.

   (b) Show the mapping from S to S_X, the range of X.

   (c) Find the probabilities for the various values of X.

   (d) Let Y be the integer value of a hex character but suppose that the most significant bit is three times as likely to be a "0" as a "1". Find the probabilities for the values of Y.

3.5. Two transmitters send messages through bursts of radio signals to an antenna. During each time slot each transmitter sends a message with probability 1/2. Simultaneous transmissions result in loss of the messages. Let X be the number of time slots until the first message gets through.

    (a) Describe the underlying sample space S of this random experiment and specify the probabilities of its elementary events.

    (b) Show the mapping from S to S_X, the range of X.

    (c) Find the probabilities for the various values of X.

3.6. An information source produces binary triplets $\{000, 111, 010, 101, 001, 110, 100, 011\}$ with corresponding probabilities $\{1/4, 1/4, 1/8, 1/8, 1/16, 1/16, 1/16, 1/16\}$. A binary code assigns a codeword of length $-\log_2 p_k$ to triplet k. Let X be the length of the string assigned to the output of the information source.

    (a) Show the mapping from S to S_X, the range of X.

    (b) Find the probabilities for the various values of X.

3.7. An urn contains 9 \$1 bills and one \$50 bill. Let the random variable X be the total amount that results when two bills are drawn from the urn without replacement.

    (a) Describe the underlying space S of this random experiment and specify the probabilities of its elementary events.

    (b) Show the mapping from S to S_X, the range of X.

    (c) Find the probabilities for the various values of X.

3.8. An urn contains 9 \$1 bills and one \$50 bill. Let the random variable X be the total amount that results when two bills are drawn from the urn *with* replacement.

    (a) Describe the underlying space S of this random experiment and specify the probabilities of its elementary events.

    (b) Show the mapping from S to S_X, the range of X.

    (c) Find the probabilities for the various values of X.

3.9. A coin is tossed n times. Let the random variable Y be the difference between the number of heads and the number of tails in the n tosses of a coin. Assume $P[\text{heads}] = p$.

    (a) Describe the sample space of S.

    (b) Find the probability of the event $\{Y = 0\}$.

    (c) Find the probabilities for the other values of Y.

3.10. An m-bit password is required to access a system. A hacker systematically works through all possible m-bit patterns. Let X be the number of patterns tested until the correct password is found.

    (a) Describe the sample space of S.

    (b) Show the mapping from S to S_X, the range of X.

    (c) Find the probabilities for the various values of X.

Section 3.2: Discrete Random Variables and Probability Mass Function

3.11. Let X be the maximum of the coin tosses in Problem 3.1.

    (a) Compare the pmf of X with the pmf of Y, the number of heads in two tosses of a fair coin. Explain the difference.

    (b) Suppose that Carlos uses a coin with probability of heads $p = 3/4$. Find the pmf of X.

3.12. Consider an information source that produces binary pairs that we designate as $S_X = \{1, 2, 3, 4\}$. Find and plot the pmf in the following cases:

    (a) $p_k = p_1/k$ for all k in S_X.

    (b) $p_{k+1} = p_k/2$ for $k = 2, 3, 4$.

(c) $p_{k+1} = p_k/2^k$ for $k = 2, 3, 4$.

(d) Can the random variables in parts a, b, and c be extended to take on values in the set $\{1, 2, \ldots\}$? If yes, specify the pmf of the resulting random variables. If no, explain why not.

3.13. Let X be a random variable with pmf $p_k = c/k^2$ for $k = 1, 2, \ldots$.

    (a) Estimate the value of c numerically. Note that the series converges.

    (b) Find $P[X > 4]$.

    (c) Find $P[6 \leq X \leq 8]$.

3.14. Compare $P[X \geq 8]$ and $P[Y \geq 8]$ for outputs of the data source in Problem 3.4.

3.15. In Problem 3.5 suppose that terminal 1 transmits with probability 1/2 in a given time slot, but terminal 2 transmits with probability p.

    (a) Find the pmf for the number of transmissions X until a message gets through.

    (b) Given a successful transmission, find the probability that terminal 2 transmitted.

3.16. **(a)** In Problem 3.7 what is the probability that the amount drawn from the urn is more than $2? More than $50?

    (b) Repeat part a for Problem 3.8.

3.17. A modem transmits a +2 voltage signal into a channel. The channel adds to this signal a noise term that is drawn from the set $\{0, -1, -2, -3\}$ with respective probabilities $\{4/10, 3/10, 2/10, 1/10\}$.

    (a) Find the pmf of the output Y of the channel.

    (b) What is the probability that the output of the channel is equal to the input of the channel?

    (c) What is the probability that the output of the channel is positive?

3.18. A computer reserves a path in a network for 10 minutes. To extend the reservation the computer must successfully send a "refresh" message before the expiry time. However, messages are lost with probability 1/2. Suppose that it takes 10 seconds to send a refresh request and receive an acknowledgment. When should the computer start sending refresh messages in order to have a 99% chance of successfully extending the reservation time?

3.19. A modem transmits over an error-prone channel, so it repeats every "0" or "1" bit transmission five times. We call each such group of five bits a "codeword." The channel changes an input bit to its complement with probability $p = 1/10$ and it does so independently of its treatment of other input bits. The modem receiver takes a majority vote of the five received bits to estimate the input signal. Find the probability that the receiver makes the wrong decision.

3.20. Two dice are tossed and we let X be the difference in the number of dots facing up.

    (a) Find and plot the pmf of X.

    (b) Find the probability that $|X| \leq k$ for all k.

Section 3.3: Expected Value and Moments of Discrete Random Variable

3.21. **(a)** In Problem 3.11, compare $E[Y]$ to $E[X]$ where X is the maximum of coin tosses.

    (b) Compare VAR[X] and VAR[Y].

3.22. Find the expected value and variance of the output of the information sources in Problem 3.12, parts a, b, and c.

3.23. **(a)** Find $E[X]$ for the hex integers in Problem 3.4.

    (b) Find VAR[X].

3.24. Find the mean codeword length in Problem 3.6. How can this average be interpreted in a very large number of encodings of binary triplets?

3.25. **(a)** Find the mean and variance of the amount drawn from the urn in Problem 3.7.

   (b) Find the mean and variance of the amount drawn from the urn in Problem 3.8.

3.26. Find $E[Y]$ and $\text{VAR}[Y]$ for the difference between the number of heads and tails in Problem 3.9. In a large number of repetitions of this random experiment, what is the meaning of $E[Y]$?

3.27. Find $E[X]$ and $\text{VAR}[X]$ in Problem 3.13.

3.28. Find the expected value and variance of the modem signal in Problem 3.17.

3.29. Find the mean and variance of the time that it takes to renew the reservation in Problem 3.18.

3.30. The modem in Problem 3.19 transmits 1000 5-bit codewords. What is the average number of codewords in error? If the modem transmits 1000 bits individually without repetition, what is the average number of bits in error? Explain how error rate is traded off against transmission speed.

3.31. **(a)** Suppose a fair coin is tossed n times. Each coin toss costs d dollars and the reward in obtaining X heads is $aX^2 + bX$. Find the expected value of the net reward.

   (b) Suppose that the reward in obtaining X heads is a^X, where $a > 0$. Find the expected value of the reward.

3.32. Let $g(X) = I_A$, where $A = \{X > 10\}$.

   (a) Find $E[g(X)]$ for X as in Problem 3.12a with $S_X = \{1, 2, \ldots, 15\}$.

   (b) Repeat part a for X as in Problem 3.12b with $S_X = \{1, 2, \ldots, 15\}$.

   (c) Repeat part a for X as in Problem 3.12c with $S_X = \{1, 2, \ldots, 15\}$.

3.33. Let $g(X) = (X - 10)^+$ (see Example 3.19).

   (a) Find $E[X]$ for X as in Problem 3.12a with $S_X = \{1, 2, \ldots, 15\}$.

   (b) Repeat part a for X as in Problem 3.12b with $S_X = \{1, 2, \ldots, 15\}$.

   (c) Repeat part a for X as in Problem 3.12c with $S_X = \{1, 2, \ldots, 15\}$.

3.34. Consider the St. Petersburg Paradox in Example 3.16. Suppose that the casino has a total of $M = 2^m$ dollars, and so it can only afford a finite number of coin tosses.

   (a) How many tosses can the casino afford?

   (b) Find the expected payoff to the player.

   (c) How much should a player be willing to pay to play this game?

Section 3.4: Conditional Probability Mass Function

3.35. **(a)** In Problem 3.11a, find the conditional pmf of X, the maximum of coin tosses, given that $X > 0$.

   (b) Find the conditional pmf of X given that Michael got one head in two tosses.

   (c) Find the conditional pmf of X given that Michael got one head in the first toss.

   (d) In Problem 3.11b, find the probability that Carlos got the maximum given that $X = 2$.

3.36. Find the conditional pmf for the quaternary information source in Problem 3.12, parts a, b, and c given that $X < 4$.

3.37. **(a)** Find the conditional pmf of the hex integer X in Problem 3.4 given that $X < 8$.

   (b) Find the conditional pmf of X given that the first bit is 0.

   (c) Find the conditional pmf of X given that the 4th bit is 0.

3.38. **(a)** Find the conditional pmf of X in Problem 3.5 given that no message gets through in time slot 1.

   (b) Find the conditional pmf of X given that the first transmitter transmitted in time slot 1.

3.39. (a) Find the conditional expected value of X in Problem 3.5 given that no message gets through in the first time slot. Show that $E[X \mid X > 1] = E[X] + 1$.

  (b) Find the conditional expected value of X in Problem 3.5 given that a message gets through in the first time slot.

  (c) Find $E[X]$ by using the results of parts a and b.

  (d) Find $E[X^2]$ and VAR$[X]$ using the approach in parts b and c.

3.40. Explain why Eq. (3.31b) can be used to find $E[X^2]$, but it cannot be used to directly find VAR$[X]$.

3.41. (a) Find the conditional pmf for X in Problem 3.7 given that the first draw produced k dollars.

  (b) Find the conditional expected value corresponding to part a.

  (c) Find $E[X]$ using the results from part b.

  (d) Find $E[X^2]$ and VAR$[X]$ using the approach in parts b and c.

3.42. Find $E[Y]$ and VAR$[Y]$ for the difference between the number of heads and tails in n tosses in Problem 3.9. *Hint:* Condition on the number of heads.

3.43. (a) In Problem 3.10 find the conditional pmf of X given that the password has not been found after k tries.

  (b) Find the conditional expected value of X given $X > k$.

  (c) Find $E[X]$ from the results in part b.

Section 3.5: Important Discrete Random Variables

3.44. Indicate the value of the indicator function for the event A, $I_A(\zeta)$, for each ζ in the sample space S. Find the pmf and expected of I_A.

  (a) $S = \{1, 2, 3, 4, 5\}$ and $A = \{\zeta > 3\}$.

  (b) $S = [0, 1]$ and $A = \{0.3 < \zeta \leq 0.7\}$.

  (c) $S = \{\zeta = (x, y): 0 < x < 1, 0 < y < 1\}$ and
  $A = \{\zeta = (x, y): 0.25 < x + y < 1.25\}$.

  (d) $S = (-\infty, \infty)$ and $A = \{\zeta > a\}$.

3.45. Let A and B be events for a random experiment with sample space S. Show that the Bernoulli random variable satisfies the following properties:

  (a) $I_S = 1$ and $I_\varnothing = 0$.

  (b) $I_{A \cap B} = I_A I_B$ and $I_{A \cup B} = I_A + I_B - I_A I_B$.

  (c) Find the expected value of the indicator functions in parts a and b.

3.46. Heat must be removed from a system according to how fast it is generated. Suppose the system has eight components each of which is active with probability 0.25, independently of the others. The design of the heat removal system requires finding the probabilities of the following events:

  (a) None of the systems is active.

  (b) Exactly one is active.

  (c) More than four are active.

  (d) More than two and fewer than six are active.

3.47. Eight numbers are selected at random from the unit interval.

  (a) Find the probability that the first four numbers are less than 0.25 and the last four are greater than 0.25.

(b) Find the probability that four numbers are less than 0.25 and four are greater than 0.25.

(c) Find the probability that the first three numbers are less than 0.25, the next two are between 0.25 and 0.75, and the last three are greater than 0.75.

(d) Find the probability that three numbers are less than 0.25, two are between 0.25 and 0.75, and three are greater than 0.75.

(e) Find the probability that the first four numbers are less than 0.25 and the last four are greater than 0.75.

(f) Find the probability that four numbers are less than 0.25 and four are greater than 0.75.

3.48. **(a)** Plot the pmf of the binomial random variable with $n = 4$ and $n = 5$, and $p = 0.10$, $p = 0.5$, and $p = 0.90$.

(b) Use Octave to plot the pmf of the binomial random variable with $n = 100$ and $p = 0.10$, $p = 0.5$, and $p = 0.90$.

3.49. Let X be a binomial random variable that results from the performance of n Bernoulli trials with probability of success p.

(a) Suppose that $X = 1$. Find the probability that the single event occurred in the kth Bernoulli trial.

(b) Suppose that $X = 2$. Find the probability that the two events occurred in the jth and kth Bernoulli trials where $j < k$.

(c) In light of your answers to parts a and b in what sense are the successes distributed "completely at random" over the n Bernoulli trials?

3.50. Let X be the binomial random variable.

(a) Show that

$$\frac{p_X(k + 1)}{p_X(k)} = \frac{n - k}{k + 1}\frac{p}{1 - p} \qquad \text{where} \qquad p_X(0) = (1 - p)^n.$$

(b) Show that part a implies that: (1) $P[X = k]$ is maximum at $k_{\max} = [(n + 1)p]$, where $[x]$ denotes the largest integer that is smaller than or equal to x; and (2) when $(n + 1)p$ is an integer, then the maximum is achieved at $k_{\max}$ and $k_{\max} - 1$.

3.51. Consider the expression $(a + b + c)^n$.

(a) Use the binomial expansion for $(a + b)$ and c to obtain an expression for $(a + b + c)^n$.

(b) Now expand all terms of the form $(a + b)^k$ and obtain an expression that involves the multinomial coefficient for $M = 3$ mutually exclusive events, A_1, A_2, A_3.

(c) Let $p_1 = P[A_1]$, $p_2 = P[A_2]$, $p_3 = P[A_3]$. Use the result from part b to show that the multinomial probabilities add to one.

3.52. A sequence of characters is transmitted over a channel that introduces errors with probability $p = 0.01$.

(a) What is the pmf of N, the number of error-free characters between erroneous characters?

(b) What is $E[N]$?

(c) Suppose we want to be 99% sure that at least 1000 characters are received correctly before a bad one occurs. What is the appropriate value of p?

3.53. Let N be a geometric random variable with $S_N = \{1, 2, \dots\}$.

(a) Find $P[N = k | N \leq m]$.

(b) Find the probability that N is odd.

3.54. Let M be a geometric random variable. Show that M satisfies the memoryless property: $P[M \geq k + j | M \geq j + 1] = P[M \geq k]$ for all $j, k > 1$.

3.55. Let X be a discrete random variable that assumes only nonnegative integer values and that satisfies the memoryless property. Show that X must be a geometric random variable. *Hint:* Find an equation that must be satisfied by $g(m) = P[M \geq m]$.

3.56. An audio player uses a low-quality hard drive. The initial cost of building the player is $50. The hard drive fails after each month of use with probability 1/12. The cost to repair the hard drive is $20. If a 1-year warranty is offered, how much should the manufacturer charge so that the probability of losing money on a player is 1% or less? What is the average cost per player?

3.57. A Christmas fruitcake has Poisson-distributed independent numbers of sultana raisins, iridescent red cherry bits, and radioactive green cherry bits with respective averages 48, 24, and 12 bits per cake. Suppose you politely accept 1/12 of a slice of the cake.

   (a) What is the probability that you get lucky and get no green bits in your slice?

   (b) What is the probability that you get really lucky and get no green bits and two or fewer red bits in your slice?

   (c) What is the probability that you get extremely lucky and get no green or red bits and more than five raisins in your slice?

3.58. The number of orders waiting to be processed is given by a Poisson random variable with parameter $\alpha = \lambda/n\mu$, where λ is the average number of orders that arrive in a day, μ is the number of orders that can be processed by an employee per day, and n is the number of employees. Let $\lambda = 5$ and $\mu = 1$. Find the number of employees required so the probability that more than four orders are waiting is less than 10%. What is the probability that there are no orders waiting?

3.59. The number of page requests that arrive at a Web server is a Poisson random variable with an average of 6000 requests per minute.

   (a) Find the probability that there are no requests in a 100-ms period.

   (b) Find the probability that there are between 5 and 10 requests in a 100-ms period.

3.60. Use Octave to plot the pmf of the Poisson random variable with $\alpha = 0.1, 0.75, 2, 20$.

3.61. Find the mean and variance of a Poisson random variable.

3.62. For the Poisson random variable, show that for $\alpha < 1$, $P[N = k]$ is maximum at $k = 0$; for $\alpha > 1$, $P[N = k]$ is maximum at $[\alpha]$; and if α is a positive integer, then $P[N = k]$ is maximum at $k = \alpha$, and at $k = \alpha - 1$. *Hint:* Use the approach of Problem 3.50.

3.63. Compare the Poisson approximation and the binomial probabilities for $k = 0, 1, 2, 3$ and $n = 10$, $p = 0.1$; $n = 20$ and $p = 0.05$; and $n = 100$ and $p = 0.01$.

3.64. At a given time, the number of households connected to the Internet is a Poisson random variable with mean 50. Suppose that the transmission bit rate available for the household is 20 Megabits per second.

   (a) Find the probability of the distribution of the transmission bit rate per user.

   (b) Find the transmission bit rate that is available to a user with probability 90% or higher.

   (c) What is the probability that a user has a share of 1 Megabit per second or higher?

3.65. An LCD display has 1000×750 pixels. A display is accepted if it has 15 or fewer faulty pixels. The probability that a pixel is faulty coming out of the production line is 10^{-5}. Find the proportion of displays that are accepted.

3.66. A data center has 10,000 disk drives. Suppose that a disk drive fails in a given day with probability 10^{-3}.

  (a) Find the probability that there are no failures in a given day.

  (b) Find the probability that there are fewer than 10 failures in two days.

  (c) Find the number of spare disk drives that should be available so that all failures in a day can be replaced with probability 99%.

3.67. A binary communication channel has a probability of bit error of 10^{-6}. Suppose that transmissions occur in blocks of 10,000 bits. Let N be the number of errors introduced by the channel in a transmission block.

  (a) Find $P[N = 0]$, $P[N \leq 3]$.

  (b) For what value of p will the probability of 1 or more errors in a block be 99%?

3.68. Find the mean and variance of the uniform discrete random variable that takes on values in the set $\{1, 2, \ldots, L\}$ with equal probability. You will need the following formulas:

$$\sum_{i=1}^{n} i = \frac{n(n + 1)}{2} \qquad \sum_{i=1}^{n} i^2 = \frac{n(n + 1)(2n + 1)}{6}.$$

3.69. A voltage X is uniformly distributed in the set $\{-3, \ldots, 3, 4\}$.

  (a) Find the mean and variance of X.

  (b) Find the mean and variance of $Y = -2X^2 + 3$.

  (c) Find the mean and variance of $W = \cos(\pi X/8)$.

  (d) Find the mean and variance of $Z = \cos^2(\pi X/8)$.

3.70. Ten news Web sites are ranked in terms of popularity, and the frequency of requests to these sites are known to follow a Zipf distribution.

  (a) What is the probability that a request is for the top-ranked site?

  (b) What is the probability that a request is for one of the bottom five sites?

3.71. A collection of 1000 words is known to have a Zipf distribution.

  (a) What is the probability of the 10 top-ranked words?

  (b) What is the probability of the 10 lowest-ranked words?

3.72. What is the shape of the log of the Zipf probability vs. the log of the rank?

3.73. Plot the mean and variance of the Zipf random variable for $L = 1$ to $L = 100$.

3.74. An online video store has 10,000 titles. In order to provide fast response, the store caches the most popular titles. How many titles should be in the cache so that with probability 99% an arriving video request will be in the cache?

3.75. (a) Income distribution is perfectly equal if every individual has the same income. What is the Lorenz curve in this case?

  (b) In a perfectly unequal income distribution, one individual has all the income and all others have none. What is the Lorenz curve in this case?

3.76. Let X be a geometric random variable in the set $\{1, 2, \ldots\}$.

  (a) Find the pmf of X.

  (b) Find the Lorenz curve of X. Assume L is infinite.

  (c) Plot the curve for $p = 0.1, 0.5, 0.9$.

3.77. Let X be a zeta random variable with parameter α.

  (a) Find an expression for $P[X \leq k]$.

    (b) Plot the pmf of X for $\alpha = 1.5, 2,$ and 3.

    (c) Plot $P[X \le k]$ for $\alpha = 1.5, 2,$ and 3.

Section 3.6: Generation of Discrete Random Variables

3.78. Octave provides function calls to evaluate the pmf of important discrete random variables. For example, the function `Poisson_pdf`(x, *lambda*) computes the pmf at x for the Poisson random variable.

    (a) Plot the Poisson pmf for $\lambda = 0.5, 5, 50$, as well as $P[X \le k]$ and $P[X > k]$.

    (b) Plot the binomial pmf for $n = 48$ and $p = 0.10, 0.30, 0.50, 0.75$, as well as $P[X \le k]$ and $P[X > k]$.

    (c) Compare the binomial probabilities with the Poisson approximation for $n = 100$, $p = 0.01$.

3.79. The `discrete_pdf` function in Octave makes it possible to specify an arbitrary pmf for a specified S_X.

    (a) Plot the pmf for Zipf random variables with $L = 10, 100, 1000$, as well as $P[X \le k]$ and $P[X > k]$.

    (b) Plot the pmf for the reward in the St. Petersburg Paradox for $m = 20$ in Problem 3.34, as well as $P[X \le k]$ and $P[X > k]$. (You will need to use a log scale for the values of k.)

3.80. Use Octave to plot the Lorenz curve for the Zipf random variables in Problem 3.79a.

3.81. Repeat Problem 3.80 for the binomial random variable with $n = 100$ and $p = 0.1, 0.5$, and 0.9.

3.82. **(a)** Use the `discrete_rnd` function in Octave to simulate the urn experiment discussed in Section 1.3. Compute the relative frequencies of the outcomes in 1000 draws from the urn.

    (b) Use the `discrete_pdf` function in Octave to specify a pmf for a binomial random variable with $n = 5$ and $p = 0.2$. Use `discrete_rnd` to generate 100 samples and plot the relative frequencies.

    (c) Use `binomial_rnd` to generate the 100 samples in part b.

3.83. Use the `discrete_rnd` function to generate 200 samples of the Zipf random variable in Problem 3.79a. Plot the sequence of outcomes as well as the overall relative frequencies.

3.84. Use the `discrete_rnd` function to generate 200 samples of the St. Petersburg Paradox random variable in Problem 3.79b. Plot the sequence of outcomes as well as the overall relative frequencies.

3.85. Use Octave to generate 200 pairs of numbers, (X_i, Y_i), in which the components are independent, and each component is uniform in the set $\{1, 2, \ldots, 9, 10\}$.

    (a) Plot the relative frequencies of the X and Y outcomes.

    (b) Plot the relative frequencies of the random variable $Z = X + Y$. Can you discern the pmf of Z?

    (c) Plot the relative frequencies of $W = XY$. Can you discern the pmf of Z?

    (d) Plot the relative frequencies of $V = X/Y$. Is the pmf discernable?

3.86. Use Octave function `binomial_rnd` to generate 200 pairs of numbers, (X_i, Y_i), in which the components are independent, and where X_i are binomial with parameter $n = 8$, $p = 0.5$ and Y_i are binomial with parameter $n = 4$, $p = 0.5$.

(a) Plot the relative frequencies of the X and Y outcomes.

(b) Plot the relative frequencies of the random variable $Z = X + Y$. Does this correspond to the pmf you would expect? Explain.

3.87. Use Octave function `Poisson_rnd` to generate 200 pairs of numbers, (X_i, Y_i), in which the components are independent, and where X_i are the number of arrivals to a system in one second and Y_i are the number of arrivals to the system in the next two seconds. Assume that the arrival rate is five customers per second.

(a) Plot the relative frequencies of the X and Y outcomes.

(b) Plot the relative frequencies of the random variable $Z = X + Y$. Does this correspond to the pmf you would expect? Explain.

Problems Requiring Cumulative Knowledge

3.88. The fraction of defective items in a production line is p. Each item is tested and defective items are identified correctly with probability a.

(a) Assume nondefective items always pass the test. What is the probability that k items are tested until a defective item is identified?

(b) Suppose that the identified defective items are removed. What proportion of the remaining items is defective?

(c) Now suppose that nondefective items are identified as defective with probability b. Repeat part b.

3.89. A data transmission system uses messages of duration T seconds. After each message transmission, the transmitter stops and waits T seconds for a reply from the receiver. The receiver immediately replies with a message indicating that a message was received correctly. The transmitter proceeds to send a new message if it receives a reply within T seconds; otherwise, it retransmits the previous message. Suppose that messages can be completely garbled while in transit and that this occurs with probability p. Find the maximum possible rate at which messages can be successfully transmitted from the transmitter to the receiver.

3.90. An inspector selects every nth item in a production line for a detailed inspection. Suppose that the time between item arrivals is an exponential random variable with mean 1 minute, and suppose that it takes 2 minutes to inspect an item. Find the smallest value of n such that with a probability of 90% or more, the inspection is completed before the arrival of the next item that requires inspection.

3.91. The number X of photons counted by a receiver in an optical communication system is a Poisson random variable with rate λ_1 when a signal is present and a Poisson random variable with rate $\lambda_0 < \lambda_1$ when a signal is absent. Suppose that a signal is present with probability p.

(a) Find $P[\text{signal present} \mid X = k]$ and $P[\text{signal absent} \mid X = k]$.

(b) The receiver uses the following decision rule:

If $P[\text{signal present} \mid X = k] > P[\text{signal absent} \mid X = k]$, decide signal present; otherwise, decide signal absent.

Show that this decision rule leads to the following threshold rule:

If $X > T$, decide signal present; otherwise, decide signal absent.

(c) What is the probability of error for the above decision rule?

3.92. A binary information source (e.g., a document scanner) generates very long strings of 0's followed by occasional 1's. Suppose that symbols are independent and that $p = P[\text{symbol} = 0]$ is very close to one. Consider the following scheme for encoding the run X of 0's between consecutive 1's:

1. If $X = n$, express n as a multiple of an integer $M = 2^m$ and a remainder r, that is, find k and r such that $n = kM + r$, where $0 \le r < M - 1$;

2. The binary codeword for n then consists of a prefix consisting of k 0's followed by a 1, and a suffix consisting of the m-bit representation of the remainder r. The decoder can deduce the value of n from this binary string.

(a) Find the probability that the prefix has k zeros, assuming that $p^M = 1/2$.

(b) Find the average codeword length when $p^M = 1/2$.

(c) Find the compression ratio, which is defined as the ratio of the average run length to the average codeword length when $p^M = 1/2$.

One Random Variable

In Chapter 3 we introduced the notion of a random variable and we developed methods for calculating probabilities and averages for the case where the random variable is discrete. In this chapter we consider the general case where the random variable may be discrete, continuous, or of mixed type. We introduce the cumulative distribution function which is used in the formal definition of a random variable, and which can handle all three types of random variables. We also introduce the probability density function for continuous random variables. The probabilities of events involving a random variable can be expressed as integrals of its probability density function. The expected value of continuous random variables is also introduced and related to our intuitive notion of average. We develop a number of methods for calculating probabilities and averages that are the basic tools in the analysis and design of systems that involve randomness.

4.1 THE CUMULATIVE DISTRIBUTION FUNCTION

The probability mass function of a discrete random variable was defined in terms of events of the form $\{X = b\}$. The cumulative distribution function is an alternative approach which uses events of the form $\{X \leq b\}$. The cumulative distribution function has the advantage that it is not limited to discrete random variables and applies to all types of random variables. We begin with a formal definition of a random variable.

> **Definition:** Consider a random experiment with sample space S and event class $\mathcal{F}$. A **random variable** X is a function from the sample space S to R with the property that the set $A_b = \{\zeta: X(\zeta) \leq b\}$ is in $\mathcal{F}$ for every b in R.

The definition simply requires that every set A_b have a well defined probability in the underlying random experiment, and this is not a problem in the cases we will consider. Why does the definition use sets of the form $\{\zeta: X(\zeta) \leq b\}$ and not $\{\zeta: X(\zeta) = x_b\}$? We will see that all events of interest in the real line can be expressed in terms of sets of the form $\{\zeta: X(\zeta) \leq b\}$.

The **cumulative distribution function** (cdf) of a random variable X is defined as the probability of the event $\{X \leq x\}$:

$$F_X(x) = P[X \leq x] \qquad \text{for } -\infty < x < +\infty, \tag{4.1}$$

that is, it is the probability that the random variable X takes on a value in the set $(-\infty, x]$. In terms of the underlying sample space, the cdf is the probability of the event $\{\zeta : X(\zeta) \leq x\}$. The event $\{X \leq x\}$ and its probability vary as x is varied; in other words, $F_X(x)$ is a function of the variable x.

The cdf is simply a convenient way of specifying the probability of all semi-infinite intervals of the real line of the form $(-\infty, b]$. The events of interest when dealing with numbers are intervals of the real line, and their complements, unions, and intersections. We show below that the probabilities of all of these events can be expressed in terms of the cdf.

The cdf has the following interpretation in terms of relative frequency. Suppose that the experiment that yields the outcome ζ, and hence $X(\zeta)$, is performed a large number of times. $F_X(b)$ is then the long-term proportion of times in which $X(\zeta) \leq b$.

Before developing the general properties of the cdf, we present examples of the cdfs for three basic types of random variables.

Example 4.1   Three Coin Tosses

Figure 4.1(a) shows the cdf X, the number of heads in three tosses of a fair coin. From Example 3.1 we know that X takes on only the values 0, 1, 2, and 3 with probabilities 1/8, 3/8, 3/8, and 1/8, respectively, so $F_X(x)$ is simply the sum of the probabilities of the outcomes from $\{0, 1, 2, 3\}$ that are less than or equal to x. The resulting cdf is seen to be a nondecreasing staircase function that grows from 0 to 1. The cdf has jumps at the points 0, 1, 2, 3 of magnitudes 1/8, 3/8, 3/8, and 1/8, respectively.

Let us take a closer look at one of these discontinuities, say, in the vicinity of $x = 1$. For δ a small positive number, we have

$$F_X(1 - \delta) = P[X \leq 1 - \delta] = P\{0 \text{ heads}\} = \frac{1}{8}$$

so the limit of the cdf as x approaches 1 from the left is 1/8. However,

$$F_X(1) = P[X \leq 1] = P[0 \text{ or } 1 \text{ heads}] = \frac{1}{8} + \frac{3}{8} = \frac{1}{2},$$

and furthermore the limit from the right is

$$F_X(1 + \delta) = P[X \leq 1 + \delta] = P[0 \text{ or } 1 \text{ heads}] = \frac{1}{2}.$$
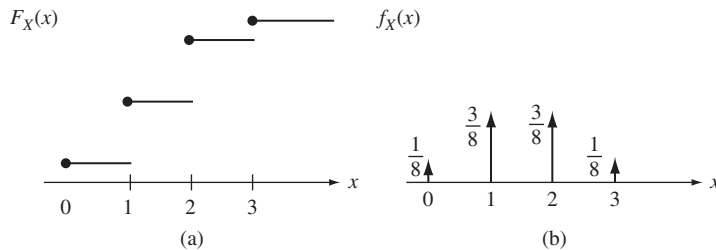
FIGURE 4.1
cdf (a) and pdf (b) of a discrete random variable.

Thus the cdf is continuous from the right and equal to 1/2 at the point $x = 1$. Indeed, we note the magnitude of the jump at the point $x = 1$ is equal to $P[X = 1] = 1/2 - 1/8 = 3/8$. Henceforth we will use dots in the graph to indicate the value of the cdf at the points of discontinuity.

The cdf can be written compactly in terms of the unit step function:

$$u(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0, \end{cases} \tag{4.2}$$

then

$$F_X(x) = \frac{1}{8}u(x) + \frac{3}{8}u(x - 1) + \frac{3}{8}u(x - 2) + \frac{1}{8}u(x - 3).$$

Example 4.2   Uniform Random Variable in the Unit Interval

Spin an arrow attached to the center of a circular board. Let θ be the final angle of the arrow, where $0 < \theta \leq 2\pi$. The probability that θ falls in a subinterval of $(0, 2\pi]$ is proportional to the length of the subinterval. The random variable X is defined by $X(\theta) = \theta/2\pi$. Find the cdf of X:

As θ increases from 0 to 2π, X increases from 0 to 1. No outcomes θ lead to values $x \leq 0$, so

$$F_X(x) = P[X \leq x] = P[\varnothing] = 0 \qquad \text{for } x < 0.$$

For $0 < x \leq 1$, $\{X \leq x\}$ occurs when $\{\theta \leq 2\pi x\}$ so

$$F_X(x) = P[X \leq x] = P[\{\theta \leq 2\pi x\}] = 2\pi x/2\pi = x \qquad 0 < x \leq 1. \tag{4.3}$$

Finally, for $x > 1$, all outcomes θ lead to $\{X(\theta) \leq 1 < x\}$, therefore:

$$F_X(x) = P[X \leq x] = P[0 < \theta \leq 2\pi] = 1 \qquad \text{for } x > 1.$$

We say that X is a **uniform random variable** in the unit interval. Figure 4.2(a) shows the cdf of the general uniform random variable X. We see that $F_X(x)$ is a nondecreasing continuous function that grows from 0 to 1 as x ranges from its minimum values to its maximum values.
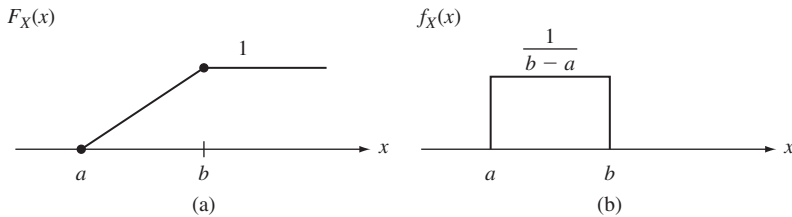
FIGURE 4.2
cdf (a) and pdf (b) of a continuous random variable.

Example 4.3

The waiting time X of a customer at a taxi stand is zero if the customer finds a taxi parked at the stand, and a uniformly distributed random length of time in the interval $[0, 1]$ (in hours) if no taxi is found upon arrival. The probability that a taxi is at the stand when the customer arrives is p. Find the cdf of X.

The cdf is found by applying the theorem on total probability:

$$F_X(x) = P[X \leq x] = P[X \leq x \mid \text{find taxi}]p + P[X \leq x \mid \text{no taxi}](1 - p).$$

Note that $P[X \leq x \mid \text{find taxi}] = 1$ when $x \geq 0$ and 0 otherwise. Furthermore $P[X \leq x \mid \text{no taxi}]$ is given by Eq. (4.3), therefore

$$F_X(x) = \begin{cases} 0 & x < 0 \\ p + (1 - p)x & 0 \leq x \leq 1 \\ 1 & x > 1. \end{cases}$$

The cdf, shown in Fig. 4.3(a), combines some of the properties of the cdf in Example 4.1 (discontinuity at 0) and the cdf in Example 4.2 (continuity over intervals). Note that $F_X(x)$ can be expressed as the sum of a step function with amplitude p and a continuous function of x.

We are now ready to state the basic properties of the cdf. The axioms of probability and their corollaries imply that the cdf has the following properties:

(i) $0 \leq F_X(x) \leq 1$.

(ii) $\lim_{x \to \infty} F_X(x) = 1$.

(iii) $\lim_{x \to -\infty} F_X(x) = 0$.

(iv) $F_X(x)$ is a nondecreasing function of x, that is, if $a < b$, then $F_X(a) \leq F_X(b)$.

(v) $F_X(x)$ is continuous from the right, that is, for $h > 0$, $F_X(b) = \lim_{h \to 0} F_X(b + h) = F_X(b^+)$.

These five properties confirm that, in general, the cdf is a nondecreasing function that grows from 0 to 1 as x increases from $-\infty$ to ∞. We already observed these properties in Examples 4.1, 4.2, and 4.3. Property (v) implies that at points of discontinuity, the cdf

(a)   (b)
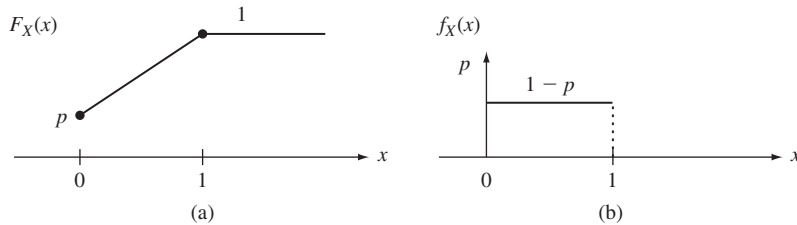
FIGURE 4.3
cdf (a) and pdf (b) of a random variable of mixed type.

is equal to the limit from the right. We observed this property in Examples 4.1 and 4.3. In Example 4.2 the cdf is continuous for all values of x, that is, the cdf is continuous both from the right and from the left for all x.

The cdf has the following properties which allow us to calculate the probability of events involving intervals and single values of X:

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

(vii) $P[X = b] = F_X(b) - F_X(b^-)$.

(viii) $P[X > x] = 1 - F_X(x)$.

Property (vii) states that the probability that $X = b$ is given by the magnitude of the jump of the cdf at the point b. This implies that *if the cdf is continuous at a point b, then* $P[X = b] = 0$. Properties (vi) and (vii) can be combined to compute the probabilities of other types of intervals. For example, since $\{a \leq X \leq b\} = \{X = a\} \cup \{a < X \leq b\}$, then

$$P[a \leq X \leq b] = P[X = a] + P[a < X \leq b]$$
$$= F_X(a) - F_X(a^-) + F_X(b) - F_X(a) = F_X(b) - F_X(a^-). \quad (4.4)$$

If the cdf is continuous at the endpoints of an interval, then the endpoints have zero probability, and therefore they can be included in, or excluded from, the interval without affecting the probability.

Example 4.4

Let X be the number of heads in three tosses of a fair coin. Use the cdf to find the probability of the events $A = \{1 < X \leq 2\}$, $B = \{0.5 \leq X < 2.5\}$, and $C = \{1 \leq X < 2\}$.

From property (vi) and Fig. 4.1 we have

$$P[1 < X \leq 2] = F_X(2) - F_X(1) = 7/8 - 1/2 = 3/8.$$

The cdf is continuous at $x = 0.5$ and $x = 2.5$, so

$$P[0.5 \leq X < 2.5] = F_X(2.5) - F_X(0.5) = 7/8 - 1/8 = 6/8.$$

Since $\{1 \leq X < 2\} \cup \{X = 2\} = \{1 \leq X \leq 2\}$, from Eq. (4.4) we have

$$P\{1 \leq X < 2\] + P[X = 2] = F_X(2) - F_X(1^-),$$

and using property (vii) for $P[X = 2]$:

$$P\{1 \leq X < 2\] = F_X(2) - F_X(1^-) - P[X = 2] = F_X(2) - F_X(1^-) - (F_X(2) - F_X(2^-))$$
$$= F_X(2^-) - F_X(1^-) = 4/8 - 1/8 = 3/8.$$

Example 4.5

Let X be the uniform random variable from Example 4.2. Use the cdf to find the probability of the events $\{-0.5 < X < 0.25\}$, $\{0.3 < X < 0.65\}$, and $\{|X - 0.4| > 0.2\}$.

The cdf of X is continuous at every point so we have:

$P[-0.5 < X \leq 0.25] = F_X(0.25) - F_X(-0.5) = 0.25 - 0 = 0.25,$

$P[0.3 < X < 0.65] = F_X(0.65) - F_X(0.3) = 0.65 - 0.3 = 0.35,$

$P[|X - 0.4| > 0.2] = P[\{X < 0.2\} \cup \{X > 0.6\}] = P[X < 0.2] + P[X > 0.6]$

$$= F_X(0.2) + (1 - F_X(0.6)) = 0.2 + 0.4 = 0.6.$$

We now consider the proof of the properties of the cdf.

- Property (i) follows from the fact that the cdf is a probability and hence must satisfy Axiom I and Corollary 2.
- To obtain property (iv), we note that the event $\{X \leq a\}$ is a subset of $\{X \leq b\}$, and so it must have smaller or equal probability (Corollary 7).
- To show property (vi), we note that $\{X \leq b\}$ can be expressed as the union of mutually exclusive events: $\{X \leq a\} \cup \{a < X \leq b\} = \{X \leq b\}$, and so by Axiom III, $F_X(a) + P[a < X \leq b] = F_X(b)$.
- Property (viii) follows from $\{X > x\} = \{X \leq x\}^c$ and Corollary 1.

While intuitively clear, properties (ii), (iii), (v), and (vii) require more advanced limiting arguments that are discussed at the end of this section.

4.1.1 The Three Types of Random Variables

The random variables in Examples 4.1, 4.2, and 4.3 are typical of the three most basic types of random variable that we are interested in.

Discrete random variables have a cdf that is a right-continuous, staircase function of x, with jumps at a countable set of points $x_0, x_1, x_2, \ldots$. The random variable in Example 4.1 is a typical example of a discrete random variable. The cdf $F_X(x)$ of a discrete random variable is the sum of the probabilities of the outcomes less than x and can be written as the weighted sum of unit step functions as in Example 4.1:

$$F_X(x) = \sum_{x_k \leq x} p_X(x_k) = \sum_k p_X(x_k) u(x - x_k), \tag{4.5}$$

where the pmf $p_X(x_k) = P[X = x_k]$ gives the magnitude of the jumps in the cdf. We see that the pmf can be obtained from the cdf and vice versa.

A **continuous random variable** is defined as a random variable whose cdf $F_X(x)$ is continuous everywhere, and which, in addition, is sufficiently smooth that it can be written as an integral of some nonnegative function $f(x)$:

$$F_X(x) = \int_{-\infty}^{x} f(t) \, dt. \tag{4.6}$$

The random variable discussed in Example 4.2 can be written as an integral of the function shown in Fig. 4.2(b). The continuity of the cdf and property (vii) implies that continuous

random variables have $P[X = x] = 0$ for all x. Every possible outcome has probability zero! An immediate consequence is that the pmf cannot be used to characterize the probabilities of X. A comparison of Eqs. (4.5) and (4.6) suggests how we can proceed to characterize continuous random variables. For discrete random variables, (Eq. 4.5), we calculate probabilities as summations of probability masses at discrete points. For continuous random variables, (Eq. 4.6), we calculate probabilities as integrals of "probability densities" over intervals of the real line.

A **random variable of mixed type** is a random variable with a cdf that has jumps on a countable set of points $x_0, x_1, x_2, \ldots$, but that also increases continuously over at least one interval of values of x. The cdf for these random variables has the form

$$F_X(x) = pF_1(x) + (1 - p)F_2(x),$$

where $0 < p < 1$, and $F_1(x)$ is the cdf of a discrete random variable and $F_2(x)$ is the cdf of a continuous random variable. The random variable in Example 4.3 is of mixed type.

Random variables of mixed type can be viewed as being produced by a two-step process: A coin is tossed; if the outcome of the toss is heads, a discrete random variable is generated according to $F_1(x)$; otherwise, a continuous random variable is generated according to $F_2(x)$.

*4.1.2   Fine Point: Limiting properties of cdf

Properties (ii), (iii), (v), and (vii) require the continuity property of the probability function discussed in Section 2.9. For example, for property (ii), we consider the sequence of events $\{X \leq n\}$ which increases to include all of the sample space S as n approaches ∞, that is, all outcomes lead to a value of X less than infinity. The continuity property of the probability function (Corollary 8) implies that:

$$\lim_{n\to\infty} F_X(n) = \lim_{n\to\infty} P[X \leq n] = P[\lim_{n\to\infty} \{X \leq n\}] = P[S] = 1.$$

For property (iii), we take the sequence $\{X \leq -n\}$ which decreases to the empty set $\varnothing$, that is, no outcome leads to a value of X less than $-\infty$:

$$\lim_{n\to\infty} F_X(-n) = \lim_{n\to\infty} P[X \leq -n] = P[\lim_{n\to\infty} \{X \leq -n\}] = P[\varnothing] = 0.$$

For property (v), we take the sequence of events $\{X \leq x + 1/n\}$ which decreases to $\{X \leq x\}$ from the right:

$$\lim_{n\to\infty} F_X(x + 1/n) = \lim_{n\to\infty} P[X \leq x + 1/n]$$

$$= P[\lim_{n\to\infty} \{X \leq x + 1/n\}] = P[\{X \leq x\}] = F_X(x).$$

Finally, for property (vii), we take the sequence of events, $\{b - 1/n < X \leq b\}$ which decreases to $\{b\}$ from the left:

$$\lim_{n\to\infty} (F_X(b) - F_X(b - 1/n)) = \lim_{n\to\infty} P[b - 1/n < X \leq b]$$

$$= P[\lim_{n\to\infty} \{b - 1/n < X \leq b\}] = P[X = b].$$

4.2   THE PROBABILITY DENSITY FUNCTION

The **probability density function of X** (pdf), if it exists, is defined as the derivative of $F_X(x)$:

$$f_X(x) = \frac{dF_X(x)}{dx}. \tag{4.7}$$

In this section we show that the pdf is an alternative, and more useful, way of specifying the information contained in the cumulative distribution function.

   The pdf represents the "density" of probability at the point x in the following sense: The probability that X is in a small interval in the vicinity of x—that is, $\{x < X \le x + h\}$—is

$$P[x < X \le x + h] = F_X(x + h) - F_X(x)$$

$$= \frac{F_X(x + h) - F_X(x)}{h} h. \tag{4.8}$$

If the cdf has a derivative at x, then as h becomes very small,

$$P[x < X \le x + h] \simeq f_X(x)h. \tag{4.9}$$

Thus $f_X(x)$ represents the "density" of probability at the point x in the sense that the probability that X is in a small interval in the vicinity of x is approximately $f_X(x)h$. The derivative of the cdf, when it exists, is positive since the cdf is a nondecreasing function of x, thus

   (i) $f_X(x) \ge 0.$ $\hspace{2cm}$ (4.10)

   Equations (4.9) and (4.10) provide us with an alternative approach to specifying the probabilities involving the random variable X. We can begin by stating a nonnegative function $f_X(x)$, called the probability density function, which specifies the probabilities of events of the form "X falls in a small interval of width dx about the point x," as shown in Fig. 4.4(a). The probabilities of events involving X are then expressed in terms of the pdf by adding the probabilities of intervals of width dx. As the widths of the intervals approach zero, we obtain an integral in terms of the pdf. For example, the probability of an interval $[a, b]$ is

   (ii) $P[a \le X \le b] = \displaystyle\int_a^b f_X(x)\, dx.$ $\hspace{2cm}$ (4.11)

The probability of an interval is therefore the area under $f_X(x)$ in that interval, as shown in Fig. 4.4(b). The probability of any event that consists of the union of disjoint intervals can thus be found by adding the integrals of the pdf over each of the intervals.

   The cdf of X can be obtained by integrating the pdf:

   (iii) $F_X(x) = \displaystyle\int_{-\infty}^x f_X(t)\, dt.$ $\hspace{2cm}$ (4.12)

In Section 4.1, we defined a *continuous random variable* as a random variable X whose cdf was given by Eq. (4.12). Since the probabilities of all events involving X can be written in terms of the cdf, it then follows that these probabilities can be written in
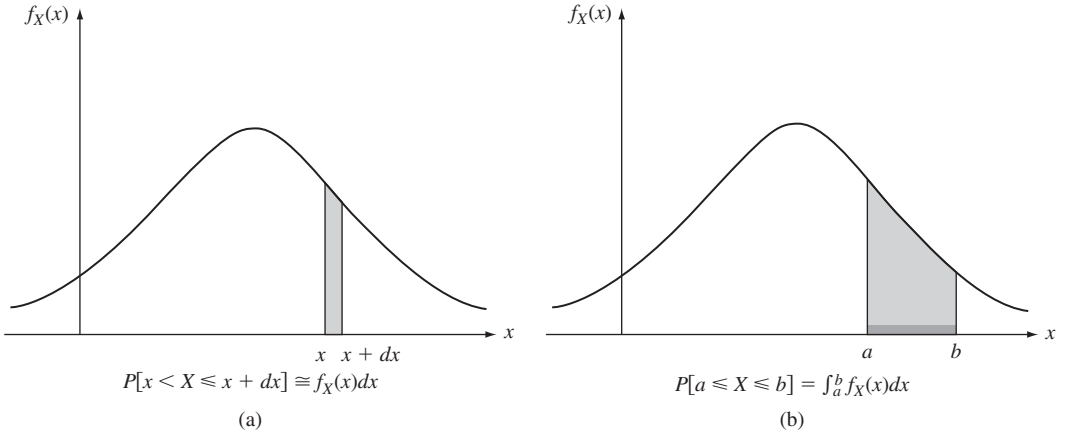
FIGURE 4.4
(a) The probability density function specifies the probability of intervals of infinitesimal width. (b) The probability of an interval [a, b] is the area under the pdf in that interval.

terms of the pdf. Thus *the pdf completely specifies the behavior of continuous random variables.*

By letting x tend to infinity in Eq. (4.12), we obtain a *normalization* condition for pdf's:

(iv) $$1 = \int_{-\infty}^{+\infty} f_X(t)\, dt. \tag{4.13}$$

The pdf reinforces the intuitive notion of probability as having attributes similar to "physical mass." Thus Eq. (4.11) states that the probability "mass" in an interval is the integral of the "density of probability mass" over the interval. Equation (4.13) states that the total mass available is one unit.

A valid pdf can be formed from any nonnegative, piecewise continuous function g(x) that has a finite integral:

$$\int_{-\infty}^{\infty} g(x)\, dx = c < \infty. \tag{4.14}$$

By letting $f_X(x) = g(x)/c$, we obtain a function that satisfies the normalization condition. Note that the pdf must be defined for all real values of x; if X does not take on values from some region of the real line, we simply set $f_X(x) = 0$ in the region.

Example 4.6   Uniform Random Variable

The pdf of the uniform random variable is given by:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & x < a \quad \text{and} \quad x > b \end{cases} \tag{4.15a}$$

and is shown in Fig. 4.2(b). The cdf is found from Eq. (4.12):

$$F_X(x) = \begin{cases} 0 & x < a \\ \dfrac{x - a}{b - a} & a \le x \le b \\ 1 & x > b. \end{cases} \tag{4.15b}$$

The cdf is shown in Fig. 4.2(a).

Example 4.7   Exponential Random Variable

The transmission time X of messages in a communication system has an exponential distribution:

$$P[X > x] = e^{-\lambda x} \qquad x > 0.$$

Find the cdf and pdf of X.

The cdf is given by $F_X(x) = 1 - P[X > x]$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0. \end{cases} \tag{4.16a}$$

The pdf is obtained by applying Eq. (4.7):

$$f_X(x) = F'_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \ge 0. \end{cases} \tag{4.16b}$$

Example 4.8   Laplacian Random Variable

The pdf of the samples of the amplitude of speech waveforms is found to decay exponentially at a rate α, so the following pdf is proposed:

$$f_X(x) = c e^{-\alpha|x|} \qquad -\infty < x < \infty. \tag{4.17}$$

Find the constant c, and then find the probability $P[|X| < v]$.

We use the normalization condition in (iv) to find c:

$$1 = \int_{-\infty}^{\infty} c e^{-\alpha|x|}\, dx = 2 \int_{0}^{\infty} c e^{-\alpha x}\, dx = \frac{2c}{\alpha}.$$

Therefore $c = \alpha/2$. The probability $P[|X| < v]$ is found by integrating the pdf:

$$P[|X| < v] = \frac{\alpha}{2} \int_{-v}^{v} e^{-\alpha|x|}\, dx = 2\left(\frac{\alpha}{2}\right) \int_{0}^{v} e^{-\alpha x}\, dx = 1 - e^{-\alpha v}.$$

4.2.1   pdf of Discrete Random Variables

The derivative of the cdf does not exist at points where the cdf is not continuous. Thus the notion of pdf as defined by Eq. (4.7) does not apply to discrete random variables at the points where the cdf is discontinuous. We can generalize the definition of the

probability density function by noting the relation between the unit step function and the delta function. The **unit step function** is defined as

$$u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0. \end{cases} \tag{4.18a}$$

The **delta function** $\delta(t)$ is related to the unit step function by the following equation:

$$u(x) = \int_{-\infty}^{x} \delta(t)\, dt. \tag{4.18b}$$

A translated unit step function is then:

$$u(x - x_0) = \int_{-\infty}^{x - x_0} \delta(t)\, dt = \int_{-\infty}^{x} \delta(t' - x_0)\, dt'. \tag{4.18c}$$

Substituting Eq. (4.18c) into the cdf of a discrete random variables:

$$F_X(x) = \sum_k p_X(x_k) u(x - x_k) = \sum_k p_X(x_k) \int_{-\infty}^{x} \delta(t - x_k)\, dt$$

$$= \int_{-\infty}^{x} \sum_k p_X(x_k) \delta(t - x_k)\, dt. \tag{4.19}$$

This suggests that we define the **pdf for a discrete random variable** by

$$f_X(x) = \frac{d}{dx} F_X(x) = \sum_k p_X(x_k) \delta(x - x_k). \tag{4.20}$$

Thus the generalized definition of pdf places a delta function of weight $P[X = x_k]$ at the points x_k where the cdf is discontinuous.

To provide some intuition on the delta function, consider a narrow rectangular pulse of unit area and width Δ centered at $t = 0$:

$$\pi_\Delta(t) = \begin{cases} 1/\Delta & -\Delta/2 \leq t \leq \Delta/2 \\ 0 & |t| > \Delta. \end{cases}$$

Consider the integral of $\pi_\Delta(t)$:

$$\int_{-\infty}^{x} \pi_\Delta(t)\, dt = \begin{cases} \int_{-\infty}^{x} \pi_\Delta(t)\, dt = \int_{-\infty}^{x} 0\, dt = 0 & \text{for } x < -\Delta/2 \\ \\ \int_{-\infty}^{x} \pi_\Delta(t)\, dt = \int_{-\Delta/2}^{\Delta/2} 1/\Delta\, dt = 1 & \text{for } x > \Delta/2 \end{cases} \rightarrow u(x). \tag{4.21}$$

As $\Delta \rightarrow 0$, we see that the integral of the narrow pulse approaches the unit step function. For this reason, we visualize the delta function $\delta(t)$ as being zero everywhere

except at $x = 0$ where it is unbounded. The above equation does not apply at the value $x = 0$. To maintain the right continuity in Eq. (4.18a), we use the convention:

$$u(0) = 1 = \int_{-\infty}^{0} \delta(t) \, dt.$$

If we replace $\pi_\Delta(t)$ in the above derivation with $g(t)\pi_\Delta(t)$, we obtain the "sifting" property of the delta function:

$$g(0) = \int_{-\infty}^{\infty} g(t)\delta(t) \, dt \quad \text{and} \qquad g(x_0) = \int_{-\infty}^{\infty} g(t)\delta(t - x_0) \, dt. \qquad (4.22)$$

The delta function is viewed as sifting through x and picking out the value of g at the point where the delta functions is centered, that is, $g(x_0)$ for the expression on the right.

The pdf for the discrete random variable discussed in Example 4.1 is shown in Fig. 4.1(b). The pdf of a random variable of mixed type will also contain delta functions at the points where its cdf is not continuous. The pdf for the random variable discussed in Example 4.3 is shown in Fig. 4.3(b).

Example 4.9

Let X be the number of heads in three coin tosses as in Example 4.1. Find the pdf of X. Find $P[1 < X \le 2]$ and $P[2 \le X < 3]$ by integrating the pdf.

In Example 4.1 we found that the cdf of X is given by

$$F_X(x) = \frac{1}{8}u(x) + \frac{3}{8}u(x - 1) + \frac{3}{8}u(x - 2) + \frac{1}{8}u(x - 3).$$

It then follows from Eqs. (4.18) and (4.19) that

$$f_X(x) = \frac{1}{8}\delta(x) + \frac{3}{8}\delta(x - 1) + \frac{3}{8}\delta(x - 2) + \frac{1}{8}\delta(x - 3).$$

When delta functions appear in the limits of integration, we must indicate whether the delta functions are to be included in the integration. Thus in $P[1 < X \le 2] = P[X \text{ in } (1, 2]]$, the delta function located at 1 is excluded from the integral and the delta function at 2 is included:

$$P[1 < X \le 2] = \int_{1+}^{2+} f_X(x) \, dx = \frac{3}{8}.$$

Similarly, we have that

$$P[2 \le X < 3] = \int_{2-}^{3-} f_X(x) \, dx = \frac{3}{8}.$$

4.2.2 Conditional cdf's and pdf's

Conditional cdf's can be defined in a straightforward manner using the same approach we used for conditional pmf's. Suppose that event C is given and that $P[C] > 0$. The **conditional cdf of X given C** is defined by

$$F_X(x|C) = \frac{P[\{X \le x\} \cap C]}{P[C]} \quad \text{if } P[C] > 0. \qquad (4.23)$$

It is easy to show that $F_X(x|C)$ satisfies all the properties of a cdf. (See Problem 4.29.) The **conditional pdf of X given C** is then defined by

$$f_X(x|C) = \frac{d}{dx} F_X(x|C). \tag{4.24}$$

Example 4.10

The lifetime X of a machine has a continuous cdf $F_X(x)$. Find the conditional cdf and pdf given the event $C = \{X > t\}$ (i.e., "machine is still working at time t").

   The conditional cdf is

$$F_X(x|X > t) = P[X \le x | X > t] = \frac{P[\{X \le x\} \cap \{X > t\}]}{P[X > t]}.$$

The intersection of the two events in the numerator is equal to the empty set when $x < t$ and to $\{t < X \le x\}$ when $x \ge t$. Thus

$$F_X(x|X > t) = \begin{cases} 0 & x \le t \\ \dfrac{F_X(x) - F_X(t)}{1 - F_X(t)} & x > t. \end{cases}$$

The conditional pdf is found by differentiating with respect to x:

$$f_X(x | X > t) = \frac{f_X(x)}{1 - F_X(t)} \quad x \ge t.$$

Now suppose that we have a partition of the sample space S into the union of disjoint events $B_1, B_2, \ldots, B_n$. Let $F_X(x|B_i)$ be the conditional cdf of X given event B_i. The theorem on total probability allows us to find the cdf of X in terms of the conditional cdf's:

$$F_X(x) = P[X \le x] = \sum_{i=1}^{n} P[X \le x | B_i] P[B_i] = \sum_{i=1}^{n} F_X(x | B_i) P[B_i]. \tag{4.25}$$

The pdf is obtained by differentiation:

$$f_X(x) = \frac{d}{dx} F_X(x) = \sum_{i=1}^{n} f_X(x | B_i) P[B_i]. \tag{4.26}$$

Example 4.11

A binary transmission system sends a "0" bit by transmitting a $-v$ voltage signal, and a "1" bit by transmitting a $+v$. The received signal is corrupted by Gaussian noise and given by:

$$Y = X + N$$

where X is the transmitted signal, and N is a noise voltage with pdf $f_N(x)$. Assume that $P["1"] = p = 1 - P["0"]$. Find the pdf of Y.

Let B_0 be the event "0" is transmitted and B_1 be the event "1" is transmitted, then B_0, B_1 form a partition, and

$$F_Y(x) = F_Y(x \mid B_0)[B_0] + F_Y(x \mid B_1)[B_1]$$

$$= P[Y \leq x \mid X = -v](1 - p) + P[Y \leq x \mid X = v]p.$$

Since $Y = X + N$, the event $\{Y < x \mid X = v\}$ is equivalent to $\{v + N < x\}$ and $\{N < x - v\}$, and the event $\{Y < x \mid X = -v\}$ is equivalent to $\{N < x + v\}$. Therefore the conditional cdf's are:

$$F_Y(x \mid B_0) = P[N \leq x + v] = F_N(x + v)$$

and

$$F_Y(x \mid B_1) = P[N \leq x - v] = F_N(x - v).$$

The cdf is:

$$F_Y(x) = F_N(x + v)(1 - p) + F_N(x - v)p.$$

The pdf of N is then:

$$f_Y(x) = \frac{d}{dx} F_Y(x)$$

$$= \frac{d}{dx} F_N(x + v)(1 - p) + \frac{d}{dx} F_N(x - v)p$$

$$= f_N(x + v)(1 - p) + f_N(x - v)p.$$

The Gaussian random variable has pdf:

$$f_N(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2} \qquad -\infty < x < \infty.$$

The conditional pdfs are:

$$f_Y(x \mid B_0) = f_N(x + v) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x+v)^2/2\sigma^2}$$
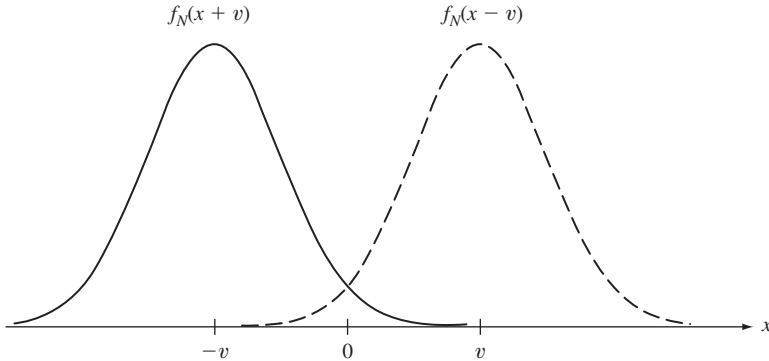
FIGURE 4.5
The conditional pdfs given the input signal

and

$$f_Y(x|B_1) = f_N(x - v) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-v)^2/2\sigma^2}.$$

The pdf of the received signal Y is then:

$$f_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x+v)^2/2\sigma^2}(1 - p) + \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(x-v)^2/2\sigma^2}p.$$

Figure 4.5 shows the two conditional pdfs. We can see that the transmitted signal X shifts the center of mass of the Gaussian pdf.

4.3   THE EXPECTED VALUE OF X

We discussed the expected value for discrete random variables in Section 3.3, and found that the sample mean of independent observations of a random variable approaches $E[X]$. Suppose we perform a series of such experiments for continuous random variables. Since continuous random variables have $P[X = x] = 0$ for any specific value of x, we divide the real line into small intervals and count the number of times $N_k(n)$ the observations fall in the interval $\{x_k < X < x_k + \Delta\}$. As n becomes large, then the relative frequency $f_k(n) = N_k(n)/n$ will approach $f_X(x_k)\Delta$, the probability of the interval. We calculate the sample mean in terms of the relative frequencies and let $n \to \infty$:

$$\langle X \rangle_n = \sum_k x_k f_k(n) \to \sum_k x_k f_X(x_k)\Delta.$$

The expression on the right-hand side approaches an integral as we decrease Δ.

The **expected value** or **mean** of a random variable X is defined by

$$E[X] = \int_{-\infty}^{+\infty} t f_X(t)\, dt. \tag{4.27}$$

The expected value $E[X]$ is defined if the above integral converges absolutely, that is,

$$E[|X|] = \int_{-\infty}^{+\infty} |t| f_X(t)\, dt \qquad < \infty.$$

If we view $f_X(x)$ as the distribution of mass on the real line, then $E[X]$ represents the center of mass of this distribution.

We already discussed $E[X]$ for discrete random variables in detail, but it is worth noting that the definition in Eq. (4.27) is applicable if we express the pdf of a discrete random variable using delta functions:

$$E[X] = \int_{-\infty}^{+\infty} t \sum_k p_X(x_k)\delta(t - x_k)\, dt$$

$$= \sum_k p_X(x_k) \int_{-\infty}^{+\infty} t \sum_k \delta(t - x_k)\, dt$$

$$= \sum_k p_X(x_k)x_k.$$

Example 4.12   Mean of a Uniform Random Variable

The mean for a uniform random variable is given by

$$E[X] = (b - a)^{-1} \int_a^b t \, dt = \frac{a + b}{2},$$

which is exactly the midpoint of the interval $[a, b]$. The results shown in Fig. 3.6 were obtained by repeating experiments in which outcomes were random variables Y and X that had uniform cdf's in the intervals $[-1, 1]$ and $[3, 7]$, respectively. The respective expected values, 0 and 5, correspond to the values about which X and Y tend to vary.

The result in Example 4.12 could have been found immediately by noting that $E[X] = m$ when the pdf is symmetric about a point m. That is, if

$$f_X(m - x) = f_X(m + x) \qquad \text{for all } x,$$

then, assuming that the mean exists,

$$0 = \int_{-\infty}^{+\infty} (m - t) f_X(t) \, dt = m - \int_{-\infty}^{+\infty} t f_X(t) \, dt.$$

The first equality above follows from the symmetry of $f_X(t)$ about $t = m$ and the odd symmetry of $(m - t)$ about the same point. We then have that $E[X] = m$.

Example 4.13   Mean of a Gaussian Random Variable

The pdf of a Gaussian random variable is symmetric about the point $x = m$. Therefore $E[X] = m$.

The following expressions are useful when X is a nonnegative random variable:

$$E[X] = \int_0^\infty (1 - F_X(t)) \, dt \qquad \text{if } X \text{ continuous and nonnegative} \qquad (4.28)$$

and

$$E[X] = \sum_{k=0}^\infty P[X > k] \qquad \text{if } X \text{ nonnegative, integer-valued.} \qquad (4.29)$$

The derivation of these formulas is discussed in Problem 4.47.

Example 4.14   Mean of Exponential Random Variable

The time X between customer arrivals at a service station has an exponential distribution. Find the mean interarrival time.

Substituting Eq. (4.17) into Eq. (4.27) we obtain

$$E[X] = \int_0^\infty t \lambda e^{-\lambda t} \, dt.$$

We evaluate the integral using integration by parts ($\int u\,dv = uv - \int v\,du$), with $u = t$ and $dv = \lambda e^{-\lambda t}\,dt$:

$$E[X] = -te^{-\lambda t}\Big|_0^\infty + \int_0^\infty e^{-\lambda t}\,dt$$

$$= \lim_{t \to \infty} te^{-\lambda t} - 0 + \left\{\frac{-e^{-\lambda t}}{\lambda}\right\}_0^\infty$$

$$= \lim_{t \to \infty}\frac{-e^{-\lambda t}}{\lambda} + \frac{1}{\lambda} = \frac{1}{\lambda},$$

where we have used the fact that $e^{-\lambda t}$ and $te^{-\lambda t}$ go to zero as t approaches infinity.

For this example, Eq. (4.28) is much easier to evaluate:

$$E[X] = \int_0^\infty e^{-\lambda t}\,dt = \frac{1}{\lambda}.$$

Recall that λ is the customer arrival rate in *customers per second*. The result that the mean inter-arrival time $E[X] = 1/\lambda$ *seconds per customer* then makes sense intuitively.

4.3.1   The Expected Value of $Y = g(X)$

Suppose that we are interested in finding the expected value of $Y = g(X)$. As in the case of discrete random variables (Eq. (3.16)), $E[Y]$ can be found directly in terms of the pdf of X:

$$E[Y] = \int_{-\infty}^\infty g(x)f_X(x)\,dx. \tag{4.30}$$

To see how Eq. (4.30) comes about, suppose that we divide the y-axis into intervals of length h, we index the intervals with the index k and we let y_k be the value in the center of the kth interval. The expected value of Y is approximated by the following sum:

$$E[Y] \simeq \sum_k y_k f_Y(y_k)h.$$

Suppose that $g(x)$ is strictly increasing, then the kth interval in the y-axis has a unique corresponding equivalent event of width h_k in the x-axis as shown in Fig. 4.6. Let x_k be the value in the kth interval such that $g(x_k) = y_k$, then since $f_Y(y_k)h = f_X(x_k)h_k$,

$$E[Y] \simeq \sum_k g(x_k)f_X(x_k)h_k.$$

By letting h approach zero, we obtain Eq. (4.30). This equation is valid even if $g(x)$ is not strictly increasing.
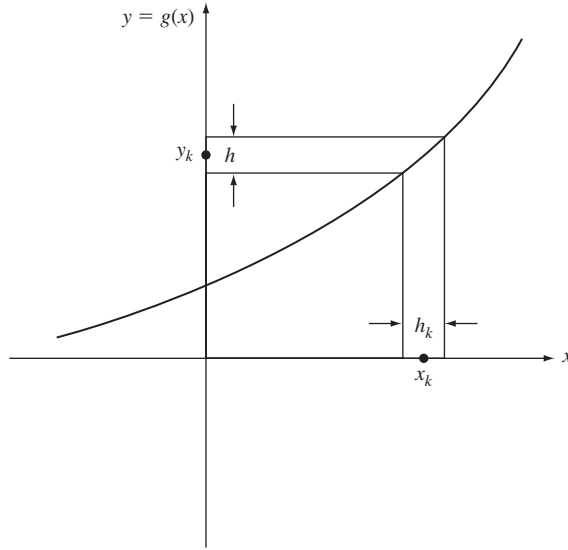
FIGURE 4.6
Two infinitesimal equivalent events.

Example 4.15   Expected Values of a Sinusoid with Random Phase

Let $Y = a \cos(\omega t + \Theta)$ where a, ω, and t are constants, and Θ is a uniform random variable in the interval $(0, 2\pi)$. The random variable Y results from sampling the amplitude of a sinusoid with random phase Θ. Find the expected value of Y and expected value of the power of Y, Y^2.

$$E[Y] = E[a \cos(\omega t + \Theta)]$$

$$= \int_0^{2\pi} a \cos(\omega t + \theta) \frac{d\theta}{2\pi} = -a \sin(\omega t + \theta)\Big|_0^{2\pi}$$

$$= -a \sin(\omega t + 2\pi) + a \sin(\omega t) = 0.$$

The average power is

$$E[Y^2] = E[a^2 \cos^2(\omega t + \Theta)] = E\left[\frac{a^2}{2} + \frac{a^2}{2}\cos(2\omega t + 2\Theta)\right]$$

$$= \frac{a^2}{2} + \frac{a^2}{2} \int_0^{2\pi} \cos(2\omega t + \theta) \frac{d\theta}{2\pi} = \frac{a^2}{2}.$$

Note that these answers are in agreement with the time averages of sinusoids: the time average ("dc" value) of the sinusoid is zero; the time-average power is $a^2/2$.

Example 4.16   Expected Values of the Indicator Function

Let $g(X) = I_C(X)$ be the indicator function for the event $\{X \text{ in } C\}$, where C is some interval or union of intervals in the real line:

$$g(X) = \begin{cases} 0 & X \text{ not in } C \\ 1 & X \text{ in } C, \end{cases}$$

then

$$E[Y] = \int_{-\infty}^{+\infty} g(X) f_X(x) \, dx = \int_C f_X(x) \, dx = P[X \text{ in } C].$$

Thus the expected value of the indicator of an event is equal to the probability of the event.

It is easy to show that Eqs. (3.17a)–(3.17e) hold for continuous random variables using Eq. (4.30). For example, let c be some constant, then

$$E[c] = \int_{-\infty}^{\infty} c f_X(x) \, dx = c \int_{-\infty}^{\infty} f_X(x) \, dx = c \tag{4.31}$$

and

$$E[cX] = \int_{-\infty}^{\infty} c x f_X(x) \, dx = c \int_{-\infty}^{\infty} x f_X(x) \, dx = c E[X]. \tag{4.32}$$

The expected value of a sum of functions of a random variable is equal to the sum of the expected values of the individual functions:

$$E[Y] = E\left[\sum_{k=1}^{n} g_k(X) \right]$$

$$= \int_{-\infty}^{\infty} \sum_{k=1}^{n} g_k(x) f_X(x) \, dx = \sum_{k=1}^{n} \int_{-\infty}^{\infty} g_k(x) f_X(x) \, dx$$

$$= \sum_{k=1}^{n} E[g_k(X)]. \tag{4.33}$$

Example 4.17

Let $Y = g(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n$, where a_k are constants, then

$$E[Y] = E[a_0] + E[a_1 X] + \cdots + E[a_n X^n]$$

$$= a_0 + a_1 E[X] + a_2 E[X^2] + \cdots + a_n E[X^n],$$

where we have used Eq. (4.33), and Eqs. (4.31) and (4.32). A special case of this result is that

$$E[X + c] = E[X] + c,$$

that is, *we can shift the mean of a random variable by adding a constant to it.*

4.3.2   Variance of *X*

The **variance of the random variable** X is defined by

$$\text{VAR}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2 \tag{4.34}$$

The **standard deviation of the random variable** X is defined by

$$\text{STD}[X] = \text{VAR}[X]^{1/2}. \tag{4.35}$$

Example 4.18   Variance of Uniform Random Variable

Find the variance of the random variable X that is uniformly distributed in the interval $[a, b]$.

Since the mean of X is $(a + b)/2$,

$$\text{VAR}[X] = \frac{1}{b - a} \int_a^b \left(x - \frac{a + b}{2} \right)^2 dx.$$

Let $y = (x - (a + b)/2)$,

$$\text{VAR}[X] = \frac{1}{b - a} \int_{-(b-a)/2}^{(b-a)/2} y^2 \, dy = \frac{(b - a)^2}{12}.$$

The random variables in Fig. 3.6 were uniformly distributed in the interval $[-1, 1]$ and $[3, 7]$, respectively. Their variances are then 1/3 and 4/3. The corresponding standard deviations are 0.577 and 1.155.

Example 4.19   Variance of Gaussian Random Variable

Find the variance of a Gaussian random variable.

First multiply the integral of the pdf of X by $\sqrt{2\pi}\,\sigma$ to obtain

$$\int_{-\infty}^{\infty} e^{-(x-m)^2/2\sigma^2} \, dx = \sqrt{2\pi}\,\sigma.$$

Differentiate both sides with respect to σ:

$$\int_{-\infty}^{\infty} \left(\frac{(x - m)^2}{\sigma^3} \right) e^{-(x-m)^2/2\sigma^2} \, dx = \sqrt{2\pi}.$$

By rearranging the above equation, we obtain

$$\text{VAR}[X] = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} (x - m)^2 e^{-(x-m)^2/2\sigma^2} \, dx = \sigma^2.$$

This result can also be obtained by direct integration. (See Problem 4.46.) Figure 4.7 shows the Gaussian pdf for several values of σ; it is evident that the "width" of the pdf increases with σ.

The following properties were derived in Section 3.3:

$$\text{VAR}[c] = 0 \tag{4.36}$$

$$\text{VAR}[X + c] = \text{VAR}[X] \tag{4.37}$$

$$\text{VAR}[cX] = c^2 \, \text{VAR}[X], \tag{4.38}$$

where c is a constant.
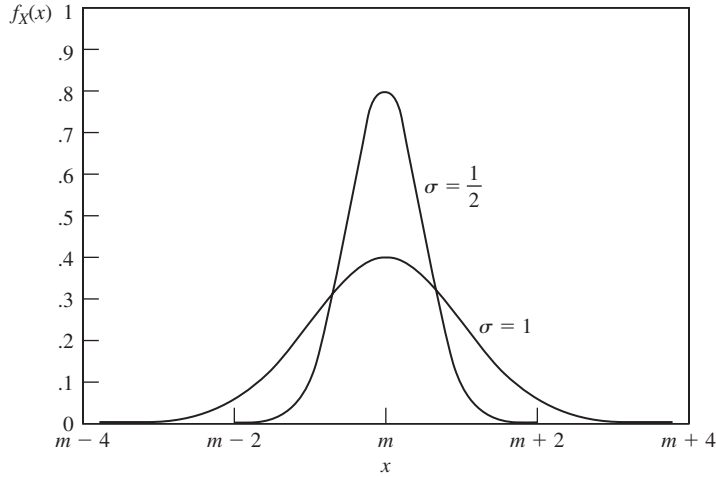
FIGURE 4.7
Probability density function of Gaussian random variable.

The mean and variance are the two most important parameters used in summarizing the pdf of a random variable. Other parameters are occasionally used. For example, the skewness defined by $E[(X - E[X])^3]/STD[X]^3$ measures the degree of asymmetry about the mean. It is easy to show that if a pdf is symmetric about its mean, then its skewness is zero. The point to note with these parameters of the pdf is that each involves the expected value of a higher power of X. Indeed we show in a later section that, under certain conditions, a pdf is completely specified if the expected values of all the powers of X are known. These expected values are called the moments of X.

The **_n_th moment of the random variable X** is defined by

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x) \, dx. \tag{4.39}$$

The mean and variance can be seen to be defined in terms of the first two moments, $E[X]$ and $E[X^2]$.

---

*Example 4.20   Analog-to-Digital Conversion: A Detailed Example

A quantizer is used to convert an analog signal (e.g., speech or audio) into digital form. A quantizer maps a random voltage X into the nearest point $q(X)$ from a set of 2^R representation values as shown in Fig. 4.8(a). The value X is then approximated by $q(X)$, which is identified by an R-bit binary number. In this manner, an "analog" voltage X that can assume a continuum of values is converted into an R-bit number.

The quantizer introduces an error $Z = X - q(X)$ as shown in Fig. 4.8(b). Note that Z is a function of X and that it ranges in value between $-d/2$ and $d/2$, where d is the quantizer step size. Suppose that X has a uniform distribution in the interval $[-x_{\max}, x_{\max}]$, that the quantizer has 2^R levels, and that $2x_{\max} = 2^R d$. It is easy to show that Z is uniformly distributed in the interval $[-d/2, d/2]$ (see Problem 4.93).
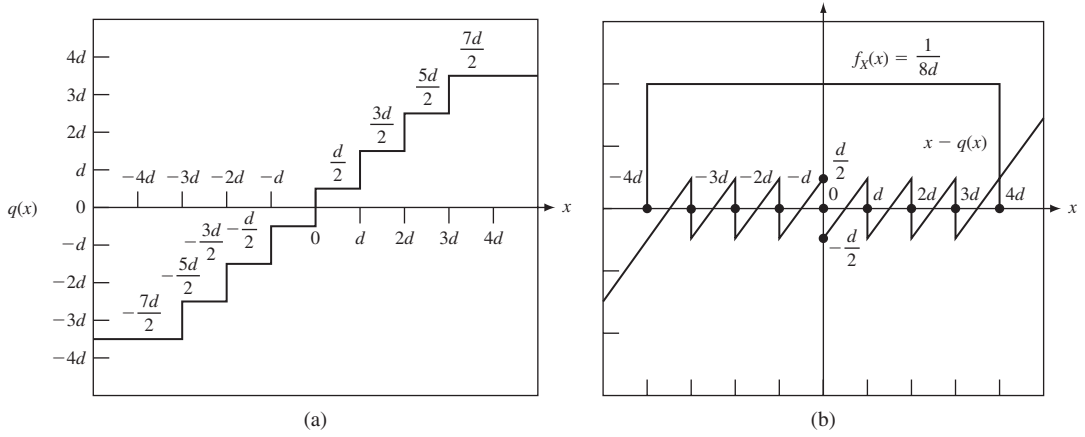
FIGURE 4.8
(a) A uniform quantizer maps the input x into the closest point from the set $\{\pm d/2, \pm 3d/2, \pm 5d/2, \pm 7d/2\}$. (b) The uniform quantizer error for the input x is $x - q(x)$.

Therefore from Example 4.12,

$$E[Z] = \frac{d/2 - d/2}{2} = 0.$$

The error Z thus has mean zero.

By Example 4.18,

$$\text{VAR}[Z] = \frac{(d/2 - (-d/2))^2}{12} = \frac{d^2}{12}.$$

This result is approximately correct for any pdf that is approximately flat over each quantizer interval. This is the case when 2^R is large.

The approximation $q(x)$ can be viewed as a "noisy" version of X since

$$Q(X) = X - Z,$$

where Z is the quantization error Z. The measure of goodness of a quantizer is specified by the SNR ratio, which is defined as the ratio of the variance of the "signal" X to the variance of the distortion or "noise" Z:

$$\text{SNR} = \frac{\text{VAR}[X]}{\text{VAR}[Z]} = \frac{\text{VAR}[X]}{d^2/12}$$

$$= \frac{\text{VAR}[X]}{x_{\max}^2/3} 2^{2R},$$

where we have used the fact that $d = 2x_{\max}/2^R$. When X is nonuniform, the value $x_{\max}$ is selected so that $P[|X| > x_{\max}]$ is small. A typical choice is $x_{\max} = 4\,\text{STD}[X]$. The SNR is then

$$\text{SNR} = \frac{3}{16} 2^{2R}.$$

This important formula is often quoted in decibels:

$$\text{SNR dB} = 10 \log_{10} \text{SNR} = 6R - 7.3 \text{ dB}.$$

The SNR increases by a factor of 4 (6 dB) with each additional bit used to represent X. This makes sense since each additional bit doubles the number of quantizer levels, which in turn reduces the step size by a factor of 2. The variance of the error should then be reduced by the square of this, namely $2^2 = 4$.

4.4   IMPORTANT CONTINUOUS RANDOM VARIABLES

We are always limited to measurements of finite precision, so in effect, every random variable found in practice is a discrete random variable. Nevertheless, there are several compelling reasons for using continuous random variable models. First, in general, continuous random variables are easier to handle analytically. Second, the limiting form of many discrete random variables yields continuous random variables. Finally, there are a number of "families" of continuous random variables that can be used to model a wide variety of situations by adjusting a few parameters. In this section we continue our introduction of important random variables. Table 4.1 lists some of the more important continuous random variables.

4.4.1   The Uniform Random Variable

The uniform random variable arises in situations where all values in an interval of the real line are equally likely to occur. The uniform random variable U in the interval $[a, b]$ has pdf:

$$f_U(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \quad \text{and} \quad x > b \end{cases} \tag{4.40}$$

and cdf

$$F_U(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b. \end{cases} \tag{4.41}$$

See Figure 4.2. The mean and variance of U are given by:

$$E[U] = \frac{a+b}{2} \quad \text{and} \quad \text{VAR}[X] = \frac{(b-a)^2}{2}. \tag{4.42}$$

The uniform random variable appears in many situations that involve equally likely continuous random variables. Obviously U can only be defined over intervals that are finite in length. We will see in Section 4.9 that the uniform random variable plays a crucial role in generating random variables in computer simulation models.

4.4.2   The Exponential Random Variable

The exponential random variable arises in the modeling of the time between occurrence of events (e.g., the time between customer demands for call connections), and in the modeling of the lifetime of devices and systems. The **exponential random variable** X with parameter λ has pdf

TABLE 4.1   Continuous random variables.

Uniform Random Variable

$S_X = [a, b]$

$f_X(x) = \dfrac{1}{b - a} \qquad a \le x \le b$

$E[X] = \dfrac{a + b}{2} \qquad \text{VAR}[X] = \dfrac{(b - a)^2}{12} \qquad \Phi_X(\omega) = \dfrac{e^{j\omega b} - e^{j\omega a}}{j\omega(b - a)}$

Exponential Random Variable

$S_X = [0, \infty)$

$f_X(x) = \lambda e^{-\lambda x} \qquad x \ge 0 \quad \text{and} \quad \lambda > 0$

$E[X] = \dfrac{1}{\lambda} \qquad \text{VAR}[X] = \dfrac{1}{\lambda^2} \qquad \Phi_X(\omega) = \dfrac{\lambda}{\lambda - j\omega}$

Remarks: The exponential random variable is the only continuous random variable with the memoryless property.

Gaussian (Normal) Random Variable

$S_X = (-\infty, +\infty)$

$f_X(x) = \dfrac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi}\sigma} \qquad -\infty < x < +\infty \quad \text{and} \quad \sigma > 0$

$E[X] = m \qquad \text{VAR}[X] = \sigma^2 \qquad \Phi_X(\omega) = e^{jm\omega - \sigma^2\omega^2/2}$

Remarks: Under a wide range of conditions X can be used to approximate the sum of a large number of independent random variables.

Gamma Random Variable

$S_X = (0, +\infty)$

$f_X(x) = \dfrac{\lambda(\lambda x)^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} \qquad x > 0 \quad \text{and} \quad \alpha > 0, \lambda > 0$

where $\Gamma(z)$ is the gamma function (Eq. 4.56).

$E[X] = \alpha/\lambda \qquad \text{VAR}[X] = \alpha/\lambda^2 \qquad \Phi_X(\omega) = \dfrac{1}{(1 - j\omega/\lambda)^{\alpha}}$

Special Cases of Gamma Random Variable

m–1 Erlang Random Variable: $\alpha = m$, a positive integer

$f_X(x) = \dfrac{\lambda e^{-\lambda x}(\lambda x)^{m-2}}{(m - 1)!} \qquad x > 0 \qquad \Phi_X(\omega) = \left(\dfrac{1}{1 - j\omega/\lambda}\right)^m$

Remarks: An m–1 Erlang random variable is obtained by adding m independent exponentially distributed random variables with parameter λ.

Chi-Square Random Variable with k degrees of freedom: $\alpha = k/2$, k a positive integer, and $\lambda = 1/2$

$f_X(x) = \dfrac{x^{(k-2)/2}e^{-x/2}}{2^{k/2}\Gamma(k/2)} \qquad x > 0 \qquad \Phi_X(\omega) = \left(\dfrac{1}{1 - 2j\omega}\right)^{k/2}$

Remarks: The sum of k mutually independent, squared zero-mean, unit-variance Gaussian random variables is a chi-square random variable with k degrees of freedom.

Laplacian Random Variable

$S_X = (-\infty, \infty)$

$f_X(x) = \dfrac{\alpha}{2} e^{-\alpha|x|}$ $-\infty < x < +\infty$ and $\alpha > 0$

$E[X] = 0$ $\text{VAR}[X] = 2/\alpha^2$ $\Phi_X(\omega) = \dfrac{\alpha^2}{\omega^2 + \alpha^2}$

Rayleigh Random Variable

$S_X = [0, \infty)$

$f_X(x) = \dfrac{x}{\alpha^2} e^{-x^2/2\alpha^2}$ $x \geq 0$ and $\alpha > 0$

$E[X] = \alpha\sqrt{\pi/2}$ $\text{VAR}[X] = (2 - \pi/2)\alpha^2$

Cauchy Random Variable

$S_X = (-\infty, +\infty)$

$f_X(x) = \dfrac{\alpha/\pi}{x^2 + \alpha^2}$ $-\infty < x < +\infty$ and $\alpha > 0$

Mean and variance do not exist. $\Phi_X(\omega) = e^{-\alpha|\omega|}$

Pareto Random Variable

$S_X = [x_m, \infty) x_m > 0.$

$$f_X(x) = \begin{cases} 0 & x < x_m \\ \alpha\dfrac{x_m^\alpha}{x^{\alpha+1}} & x \geq x_m \end{cases}$$

$E[X] = \dfrac{\alpha x_m}{\alpha - 1}$ for $\alpha > 1$ $\text{VAR}[X] = \dfrac{\alpha x_m^2}{(\alpha - 2)(\alpha - 1)^2}$ for $\alpha > 2$

Remarks: The Pareto random variable is the most prominent example of random variables with "long tails," and can be viewed as a continuous version of the Zipf discrete random variable.

Beta Random Variable

$$f_X(x) = \begin{cases} \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)} x^{\alpha-1}(1 - x)^{\beta-1} & 0 < x < 1 \quad \text{and} \quad \alpha > 0, \beta > 0 \\ 0 & \text{otherwise} \end{cases}$$

$E[X] = \dfrac{\alpha}{\alpha + \beta}$ $\text{VAR}[X] = \dfrac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

Remarks: The beta random variable is useful for modeling a variety of pdf shapes for random variables that range over finite intervals.

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases} \tag{4.43}$$

and cdf

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0. \end{cases} \tag{4.44}$$

The cdf and pdf of X are shown in Fig. 4.9.

The parameter λ is the rate at which events occur, so in Eq. (4.44) the probability of an event occurring by time x increases at the rate λ increases. Recall from Example 3.31 that the interarrival times between events in a Poisson process (Fig. 3.10) is an exponential random variable.

The mean and variance of X are given by:

$$E[U] = \frac{1}{\lambda} \quad \text{and} \quad VAR[X] = \frac{1}{\lambda^2}. \tag{4.45}$$

In event interarrival situations, λ is in units of events/second and $1/\lambda$ is in units of seconds per event interarrival.

The exponential random variable satisfies the **memoryless property**:

$$P[X > t + h \mid X > t] = P[X > h]. \tag{4.46}$$

The expression on the left side is the probability of having to wait at least h additional seconds given that one has already been waiting t seconds. The expression on the right side is the probability of waiting at least h seconds when one first begins to wait. Thus the probability of waiting at least an additional h seconds is the same regardless of how long one has already been waiting! We see later in the book that the memoryless property of the exponential random variable makes it the cornerstone for the theory of
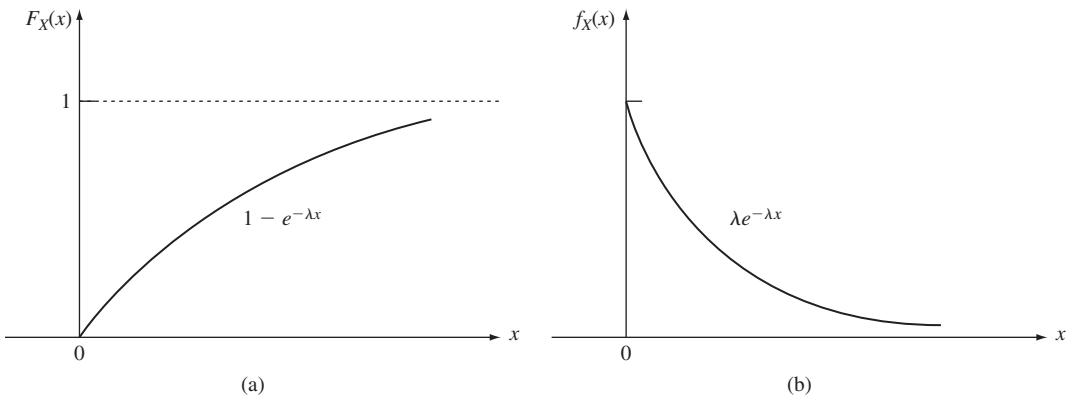
FIGURE 4.9
An example of a continuous random variable—the exponential random variable. Part (a) is the cdf and part (b) is the pdf.

Markov chains, which is used extensively in evaluating the performance of computer systems and communications networks.

We now prove the memoryless property:

$$P[X > t + h \,|\, X > t] = \frac{P[\{X > t + h\} \cap \{X > t\}]}{P[X > t]} \qquad \text{for } h > 0$$

$$= \frac{P[X > t + h]}{P[X > t]} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}}$$

$$= e^{-\lambda h} = P[X > h].$$

It can be shown that the exponential random variable is the only continuous random variable that satisfies the memoryless property.

Examples 2.13, 2.28, and 2.30 dealt with the exponential random variable.

4.4.3   The Gaussian (Normal) Random Variable

There are many situations in manmade and in natural phenomena where one deals with a random variable X that consists of the sum of a large number of "small" random variables. The exact description of the pdf of X in terms of the component random variables can become quite complex and unwieldy. However, one finds that under very general conditions, as the number of components becomes large, the cdf of X approaches that of the **Gaussian (normal) random variable.**[1] This random variable appears so often in problems involving randomness that it has come to be known as the "normal" random variable.

The pdf for the Gaussian random variable X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2} \qquad -\infty < x < \infty, \tag{4.47}$$

where m and $\sigma > 0$ are real numbers, which we showed in Examples 4.13 and 4.19 to be the mean and standard deviation of X. Figure 4.7 shows that the Gaussian pdf is a "bell-shaped" curve centered and symmetric about m and whose "width" increases with σ.

The cdf of the Gaussian random variable is given by

$$P[X \le x] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-(x'-m)^2/2\sigma^2} \, dx'. \tag{4.48}$$

The change of variable $t = (x' - m)/\sigma$ results in

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-m)/\sigma} e^{-t^2/2} \, dt$$

$$= \Phi\!\left(\frac{x - m}{\sigma}\right) \tag{4.49}$$

where $\Phi(x)$ is the cdf of a Gaussian random variable with $m = 0$ and $\sigma = 1$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} \, dt. \tag{4.50}$$

---

[1]This result, called the central limit theorem, will be discussed in Chapter 7.

Therefore any probability involving an arbitrary Gaussian random variable can be expressed in terms of $\Phi(x)$.

Example 4.21

Show that the Gaussian pdf integrates to one. Consider the square of the integral of the pdf:

$$\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\, dx\right]^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-x^2/2}\, dx \int_{-\infty}^{\infty} e^{-y^2/2}\, dy$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy.$$

Let $x = r\cos\theta$ and $y = r\sin\theta$ and carry out the change from Cartesian to polar coordinates, then we obtain:

$$\frac{1}{2\pi}\int_{0}^{\infty}\int_{0}^{2\pi} e^{-r^2/2} r\, dr\, d\theta = \int_{0}^{\infty} re^{-r^2/2}\, dr$$

$$= [-e^{-r^2/2}]_{0}^{\infty}$$

$$= 1.$$

In electrical engineering it is customary to work with the Q-function, which is defined by

$$Q(x) = 1 - \Phi(x) \tag{4.51}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{x}^{\infty} e^{-t^2/2}\, dt. \tag{4.52}$$

$Q(x)$ is simply the probability of the "tail" of the pdf. The symmetry of the pdf implies that

$$Q(0) = 1/2 \quad \text{and} \quad Q(-x) = 1 - Q(x). \tag{4.53}$$

The integral in Eq. (4.50) does not have a closed-form expression. Traditionally the integrals have been evaluated by looking up tables that list $Q(x)$ or by using approximations that require numerical evaluation [Ross]. The following expression has been found to give good accuracy for $Q(x)$ over the entire range $0 < x < \infty$:

$$Q(x) \simeq \left[\frac{1}{(1-a)x + a\sqrt{x^2 + b}}\right]\frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \tag{4.54}$$

where $a = 1/\pi$ and $b = 2\pi$ [Gallager]. Table 4.2 shows $Q(x)$ and the value given by the above approximation. In some problems, we are interested in finding the value of x for which $Q(x) = 10^{-k}$. Table 4.3 gives these values for $k = 1, \dots, 10$.

The Gaussian random variable plays a very important role in communication systems, where transmission signals are corrupted by noise voltages resulting from the thermal motion of electrons. It can be shown from physical principles that these voltages will have a Gaussian pdf.

TABLE 4.2   Comparison of $Q(x)$ and approximation given by Eq. (4.54).

x	$Q(x)$	Approximation	x	$Q(x)$	Approximation
0	5.00E-01	5.00E-01	2.7	3.47E-03	3.46E-03
0.1	4.60E-01	4.58E-01	2.8	2.56E-03	2.55E-03
0.2	4.21E-01	4.17E-01	2.9	1.87E-03	1.86E-03
0.3	3.82E-01	3.78E-01	3.0	1.35E-03	1.35E-03
0.4	3.45E-01	3.41E-01	3.1	9.68E-04	9.66E-04
0.5	3.09E-01	3.05E-01	3.2	6.87E-04	6.86E-04
0.6	2.74E-01	2.71E-01	3.3	4.83E-04	4.83E-04
0.7	2.42E-01	2.39E-01	3.4	3.37E-04	3.36E-04
0.8	2.12E-01	2.09E-01	3.5	2.33E-04	2.32E-04
0.9	1.84E-01	1.82E-01	3.6	1.59E-04	1.59E-04
1.0	1.59E-01	1.57E-01	3.7	1.08E-04	1.08E-04
1.1	1.36E-01	1.34E-01	3.8	7.24E-05	7.23E-05
1.2	1.15E-01	1.14E-01	3.9	4.81E-05	4.81E-05
1.3	9.68E-02	9.60E-02	4.0	3.17E-05	3.16E-05
1.4	8.08E-02	8.01E-02	4.5	3.40E-06	3.40E-06
1.5	6.68E-02	6.63E-02	5.0	2.87E-07	2.87E-07
1.6	5.48E-02	5.44E-02	5.5	1.90E-08	1.90E-08
1.7	4.46E-02	4.43E-02	6.0	9.87E-10	9.86E-10
1.8	3.59E-02	3.57E-02	6.5	4.02E-11	4.02E-11
1.9	2.87E-02	2.86E-02	7.0	1.28E-12	1.28E-12
2.0	2.28E-02	2.26E-02	7.5	3.19E-14	3.19E-14
2.1	1.79E-02	1.78E-02	8.0	6.22E-16	6.22E-16
2.2	1.39E-02	1.39E-02	8.5	9.48E-18	9.48E-18
2.3	1.07E-02	1.07E-02	9.0	1.13E-19	1.13E-19
2.4	8.20E-03	8.17E-03	9.5	1.05E-21	1.05E-21
2.5	6.21E-03	6.19E-03	10.0	7.62E-24	7.62E-24
2.6	4.66E-03	4.65E-03			

Example 4.22

A communication system accepts a positive voltage V as input and outputs a voltage $Y = \alpha V + N$, where $\alpha = 10^{-2}$ and N is a Gaussian random variable with parameters $m = 0$ and $\sigma = 2$. Find the value of V that gives $P[Y < 0] = 10^{-6}$.

The probability $P[Y < 0]$ is written in terms of N as follows:

$$P[Y < 0] = P[\alpha V + N < 0]$$

$$= P[N < -\alpha V] = \Phi\left(\frac{-\alpha V}{\sigma}\right) = Q\left(\frac{\alpha V}{\sigma}\right) = 10^{-6}.$$

From Table 4.3 we see that the argument of the Q-function should be $\alpha V/\sigma = 4.753$. Thus $V = (4.753)\sigma/\alpha = 950.6$.

TABLE 4.3 $Q(x) = 10^{-k}$

k	$x = Q^{-1}(10^{-k})$
1	1.2815
2	2.3263
3	3.0902
4	3.7190
5	4.2649
6	4.7535
7	5.1993
8	5.6120
9	5.9978
10	6.3613

4.4.4 The Gamma Random Variable

The gamma random variable is a versatile random variable that appears in many applications. For example, it is used to model the time required to service customers in queueing systems, the lifetime of devices and systems in reliability studies, and the defect clustering behavior in VLSI chips.

The pdf of the **gamma random variable** has two parameters, $\alpha > 0$ and $\lambda > 0$, and is given by

$$f_X(x) = \frac{\lambda(\lambda x)^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} \qquad 0 < x < \infty, \tag{4.55}$$

where $\Gamma(z)$ is the gamma function, which is defined by the integral

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}\,dx \qquad z > 0. \tag{4.56}$$

The gamma function has the following properties:

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

$$\Gamma(z + 1) = z\Gamma(z) \qquad \text{for } z > 0, \text{ and}$$
$$\Gamma(m + 1) = m! \qquad \text{for } m \text{ a nonnegative integer.}$$

The versatility of the gamma random variable is due to the richness of the gamma function $\Gamma(z)$. The pdf of the gamma random variable can assume a variety of shapes as shown in Fig. 4.10. By varying the parameters α and λ it is possible to fit the gamma pdf to many types of experimental data. In addition, many random variables are special cases of the gamma random variable. The exponential random variable is obtained by letting $\alpha = 1$. By letting $\lambda = 1/2$ and $\alpha = k/2$, where k is a positive integer, we obtain the **chi-square random variable**, which appears in certain statistical problems. The **m-Erlang random variable** is obtained when $\alpha = m$, a positive integer. The m-Erlang random variable is used in the system reliability models and in queueing systems models. Both of these random variables are discussed in later examples.
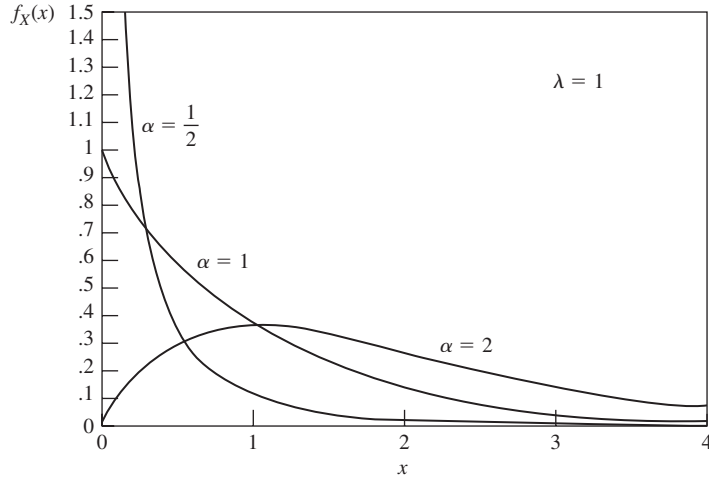
FIGURE 4.10
Probability density function of gamma random variable.

Example 4.23

Show that the pdf of a gamma random variable integrates to one.

The integral of the pdf is

$$\int_0^\infty f_X(x)\,dx = \int_0^\infty \frac{\lambda(\lambda x)^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}\,dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-\lambda x}\,dx.$$

Let $y = \lambda x$, then $dx = dy/\lambda$ and the integral becomes

$$\frac{\lambda^\alpha}{\Gamma(\alpha)\lambda^\alpha}\int_0^\infty y^{\alpha-1}e^{-y}\,dy = 1,$$

where we used the fact that the integral equals $\Gamma(\alpha)$.

In general, the cdf of the gamma random variable does not have a closed-form expression. We will show that the special case of the m-Erlang random variable does have a closed-form expression for the cdf by using its close interrelation with the exponential and Poisson random variables. The cdf can also be obtained by integration of the pdf (see Problem 4.74).

Consider once again the limiting procedure that was used to derive the Poisson random variable. Suppose that we observe the time S_m that elapses until the occurrence of the mth event. The times $X_1, X_2, \ldots, X_m$ between events are exponential random variables, so we must have

$$S_m = X_1 + X_2 + \cdots + X_m.$$

We will show that S_m is an m-Erlang random variable. To find the cdf of S_m, let $N(t)$ be the Poisson random variable for the number of events in t seconds. Note that the mth event occurs before time t—that is, $S_m \leq t$—if and only if m or more events occur in t seconds, namely $N(t) \geq m$. The reasoning goes as follows. If the mth event has occurred before time t, then it follows that m or more events will occur in time t. On the other hand, if m or more events occur in time t, then it follows that the mth event occurred by time t. Thus

$$F_{S_m}(t) = P[S_m \leq t] = P[N(t) \geq m] \tag{4.57}$$

$$= 1 - \sum_{k=0}^{m-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \tag{4.58}$$

where we have used the result of Example 3.31. If we take the derivative of the above cdf, we finally obtain the pdf of the m-Erlang random variable. Thus we have shown that S_m is an m-Erlang random variable.

Example 4.24

A factory has two spares of a critical system component that has an average lifetime of $1/\lambda = 1$ month. Find the probability that the three components (the operating one and the two spares) will last more than 6 months. Assume the component lifetimes are exponential random variables.

    The remaining lifetime of the component in service is an exponential random variable with rate λ by the memoryless property. Thus, the total lifetime X of the three components is the sum of three exponential random variables with parameter $\lambda = 1$. Thus X has a 3-Erlang distribution with $\lambda = 1$. From Eq. (4.58) the probability that X is greater than 6 is

$$P[X > 6] = 1 - P[X \leq 6]$$

$$= \sum_{k=0}^{2} \frac{6^k}{k!} e^{-6} = .06197.$$

4.4.5   The Beta Random Variable

The beta random variable X assumes values over a closed interval and has pdf:

$$f_X(x) = cx^{a-1}(1 - x)^{b-1} \qquad \text{for } 0 < x < 1 \tag{4.59}$$

where the normalization constant is the reciprocal of the beta function

$$\frac{1}{c} = B(a, b) = \int_0^1 x^{a-1}(1 - x)^{b-1} \, dx$$

and where the beta function is related to the gamma function by the following expression:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}.$$

When $a = b = 1$, we have the uniform random variable. Other choices of a and b give pdfs over finite intervals that can differ markedly from the uniform. See Problem 4.75. If

$a = b > 1$, then the pdf is symmetric about $x = 1/2$ and is concentrated about $x = 1/2$ as well. When $a = b < 1$, then the pdf is symmetric but the density is concentrated at the edges of the interval. When $a < b$ (or $a > b$) the pdf is skewed to the right (or left).

The mean and variance are given by:

$$E[X] = \frac{a}{a + b} \quad \text{and} \quad \text{VAR}[X] = \frac{ab}{(a + b)^2(a + b + 1)}. \tag{4.60}$$

The versatility of the pdf of the beta random variable makes it useful to model a variety of behaviors for random variables that range over finite intervals. For example, in a Bernoulli trial experiment, the probability of success p could itself be a random variable. The beta pdf is frequently used to model p.

4.4.6   The Cauchy Random Variable

The Cauchy random variable X assumes values over the entire real line and has pdf:

$$f_X(x) = \frac{1/\pi}{1 + x^2}. \tag{4.61}$$

It is easy to verify that this pdf integrates to 1. However, X does not have any moments since the associated integrals do not converge. The Cauchy random variable arises as the tangent of a uniform random variable in the unit interval.

4.4.7   The Pareto Random Variable

The Pareto random variable arises in the study of the distribution of wealth where it has been found to model the tendency for a small portion of the population to own a large portion of the wealth. Recently the Pareto distribution has been found to capture the behavior of many quantities of interest in the study of Internet behavior, e.g., sizes of files, packet delays, audio and video title preferences, session times in peer-to-peer networks, etc. The Pareto random variable can be viewed as a continuous version of the Zipf discrete random variable.

The Pareto random variable X takes on values in the range $x > x_m$, where x_m is a positive real number. X has complementary cdf with shape parameter $\alpha > 0$ given by:

$$P[X > x] = \begin{cases} 1 & x < x_m \\ \dfrac{x_m^{\alpha}}{x^{\alpha}} & x \geq x_m. \end{cases} \tag{4.62}$$

The tail of X decays algebraically with x which is rather slower in comparison to the exponential and Gaussian random variables. The Pareto random variable is the most prominent example of random variables with "long tails."

The cdf and pdf of X are:

$$F_X(x) = \begin{cases} 0 & x < x_m \\ 1 - \dfrac{x_m^{\alpha}}{x^{\alpha}} & x \geq x_m. \end{cases} \tag{4.63}$$

Because of its long tail, the cdf of X approaches 1 rather slowly as x increases.

$$f_X(x) = \begin{cases} 0 & x < x_m \\ \alpha \dfrac{x_m^\alpha}{x^{\alpha+1}} & x \geq x_m. \end{cases} \tag{4.64}$$

Example 4.25   Mean and Variance of Pareto Random Variable

Find the mean and variance of the Pareto random variable.

$$E[X] = \int_{x_m}^\infty t\alpha \frac{x_m^\alpha}{t^{\alpha+1}} dt = \int_{x_m}^\infty \alpha \frac{x_m^\alpha}{t^\alpha} dt = \frac{\alpha}{\alpha - 1} \frac{x_m^\alpha}{x_m^{\alpha-1}} = \frac{\alpha x_m}{\alpha - 1} \qquad \text{for } \alpha > 1 \quad (4.65)$$

where the integral is defined for $\alpha > 1$, and

$$E[X^2] = \int_{x_m}^\infty t^2 \alpha \frac{x_m^\alpha}{t^{\alpha+1}} dt = \int_{x_m}^\infty \alpha \frac{x_m^\alpha}{t^{\alpha-1}} dt = \frac{\alpha}{\alpha - 2} \frac{x_m^\alpha}{x_m^{\alpha-2}} = \frac{\alpha x_m^2}{\alpha - 2} \qquad \text{for } \alpha > 2$$

where the second moment is defined for $\alpha > 2$.

The variance of X is then:

$$\text{VAR}[X] = \frac{\alpha x_m^2}{\alpha - 2} - \left(\frac{\alpha x_m^2}{\alpha - 1} \right)^2 = \frac{\alpha x_m^2}{(\alpha - 2)(\alpha - 1)^2} \qquad \text{for } \alpha > 2. \tag{4.66}$$

4.5   FUNCTIONS OF A RANDOM VARIABLE

Let X be a random variable and let $g(x)$ be a real-valued function defined on the real line. Define $Y = g(X)$, that is, Y is determined by evaluating the function $g(x)$ at the value assumed by the random variable X. Then Y is also a random variable. The probabilities with which Y takes on various values depend on the function $g(x)$ as well as the cumulative distribution function of X. In this section we consider the problem of finding the cdf and pdf of Y.

Example 4.26

Let the function $h(x) = (x)^+$ be defined as follows:

$$(x)^+ = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0. \end{cases}$$

For example, let X be the number of active speakers in a group of N speakers, and let Y be the number of active speakers in excess of M, then $Y = (X - M)^+$. In another example, let X be a voltage input to a halfwave rectifier, then $Y = (X)^+$ is the output.

Example 4.27

Let the function $q(x)$ be defined as shown in Fig. 4.8(a), where the set of points on the real line are mapped into the nearest representation point from the set $S_Y = \{-3.5d, -2.5d, -1.5d, -0.5d, 0.5d, 1.5d, 2.5d, 3.5d\}$. Thus, for example, all the points in the interval $(0, d)$ are mapped into the point $d/2$. The function $q(x)$ represents an eight-level uniform quantizer.

Example 4.28

Consider the linear function $c(x) = ax + b$, where a and b are constants. This function arises in many situations. For example, $c(x)$ could be the cost associated with the quantity x, with the constant a being the cost per unit of x, and b being a fixed cost component. In a signal processing context, $c(x) = ax$ could be the amplified version (if $a > 1$) or attenuated version (if $a < 1$) of the voltage x.

The probability of an event C involving Y is equal to the probability of the equivalent event B of values of X such that $g(X)$ is in C:

$$P[Y \text{ in } C] = P[g(X) \text{ in } C] = P[X \text{ in } B].$$

Three types of equivalent events are useful in determining the cdf and pdf of $Y = g(X)$: (1) The event $\{g(X) = y_k\}$ is used to determine the magnitude of the jump at a point y_k where the cdf of Y is known to have a discontinuity; (2) the event $\{g(X) \le y\}$ is used to find the cdf of Y directly; and (3) the event $\{y < g(X) \le y + h\}$ is useful in determining the pdf of Y. We will demonstrate the use of these three methods in a series of examples.

The next two examples demonstrate how the pmf is computed in cases where $Y = g(X)$ is discrete. In the first example, X is discrete. In the second example, X is continuous.

Example 4.29

Let X be the number of active speakers in a group of N independent speakers. Let p be the probability that a speaker is active. In Example 2.39 it was shown that X has a binomial distribution with parameters N and p. Suppose that a voice transmission system can transmit up to M voice signals at a time, and that when X exceeds M, $X - M$ randomly selected signals are discarded. Let Y be the number of signals discarded, then

$$Y = (X - M)^+.$$

Y takes on values from the set $S_Y = \{0, 1, \ldots, N - M\}$. Y will equal zero whenever X is less than or equal to M, and Y will equal $k > 0$ when X is equal to $M + k$. Therefore

$$P[Y = 0] = P[X \text{ in } \{0, 1, \ldots, M\}] = \sum_{j=0}^{M} p_j$$

and

$$P[Y = k] = P[X = M + k] = p_{M+k} \qquad 0 < k \le N - M,$$

where p_j is the pmf of X.

Example 4.30

Let X be a sample voltage of a speech waveform, and suppose that X has a uniform distribution in the interval $[-4d, 4d]$. Let $Y = q(X)$, where the quantizer input-output characteristic is as shown in Fig. 4.10. Find the pmf for Y.

The event $\{Y = q\}$ for q in S_Y is equivalent to the event $\{X$ in $I_q\}$, where I_q is an interval of points mapped into the representation point q. The pmf of Y is therefore found by evaluating

$$P[Y = q] = \int_{I_q} f_X(t) \, dt.$$

It is easy to see that the representation point has an interval of length d mapped into it. Thus the eight possible outputs are equiprobable, that is, $P[Y = q] = 1/8$ for q in S_Y.

In Example 4.30, each constant section of the function $q(X)$ produces a delta function in the pdf of Y. In general, if the function $g(X)$ is constant during certain intervals and if the pdf of X is nonzero in these intervals, then the pdf of Y will contain delta functions. Y will then be either discrete or of mixed type.

The cdf of Y is defined as the probability of the event $\{Y \leq y\}$. In principle, it can always be obtained by finding the probability of the equivalent event $\{g(X) \leq y\}$ as shown in the next examples.

Example 4.31 A Linear Function

Let the random variable Y be defined by

$$Y = aX + b,$$

where a is a nonzero constant. Suppose that X has cdf $F_X(x)$, then find $F_Y(y)$.

The event $\{Y \leq y\}$ occurs when $A = \{aX + b \leq y\}$ occurs. If $a > 0$, then $A = \{X \leq (y - b)/a\}$ (see Fig. 4.11), and thus

$$F_Y(y) = P\left[X \leq \frac{y - b}{a}\right] = F_X\left(\frac{y - b}{a}\right) \qquad a > 0.$$

On the other hand, if $a < 0$, then $A = \{X \geq (y - b)/a\}$, and

$$F_Y(y) = P\left[X \geq \frac{y - b}{a}\right] = 1 - F_X\left(\frac{y - b}{a}\right) \qquad a < 0.$$

We can obtain the pdf of Y by differentiating with respect to y. To do this we need to use the chain rule for derivatives:

$$\frac{dF}{dy} = \frac{dF}{du}\frac{du}{dy},$$

where u is the argument of F. In this case, $u = (y - b)/a$, and we then obtain

$$f_Y(y) = \frac{1}{a}f_X\left(\frac{y - b}{a}\right) \qquad a > 0$$
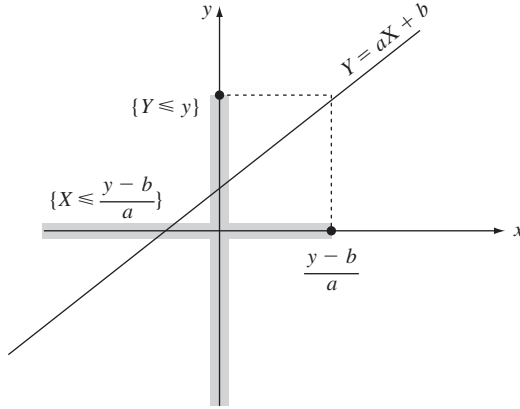
FIGURE 4.11
The equivalent event for $\{Y \leq y\}$ is the event
$\{X \leq (y - b)/a\}$, if $a > 0$.

and

$$f_Y(y) = \frac{1}{-a} f_X\left(\frac{y - b}{a}\right) \qquad a < 0.$$

The above two results can be written compactly as

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right). \tag{4.67}$$

Example 4.32   A Linear Function of a Gaussian Random Variable

Let X be a random variable with a Gaussian pdf with mean m and standard deviation σ:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m)^2/2\sigma^2} \qquad -\infty < x < \infty. \tag{4.68}$$

Let $Y = aX + b$, then find the pdf of Y.
    Substitution of Eq. (4.68) into Eq. (4.67) yields

$$f_Y(y) = \frac{1}{\sqrt{2\pi}|a\sigma|} e^{-(y-b-am)^2/2(a\sigma)^2}.$$

Note that Y also has a Gaussian distribution with mean $b + am$ and standard deviation $|a|\,\sigma$.
Therefore *a linear function of a Gaussian random variable is also a Gaussian random variable.*

Example 4.33

Let the random variable Y be defined by

$$Y = X^2,$$

where X is a continuous random variable. Find the cdf and pdf of Y.
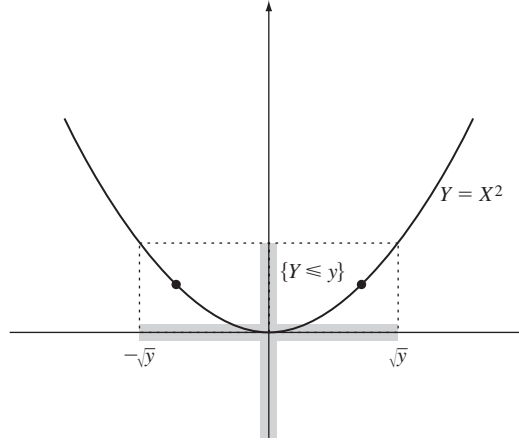
FIGURE 4.12
The equivalent event for $\{Y \le y\}$ is the event
$\{-\sqrt{y} \le X \le \sqrt{y}\}$, if $y \ge 0$.

The event $\{Y \le y\}$ occurs when $\{X^2 \le y\}$ or equivalently when $\{-\sqrt{y} \le X \le \sqrt{y}\}$ for y nonnegative; see Fig. 4.12. The event is null when y is negative. Thus

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

and differentiating with respect to y,

$$f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} - \frac{f_X(-\sqrt{y})}{-2\sqrt{y}} \qquad y > 0$$

$$= \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}. \tag{4.69}$$

Example 4.34   A Chi-Square Random Variable

Let X be a Gaussian random variable with mean $m = 0$ and standard deviation $\sigma = 1$. X is then said to be a standard normal random variable. Let $Y = X^2$. Find the pdf of Y.

Substitution of Eq. (4.68) into Eq. (4.69) yields

$$f_Y(y) = \frac{e^{-y/2}}{\sqrt{2y\pi}} \qquad y \ge 0. \tag{4.70}$$

From Table 4.1 we see that $f_Y(y)$ is the pdf of a *chi-square random variable with one degree of freedom*.

The result in Example 4.33 suggests that if the equation $y_0 = g(x)$ has n solutions, $x_0, x_1, \ldots, x_n$, then $f_Y(y_0)$ will be equal to n terms of the type on the right-hand
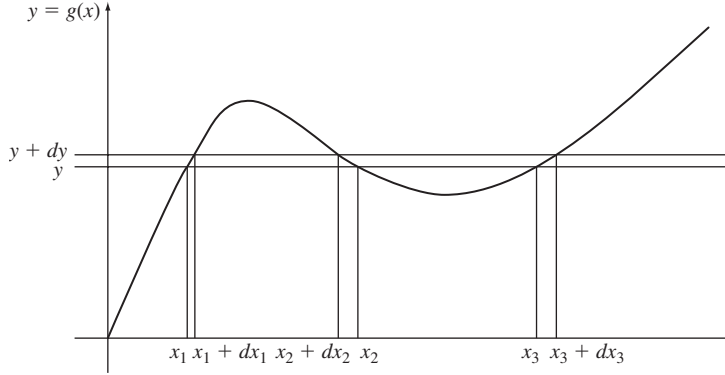
FIGURE 4.13
The equivalent event of $\{y < Y < y + dy\}$ is $\{x_1 < X < x_1 + dx_1\}$
$\cup \{x_2 + dx_2 < X < x_2\} \cup \{x_3 < X < x_3 + dx_3\}$.

side of Eq. (4.69). We now show that this is generally true by using a method for direct-ly obtaining the pdf of Y in terms of the pdf of X.

Consider a nonlinear function $Y = g(X)$ such as the one shown in Fig. 4.13. Con-sider the event $C_y = \{y < Y < y + dy\}$ and let B_y be its equivalent event. For y indi-cated in the figure, the equation $g(x) = y$ has three solutions x_1, x_2, and x_3, and the equivalent event B_y has a segment corresponding to each solution:

$$B_y = \{x_1 < X < x_1 + dx_1\} \cup \{x_2 + dx_2 < X < x_2\}$$

$$\cup \{x_3 < X < x_3 + dx_3\}.$$

The probability of the event C_y is approximately

$$P[C_y] = f_Y(y)|dy|, \tag{4.71}$$

where $|dy|$ is the length of the interval $y < Y \le y + dy$. Similarly, the probability of the event B_y is approximately

$$P[B_y] = f_X(x_1)|dx_1| + f_X(x_2)|dx_2| + f_X(x_3)|dx_3|. \tag{4.72}$$

Since C_y and B_y are equivalent events, their probabilities must be equal. By equating Eqs. (4.71) and (4.72) we obtain

$$f_Y(y) = \sum_k \frac{f_X(x)}{|dy/dx|}\bigg|_{x=x_k} \tag{4.73}$$

$$= \sum_k f_X(x)\left|\frac{dx}{dy}\right|\bigg|_{x=x_k}. \tag{4.74}$$

It is clear that if the equation $g(x) = y$ has n solutions, the expression for the pdf of Y at that point is given by Eqs. (4.73) and (4.74), and contains n terms.

Example 4.35

Let $Y = X^2$ as in Example 4.34. For $y \geq 0$, the equation $y = x^2$ has two solutions, $x_0 = \sqrt{y}$ and $x_1 = -\sqrt{y}$, so Eq. (4.73) has two terms. Since $dy/dx = 2x$, Eq. (4.73) yields

$$f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}.$$

This result is in agreement with Eq. (4.69). To use Eq. (4.74), we note that

$$\frac{dx}{dy} = \frac{d}{dy} \pm \sqrt{y} = \pm\frac{1}{2\sqrt{y}},$$

which when substituted into Eq. (4.74) then yields Eq. (4.69) again.

Example 4.36   Amplitude Samples of a Sinusoidal Waveform

Let $Y = \cos(X)$, where X is uniformly distributed in the interval $(0, 2\pi]$. Y can be viewed as the sample of a sinusoidal waveform at a random instant of time that is uniformly distributed over the period of the sinusoid. Find the pdf of Y.

It can be seen in Fig. 4.14 that for $-1 < y < 1$ the equation $y = \cos(x)$ has two solutions in the interval of interest, $x_0 = \cos^{-1}(y)$ and $x_1 = 2\pi - x_0$. Since (see an introductory calculus textbook)

$$\left.\frac{dy}{dx}\right|_{x_0} = -\sin(x_0) = -\sin(\cos^{-1}(y)) = -\sqrt{1 - y^2},$$

and since $f_X(x) = 1/2\pi$ in the interval of interest, Eq. (4.73) yields

$$f_Y(y) = \frac{1}{2\pi\sqrt{1 - y^2}} + \frac{1}{2\pi\sqrt{1 - y^2}}$$

$$= \frac{1}{\pi\sqrt{1 - y^2}} \qquad \text{for } -1 < y < 1.$$
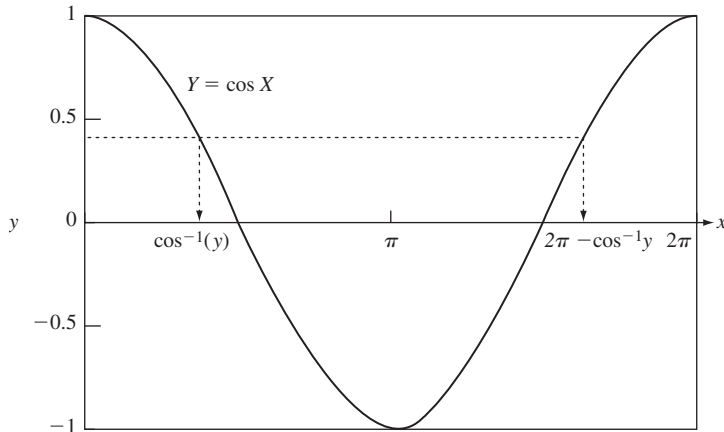
FIGURE 4.14
$y = \cos x$ has two roots in the interval $(0, 2\pi)$.

The cdf of Y is found by integrating the above:

$$F_Y(y) = \begin{cases} 0 & y < -1 \\ \dfrac{1}{2} + \dfrac{\sin^{-1}y}{\pi} & -1 \le y \le 1 \\ 1 & y > 1. \end{cases}$$

Y is said to have the **arcsine distribution**.

4.6   THE MARKOV AND CHEBYSHEV INEQUALITIES

In general, the mean and variance of a random variable do not provide enough information to determine the cdf/pdf. However, the mean and variance of a random variable X do allow us to obtain bounds for probabilities of the form $P[|X| \ge t]$. Suppose first that X is a nonnegative random variable with mean $E[X]$. The **Markov inequality** then states that

$$P[X \ge a] \le \frac{E[X]}{a} \qquad \text{for } X \text{ nonnegative.} \tag{4.75}$$

We obtain Eq. (4.75) as follows:

$$E[X] = \int_0^a tf_X(t) \, dt + \int_a^\infty tf_X(t) \, dt \ge \int_a^\infty tf_X(t) \, dt$$

$$\ge \int_a^\infty af_X(t) \, dt = aP[X \ge a].$$

The first inequality results from discarding the integral from zero to a; the second inequality results from replacing t with the smaller number a.

Example 4.37

The mean height of children in a kindergarten class is 3 feet, 6 inches. Find the bound on the probability that a kid in the class is taller than 9 feet. The Markov inequality gives $P[H \ge 9] \le 42/108 = .389$.

The bound in the above example appears to be ridiculous. However, a bound, by its very nature, must take the worst case into consideration. One can easily construct a random variable for which the bound given by the Markov inequality is exact. The reason we know that the bound in the above example is ridiculous is that we have knowledge about the variability of the children's height about their mean.

Now suppose that the mean $E[X] = m$ and the variance $\text{VAR}[X] = \sigma^2$ of a random variable are known, and that we are interested in bounding $P[|X - m| \ge a]$. The **Chebyshev inequality** states that

$$P[|X - m| \ge a] \le \frac{\sigma^2}{a^2}. \tag{4.76}$$

The Chebyshev inequality is a consequence of the Markov inequality. Let $D^2 = (X - m)^2$ be the squared deviation from the mean. Then the Markov inequality applied to D^2 gives

$$P[D^2 \geq a^2] \leq \frac{E[(X - m)^2]}{a^2} = \frac{\sigma^2}{a^2}.$$

Equation (4.76) follows when we note that $\{D^2 \geq a^2\}$ and $\{|X - m| \geq a\}$ are equivalent events.

Suppose that a random variable X has zero variance; then the Chebyshev inequality implies that

$$P[X = m] = 1, \qquad (4.77)$$

that is, the random variable is equal to its mean with probability one. In other words, X is equal to the constant m in almost all experiments.

Example 4.38

The mean response time and the standard deviation in a multi-user computer system are known to be 15 seconds and 3 seconds, respectively. Estimate the probability that the response time is more than 5 seconds from the mean.

The Chebyshev inequality with $m = 15$ seconds, $\sigma = 3$ seconds, and $a = 5$ seconds gives

$$P[|X - 15| \geq 5] \leq \frac{9}{25} = .36.$$

Example 4.39

If X has mean m and variance σ^2, then the Chebyshev inequality for $a = k\sigma$ gives

$$P[|X - m| \geq k\sigma] \leq \frac{1}{k^2}.$$

Now suppose that we know that X is a Gaussian random variable, then for $k = 2$, $P[|X - m| \geq 2\sigma] = .0456$, whereas the Chebyshev inequality gives the upper bound .25.

Example 4.40   Chebyshev Bound Is Tight

Let the random variable X have $P[X = -v] = P[X = v] = 0.5$. The mean is zero and the variance is $\text{VAR}[X] = E[X^2] = (-v)^2 0.5 + v^2 0.5 = v^2$.

Note that $P[|X| \geq v] = 1$. The Chebyshev inequality states:

$$P[|X| \geq v] \leq 1 - \frac{\text{VAR}[X]}{v^2} = 1.$$

We see that the bound and the exact value are in agreement, so the bound is tight.

We see from Example 4.38 that for certain random variables, the Chebyshev inequality can give rather loose bounds. Nevertheless, the inequality is useful in situations in which we have no knowledge about the distribution of a given random variable other than its mean and variance. In Section 7.2, we will use the Chebyshev inequality to prove that the arithmetic average of independent measurements of the same random variable is highly likely to be close to the expected value of the random variable when the number of measurements is large. Problems 4.100 and 4.101 give examples of this result.

If more information is available than just the mean and variance, then it is possible to obtain bounds that are tighter than the Markov and Chebyshev inequalities. Consider the Markov inequality again. The region of interest is $A = \{t \geq a\}$, so let $I_A(t)$ be the indicator function, that is, $I_A(t) = 1$ if $t \in A$ and $I_A(t) = 0$ otherwise. The key step in the derivation is to note that $t/a \geq 1$ in the region of interest. In effect we bounded $I_A(t)$ by t/a as shown in Fig. 4.15. We then have:

$$P[X \geq a] = \int_0^\infty I_A(t)f_X(t)\,dt \leq \int_0^\infty \frac{t}{a}f_X(t)\,dt = \frac{E[X]}{a}.$$

By changing the upper bound on $I_A(t)$, we can obtain different bounds on $P[X \geq a]$. Consider the bound $I_A(t) \leq e^{s(t-a)}$, also shown in Fig. 4.15, where $s > 0$. The resulting bound is:

$$P[X \geq a] = \int_0^\infty I_A(t)f_X(t)\,dt \leq \int_0^\infty e^{s(t-a)}f_X(t)\,dt$$

$$= e^{-sa}\int_0^\infty e^{st}f_X(t)\,dt = e^{-sa}E[e^{sX}]. \tag{4.78}$$

This bound is called the **Chernoff bound**, which can be seen to depend on the expected value of an exponential function of X. This function is called the moment generating function and is related to the transforms that are introduced in the next section. We develop the Chernoff bound further in the next section.
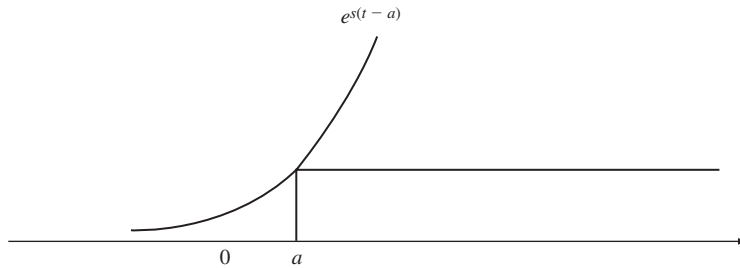
FIGURE 4.15
Bounds on indicator function for $A = \{t \geq a\}$.

4.7 TRANSFORM METHODS

In the old days, before calculators and computers, it was very handy to have logarithm tables around if your work involved performing a large number of multiplications. If you wanted to multiply the numbers x and y, you looked up $\log(x)$ and $\log(y)$, *added* $\log(x)$ and $\log(y)$, and then looked up the inverse logarithm of the result. You probably remember from grade school that longhand multiplication is more tedious and error-prone than addition. Thus logarithms were very useful as a computational aid.

Transform methods are extremely useful computational aids in the solution of equations that involve derivatives and integrals of functions. In many of these problems, the solution is given by the convolution of two functions: $f_1(x) * f_2(x)$. We will define the convolution operation later. For now, all you need to know is that finding the convolution of two functions can be more tedious and error-prone than longhand multiplication! In this section we introduce transforms that map the function $f_k(x)$ into another function $\mathcal{F}_k(\omega)$, and that satisfy the property that $\mathcal{F}[f_1(x) * f_2(x)] = \mathcal{F}_1(\omega)\mathcal{F}_2(\omega)$. In other words, the transform of the convolution is equal to the product of the individual transforms. Therefore transforms allow us to replace the convolution operation by the much simpler multiplication operation. The transform expressions introduced in this section will prove very useful when we consider sums of random variables in Chapter 7.

4.7.1 The Characteristic Function

The **characteristic function** of a random variable X is defined by

$$\Phi_X(\omega) = E[e^{j\omega X}] \tag{4.79a}$$

$$= \int_{-\infty}^{\infty} f_X(x)e^{j\omega x}\, dx, \tag{4.79b}$$

where $j = \sqrt{-1}$ is the imaginary unit number. The two expressions on the right-hand side motivate two interpretations of the characteristic function. In the first expression, $\Phi_X(\omega)$ can be viewed as the expected value of a function of $X, e^{j\omega X}$, in which the parameter ω is left unspecified. In the second expression, $\Phi_X(\omega)$ is simply the Fourier transform of the pdf $f_X(x)$ (with a reversal in the sign of the exponent). Both of these interpretations prove useful in different contexts.

If we view $\Phi_X(\omega)$ as a Fourier transform, then we have from the Fourier transform inversion formula that the pdf of X is given by

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\Phi_X(\omega)e^{-j\omega x}\, d\omega. \tag{4.80}$$

It then follows that every pdf and its characteristic function form a unique Fourier transform pair. Table 4.1 gives the characteristic function of some continuous random variables.

Example 4.41   Exponential Random Variable

The characteristic function for an exponentially distributed random variable with parameter λ is given by

$$\Phi_X(\omega) = \int_0^\infty \lambda e^{-\lambda x} e^{j\omega x}\, dx = \int_0^\infty \lambda e^{-(\lambda - j\omega)x}\, dx$$

$$= \frac{\lambda}{\lambda - j\omega}.$$

If X is a discrete random variable, substitution of Eq. (4.20) into the definition of $\Phi_X(\omega)$ gives

$$\Phi_X(\omega) = \sum_k p_X(x_k) e^{j\omega x_k} \qquad \text{discrete random variables.}$$

Most of the time we deal with discrete random variables that are integer-valued. The characteristic function is then

$$\Phi_X(\omega) = \sum_{k=-\infty}^{\infty} p_X(k) e^{j\omega k} \qquad \text{integer-valued random variables.} \qquad (4.81)$$

Equation (4.81) is the **Fourier transform of the sequence** $p_X(k)$. Note that the Fourier transform in Eq. (4.81) is a periodic function of ω with period 2π, since $e^{j(\omega + 2\pi)k} = e^{j\omega k} e^{jk2\pi}$ and $e^{jk2\pi} = 1$. Therefore the characteristic function of integer-valued random variables is a periodic function of ω. The following inversion formula allows us to recover the probabilities $p_X(k)$ from $\Phi_X(\omega)$:

$$p_X(k) = \frac{1}{2\pi} \int_0^{2\pi} \Phi_X(\omega) e^{-j\omega k}\, d\omega \qquad k = 0, \pm 1, \pm 2, \dots \qquad (4.82)$$

Indeed, a comparison of Eqs. (4.81) and (4.82) shows that the $p_X(k)$ are simply the co-efficients of the Fourier series of the periodic function $\Phi_X(\omega)$.

Example 4.42   Geometric Random Variable

The characteristic function for a geometric random variable is given by

$$\Phi_X(\omega) = \sum_{k=0}^{\infty} pq^k e^{j\omega k} = p \sum_{k=0}^{\infty} (q e^{j\omega})^k$$

$$= \frac{p}{1 - q e^{j\omega}}.$$

Since $f_X(x)$ and $\Phi_X(\omega)$ form a transform pair, we would expect to be able to obtain the moments of X from $\Phi_X(\omega)$. The **moment theorem** states that the moments of

X are given by

$$E[X^n] = \frac{1}{j^n} \frac{d^n}{d\omega^n} \Phi_X(\omega) \bigg|_{\omega=0}. \tag{4.83}$$

To show this, first expand $e^{j\omega x}$ in a power series in the definition of $\Phi_X(\omega)$:

$$\Phi_X(\omega) = \int_{-\infty}^{\infty} f_X(x) \left\{ 1 + j\omega X + \frac{(j\omega X)^2}{2!} + \cdots \right\} dx.$$

Assuming that all the moments of X are finite and that the series can be integrated term by term, we obtain

$$\Phi_X(\omega) = 1 + j\omega E[X] + \frac{(j\omega)^2 E[X^2]}{2!} + \cdots + \frac{(j\omega)^n E[X^n]}{n!} + \cdots.$$

If we differentiate the above expression once and evaluate the result at $\omega = 0$ we obtain

$$\frac{d}{d\omega} \Phi_X(\omega) \bigg|_{\omega=0} = jE[X].$$

If we differentiate n times and evaluate at $\omega = 0$, we finally obtain

$$\frac{d^n}{d\omega^n} \Phi_X(\omega) \bigg|_{\omega=0} = j^n E[X^n],$$

which yields Eq. (4.83).

Note that when the above power series converges, the characteristic function and hence the pdf by Eq. (4.80) are completely determined by the moments of X.

Example 4.43

To find the mean of an exponentially distributed random variable, we differentiate $\Phi_X(\omega) = \lambda(\lambda - j\omega)^{-1}$ once, and obtain

$$\Phi'_X(\omega) = \frac{\lambda j}{(\lambda - j\omega)^2}.$$

The moment theorem then implies that $E[X] = \Phi'_X(0)/j = 1/\lambda$.

If we take two derivatives, we obtain

$$\Phi''_X(\omega) = \frac{-2\lambda}{(\lambda - j\omega)^3},$$

so the second moment is then $E[X^2] = \Phi''_X(0)/j^2 = 2/\lambda^2$. The variance of X is then given by

$$\text{VAR}[X] = E[X^2] - E[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Example 4.44   Chernoff Bound for Gaussian Random Variable

Let X be a Gaussian random variable with mean m and variance σ^2. Find the Chernoff bound for X.

The Chernoff bound (Eq. 4.78) depends on the moment generating function:

$$E[e^{sX}] = \Phi_X(-js).$$

In terms of the characteristic function the bound is given by:

$$P[X \geq a] \leq e^{-sa}\Phi_X(-js) \quad \text{for} \quad s \geq 0.$$

The parameter s can be selected to minimize the upper bound.

The bound for the Gaussian random variable is:

$$P[X \geq a] \leq e^{-sa}e^{ms+\sigma^2 s^2/2} = e^{-s(a-m)+\sigma^2 s^2/2} \quad \text{for} \quad s \geq 0.$$

We minimize the upper bound by minimizing the exponent:

$$0 = \frac{d}{ds}(-s(a-m) + \sigma^2 s^2/2) \quad \text{which implies } s = \frac{a-m}{\sigma^2}.$$

The resulting upper bound is:

$$P[X \geq a] = Q\left(\frac{a-m}{\sigma}\right) \leq e^{-(a-m)^2/2\sigma^2}.$$

This bound is much better than the Chebyshev bound and is similar to the estimate given in Eq. (4.54).

4.7.2   The Probability Generating Function

In problems where random variables are nonnegative, it is usually more convenient to use the z-transform or the Laplace transform. The **probability generating function** $G_N(z)$ of a nonnegative integer-valued random variable N is defined by

$$G_N(z) = E[z^N] \tag{4.84a}$$

$$= \sum_{k=0}^{\infty} p_N(k)z^k. \tag{4.84b}$$

The first expression is the expected value of the function of N, z^N. The second expression is the z-transform of the pmf (with a sign change in the exponent). Table 3.1 shows the probability generating function for some discrete random variables. Note that the characteristic function of N is given by $\Phi_N(\omega) = G_N(e^{j\omega})$.

Using a derivation similar to that used in the moment theorem, it is easy to show that the pmf of N is given by

$$p_N(k) = \frac{1}{k!}\frac{d^k}{dz^k}G_N(z)\Bigg|_{z=0}. \tag{4.85}$$

This is why $G_N(z)$ is called the probability generating function. By taking the first two derivatives of $G_N(z)$ and evaluating the result at $z = 1$, it is possible to find the first

two moments of X:

$$\frac{d}{dz}G_N(z)\bigg|_{z=1} = \sum_{k=0}^{\infty} p_N(k)kz^{k-1}\bigg|_{z=1} = \sum_{k=0}^{\infty} kp_N(k) = E[N]$$

and

$$\frac{d^2}{dz^2}G_N(z)\bigg|_{z=1} = \sum_{k=0}^{\infty} p_N(k)k(k-1)z^{k-2}\bigg|_{z=1}$$

$$= \sum_{k=0}^{\infty} k(k-1)p_N(k) = E[N(N-1)] = E[N^2] - E[N].$$

Thus the mean and variance of X are given by

$$E[N] = G'_N(1) \tag{4.86}$$

and

$$\text{VAR}[N] = G''_N(1) + G'_N(1) - (G'_N(1))^2. \tag{4.87}$$

Example 4.45   Poisson Random Variable

The probability generating function for the Poisson random variable with parameter α is given by

$$G_N(z) = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}e^{-\alpha}z^k = e^{-\alpha}\sum_{k=0}^{\infty} \frac{(\alpha z)^k}{k!}$$

$$= e^{-\alpha}e^{\alpha z} = e^{\alpha(z-1)}.$$

The first two derivatives of $G_N(z)$ are given by

$$G'_N(z) = \alpha e^{\alpha(z-1)}$$

and

$$G''_N(z) = \alpha^2 e^{\alpha(z-1)}.$$

Therefore the mean and variance of the Poisson are

$$E[N] = \alpha$$

$$\text{VAR}[N] = \alpha^2 + \alpha - \alpha^2 = \alpha.$$

4.7.3   The Laplace Transform of the pdf

In queueing theory one deals with service times, waiting times, and delays. All of these are nonnegative continuous random variables. It is therefore customary to work with the **Laplace transform** of the pdf,

$$X^*(s) = \int_0^{\infty} f_X(x)e^{-sx}\,dx = E[e^{-sX}]. \tag{4.88}$$

Note that $X^*(s)$ can be interpreted as a Laplace transform of the pdf or as an expected value of a function of X, e^{-sX}.

The moment theorem also holds for $X^*(s)$:

$$E[X^n] = (-1)^n \frac{d^n}{ds^n} X^*(s) \Big|_{s=0}. \tag{4.89}$$

Example 4.46   Gamma Random Variable

The Laplace transform of the gamma pdf is given by

$$X^*(s) = \int_0^\infty \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x} e^{-sx}}{\Gamma(\alpha)} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda+s)x} dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{1}{(\lambda+s)^\alpha} \int_0^\infty y^{\alpha-1} e^{-y} dy = \frac{\lambda^\alpha}{(\lambda+s)^\alpha},$$

where we used the change of variable $y = (\lambda + s)x$. We can then obtain the first two moments of X as follows:

$$E[X] = -\frac{d}{ds} \frac{\lambda^\alpha}{(\lambda+s)^\alpha} \Big|_{s=0} = \frac{\alpha\lambda^\alpha}{(\lambda+s)^{\alpha+1}} \Big|_{s=0} = \frac{\alpha}{\lambda}$$

and

$$E[X^2] = \frac{d^2}{ds^2} \frac{\lambda^\alpha}{(\lambda+s)^\alpha} \Big|_{s=0} = \frac{\alpha(\alpha+1)\lambda^\alpha}{(\lambda+s)^{\alpha+2}} \Big|_{s=0} = \frac{\alpha(\alpha+1)}{\lambda^2}.$$

Thus the variance of X is

$$\text{VAR}(X) = E[X^2] - E[X]^2 = \frac{\alpha}{\lambda^2}.$$

4.8   BASIC RELIABILITY CALCULATIONS

In this section we apply some of the tools developed so far to the calculation of measures that are of interest in assessing the reliability of systems. We also show how the reliability of a system can be determined in terms of the reliability of its components.

4.8.1   The Failure Rate Function

Let T be the lifetime of a component, a subsystem, or a system. The **reliability** at time t is defined as the probability that the component, subsystem, or system is still functioning at time t:

$$R(t) = P[T > t]. \tag{4.90}$$

The relative frequency interpretation implies that, in a large number of components or systems, $R(t)$ is the fraction that fail after time t. The reliability can be expressed in terms of the cdf of T:

$$R(t) = 1 - P[T \le t] = 1 - F_T(t). \tag{4.91}$$

Note that the derivative of $R(t)$ gives the negative of the pdf of T:

$$R'(t) = -f_T(t).\tag{4.92}$$

The **mean time to failure (MTTF)** is given by the expected value of T:

$$E[T] = \int_0^\infty f_T(t)\,dt = \int_0^\infty R(t)\,dt,$$

where the second expression was obtained using Eqs. (4.28) and (4.91).

Suppose that we know a system is still functioning at time t; what is its future behavior? In Example 4.10, we found that the conditional cdf of T given that $T > t$ is given by

$$F_T(x|T > t) = P[T \le x|T > t]$$

$$= \begin{cases} 0 & x < t \\ \dfrac{F_T(x) - F_T(t)}{1 - F_T(t)} & x \ge t. \end{cases}\tag{4.93}$$

The pdf associated with $F_T(x|T > t)$ is

$$f_T(x|T > t) = \frac{f_T(x)}{1 - F_T(t)} \quad x \ge t.\tag{4.94}$$

Note that the denominator of Eq. (4.94) is equal to $R(t)$.

The **failure rate function** $r(t)$ is defined as $f_T(x|T > t)$ evaluated at $x = t$:

$$r(t) = f_T(t|T > t)$$

$$= \frac{-R'(t)}{R(t)},\tag{4.95}$$

since by Eq. (4.92), $R'(t) = -f_T(t)$. The failure rate function has the following meaning:

$$P[t < T \le t + dt|T > t] = f_T(t|T > t)\,dt = r(t)\,dt.\tag{4.96}$$

In words, $r(t)\,dt$ is the probability that a component that has functioned up to time t will fail in the next dt seconds.

Example 4.47   Exponential Failure Law

Suppose a component has a constant failure rate function, say $r(t) = \lambda$. Find the pdf and the MTTF for its lifetime T.

Equation (4.95) implies that

$$\frac{R'(t)}{R(t)} = -\lambda.\tag{4.97}$$

Equation (4.97) is a first-order differential equation with initial condition $R(0) = 1$. If we integrate both sides of Eq. (4.97) from 0 to t, we obtain

$$-\int_0^t \lambda\,dt' + k = \int_0^t \frac{R'(t')}{R(t')}\,dt' = \ln R(t),$$

which implies that

$$R(t) = Ke^{-\lambda t}, \quad \text{where } K = e^k.$$

The initial condition $R(0) = 1$ implies that $K = 1$. Thus

$$R(t) = e^{-\lambda t} \quad t > 0 \tag{4.98}$$

and

$$f_T(t) = \lambda e^{-\lambda t} \quad t > 0.$$

Thus if T has a constant failure rate function, then T is an exponential random variable. This is not surprising, since the exponential random variable satisfies the memoryless property. The MTTF $= E[T] = 1/\lambda$.

The derivation that was used in Example 4.47 can be used to show that, in general, the failure rate function and the reliability are related by

$$R(t) = \exp\left\{ -\int_0^t r(t') \, dt' \right\} \tag{4.99}$$

and from Eq. (4.92),

$$f_T(t) = r(t) \exp\left\{ -\int_0^t r(t') \, dt' \right\}. \tag{4.100}$$

Figure 4.16 shows the failure rate function for a typical system. Initially there may be a high failure rate due to defective parts or installation. After the "bugs" have been worked out, the system is stable and has a low failure rate. At some later point, ageing and wear effects set in, resulting in an increased failure rate. Equations (4.99) and (4.100) allow us to postulate reliability functions and the associated pdf's in terms of the failure rate function, as shown in the following example.
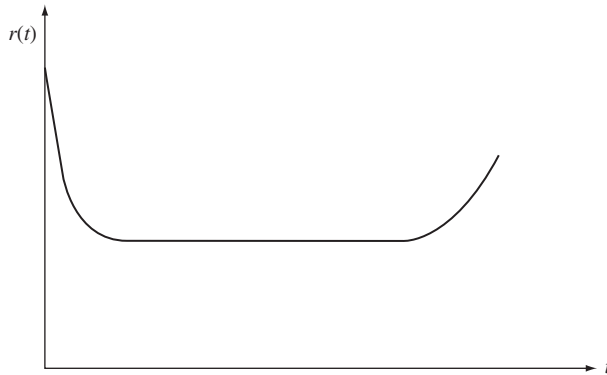
FIGURE 4.16
Failure rate function for a typical system.

Example 4.48 Weibull Failure Law

The Weibull failure law has failure rate function given by

$$r(t) = \alpha\beta t^{\beta-1}, \tag{4.101}$$

where α and β are positive constants. Equation (4.99) implies that the reliability is given by

$$R(t) = e^{-\alpha t^{\beta}}.$$

Equation (4.100) then implies that the pdf for T is

$$f_T(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}} \quad t > 0. \tag{4.102}$$

Figure 4.17 shows $f_T(t)$ for $\alpha = 1$ and several values of β. Note that $\beta = 1$ yields the exponential failure law, which has a constant failure rate. For $\beta > 1$, Eq. (4.101) gives a failure rate function that increases with time. For $\beta < 1$, Eq. (4.101) gives a failure rate function that decreases with time. Further properties of the Weibull random variable are developed in the problems.

4.8.2 Reliability of Systems

Suppose that a system consists of several components or subsystems. We now show how the reliability of a system can be computed in terms of the reliability of its subsystems if the components are assumed to fail independently of each other.
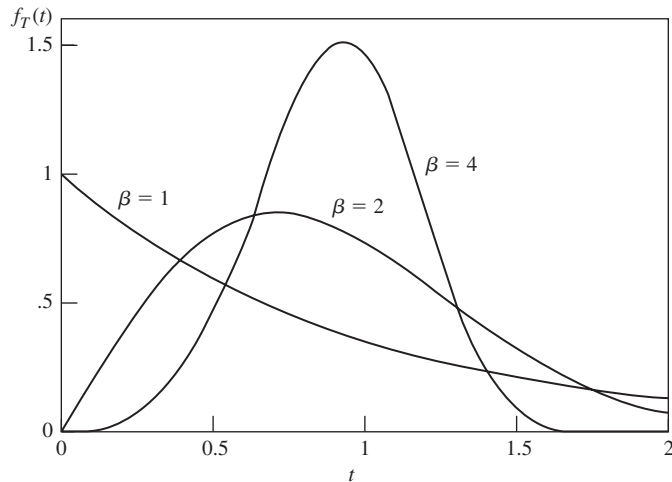
FIGURE 4.17
Probability density function of Weibull random variable, $\alpha = 1$ and $\beta = 1, 2, 4$.

(a)

(b)
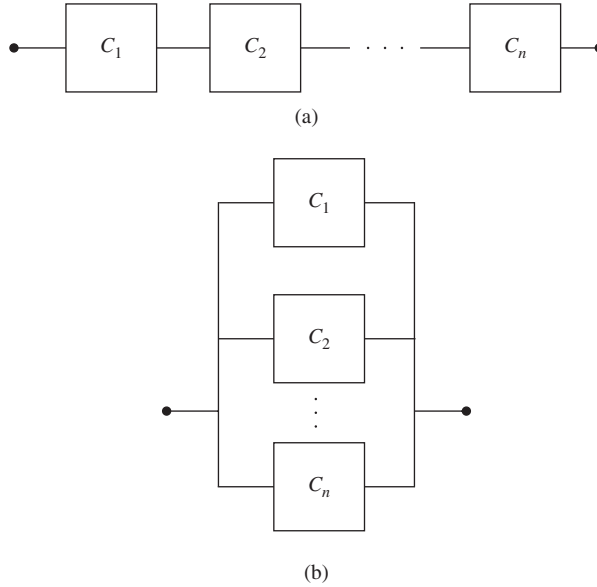
FIGURE 4.18
(a) System consisting of n components in series. (b) System consisting
of n components in parallel.

Consider first a system that consists of the series arrangement of n components as shown in Fig. 4.18(a). This system is considered to be functioning only if all the components are functioning. Let A_s be the event "system functioning at time t," and let A_j be the event "jth component is functioning at time t," then the probability that the system is functioning at time t is

$$R(t) = P[A_s]$$
$$= P[A_1 \cap A_2 \cap \cdots \cap A_n] = P[A_1]P[A_2]\ldots P[A_n]$$
$$= R_1(t)R_2(t)\ldots R_n(t), \tag{4.103}$$

since $P[A_j] = R_j(t)$, the reliability function of the jth component. Since probabilities are numbers that are less than or equal to one, we see that $R(t)$ can be no more reliable than the least reliable of the components, that is, $R(t) \leq \min_j R_j(t)$.

If we apply Eq. (4.99) to each of the $R_j(t)$ in Eq. (4.103), we then find that the failure rate function of a series system is given by the sum of the component failure rate functions:

$$R(t) = \exp\left\{-\int_0^t r_1(t')\,dt'\right\}\exp\left\{-\int_0^t r_2(t')\,dt'\right\}\ldots\exp\left\{-\int_0^t r_n(t')\,dt'\right\}$$
$$= \exp\left\{-\int_0^t [r_1(t') + r_2(t') + \cdots + r_n(t')]\,dt'\right\}.$$

Example 4.49

Suppose that a system consists of n components in series and that the component lifetimes are exponential random variables with rates $\lambda_1, \lambda_2, \ldots, \lambda_n$. Find the system reliability.

From Eqs. (4.98) and (4.103), we have

$$R(t) = e^{-\lambda_1 t} e^{-\lambda_2 t} \dots e^{-\lambda_n t}$$

$$= e^{-(\lambda_1 + \dots + \lambda_n)t}.$$

Thus the system reliability is exponentially distributed with rate $\lambda_1 + \lambda_2 + \dots + \lambda_n$.

Now suppose that a system consists of n components in parallel, as shown in Fig. 4.18(b). This system is considered to be functioning as long as at least one of the components is functioning. The system will *not* be functioning if and only if all the components have failed, that is,

$$P[A_s^c] = P[A_1^c]P[A_2^c]\dots P[A_n^c].$$

Thus

$$1 - R(t) = (1 - R_1(t))(1 - R_2(t))\dots(1 - R_n(t)),$$

and finally,

$$R(t) = 1 - (1 - R_1(t))(1 - R_2(t))\dots(1 - R_n(t)). \tag{4.104}$$

Example 4.50

Compare the reliability of a single-unit system against that of a system that operates two units in parallel. Assume all units have exponentially distributed lifetimes with rate 1.

The reliability of the single-unit system is

$$R_s(t) = e^{-t}.$$

The reliability of the two-unit system is

$$R_p(t) = 1 - (1 - e^{-t})(1 - e^{-t})$$

$$= e^{-t}(2 - e^{-t}).$$

The parallel system is more reliable by a factor of

$$(2 - e^{-t}) > 1.$$

More complex configurations can be obtained by combining subsystems consisting of series and parallel components. The reliability of such systems can then be computed in terms of the subsystem reliabilities. See Example 2.35 for an example of such a calculation.

4.9   COMPUTER METHODS FOR GENERATING RANDOM VARIABLES

The computer simulation of any random phenomenon involves the generation of random variables with prescribed distributions. For example, the simulation of a queueing system involves generating the time between customer arrivals as well as the service times of each customer. Once the cdf's that model these random quantities have been selected, an algorithm for generating random variables with these cdf's must be found. MATLAB and Octave have built-in functions for generating random variables for all

of the well known distributions. In this section we present the methods that are used for generating random variables. All of these methods are based on the availability of random numbers that are uniformly distributed between zero and one. Methods for generating these numbers were discussed in Section 2.7.

All of the methods for generating random variables require the evaluation of either the pdf, the cdf, or the inverse of the cdf of the random variable of interest. We can write programs to perform these evaluations, or we can use the functions available in programs such as MATLAB and Octave. The following example shows some typical evaluations for the Gaussian random variable.

Example 4.51   Evaluation of pdf, cdf, and Inverse cdf

Let X be a Gaussian random variable with mean 1 and variance 2. Find the pdf at $x = 7$. Find the cdf at $x = -2$. Find the value of x at which the cdf $= 0.25$.

The following commands show how these results are obtained using Octave.

```
> normal_pdf (7, 1, 2)
ans = 3.4813e-05
> normal_cdf (−2, 1, 2)
ans = 0.016947
> normal_inv (0.25, 1, 2)
ans = 0.046127
```

4.9.1   The Transformation Method

Suppose that U is uniformly distributed in the interval $[0, 1]$. Let $F_X(x)$ be the cdf of the random variable we are interested in generating. Define the random variable, $Z = F_X^{-1}(U)$; that is, first U is selected and then Z is found as indicated in Fig. 4.19. The cdf of Z is

$$P[Z \leq x] = P[F_X^{-1}(U) \leq x] = P[U \leq F_X(x)].$$

But if U is uniformly distributed in $[0, 1]$ and $0 \leq h \leq 1$, then $P[U \leq h] = h$ (see Example 4.6). Thus

$$P[Z \leq x] = F_X(x),$$

and $Z = F_X^{-1}(U)$ has the desired cdf.

Transformation Method for Generating X:

1. Generate U uniformly distributed in $[0, 1]$.
2. Let $Z = F_X^{-1}(U)$.

Example 4.52   Exponential Random Variable

To generate an exponentially distributed random variable X with parameter λ, we need to invert the expression $u = F_X(x) = 1 - e^{-\lambda x}$. We obtain

$$X = -\frac{1}{\lambda}\ln(1 - U).$$
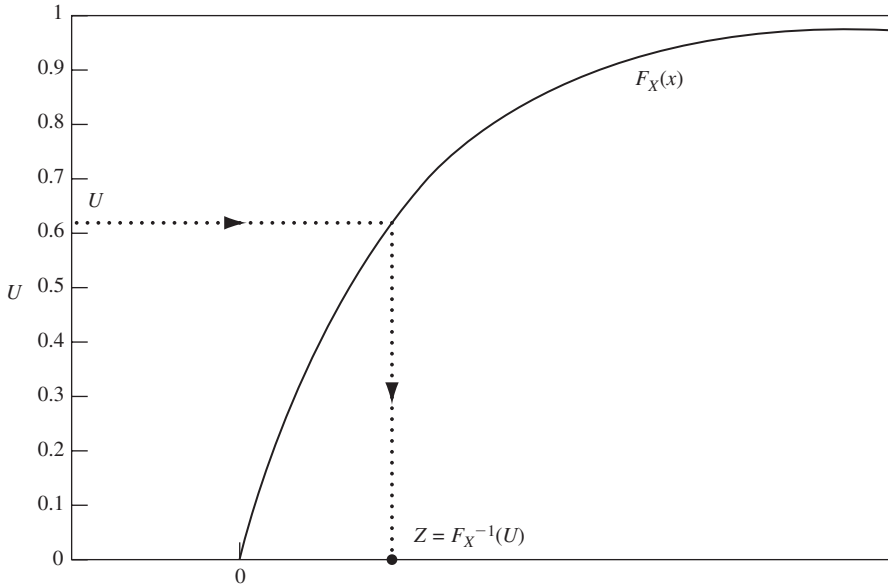
FIGURE 4.19
Transformation method for generating a random variable with cdf $F_X(x)$.

Note that we can use the simpler expression $X = -\ln(U)/\lambda$, since $1 - U$ is also uniformly distributed in $[0, 1]$. The first two lines of the Octave commands below show how to implement the transformation method to generate 1000 exponential random variables with $\lambda = 1$. Figure 4.20 shows the histogram of values obtained. In addition, the figure shows the probability that samples of the random variables fall in the corresponding histogram bins. Good correspondence between the histograms and these probabilities are observed. In Chapter 8 we introduce methods for assessing the goodness-of-fit of data to a given distribution. Both MATLAB and Octave use the transformation method in their function `exponential_rnd`.

```
> U=rand (1, 1000);          % Generate 1000 uniform random variables.
> X=-log(U);                 % Compute 1000 exponential RVs.
> K=0.25:0.5:6;
> P(1)=1-exp(-0.5)
> for i=2:12,                % The remaining lines show how to generate
> P(i)=P(i-1)*exp(-0.5)      % the histogram bins.
> end;
> stem (K, P)
> hold on
> Hist (X, K, 1)
```

4.9.2   The Rejection Method

We first consider the simple version of this algorithm and explain why it works; then we present it in its general form. Suppose that we are interested in generating a random variable Z with pdf $f_X(x)$ as shown in Fig. 4.21. In particular, we assume that: (1) the pdf is nonzero only in the interval $[0, a]$, and (2) the pdf takes on values in the range $[0, b]$. The **rejection method** in this case works as follows:
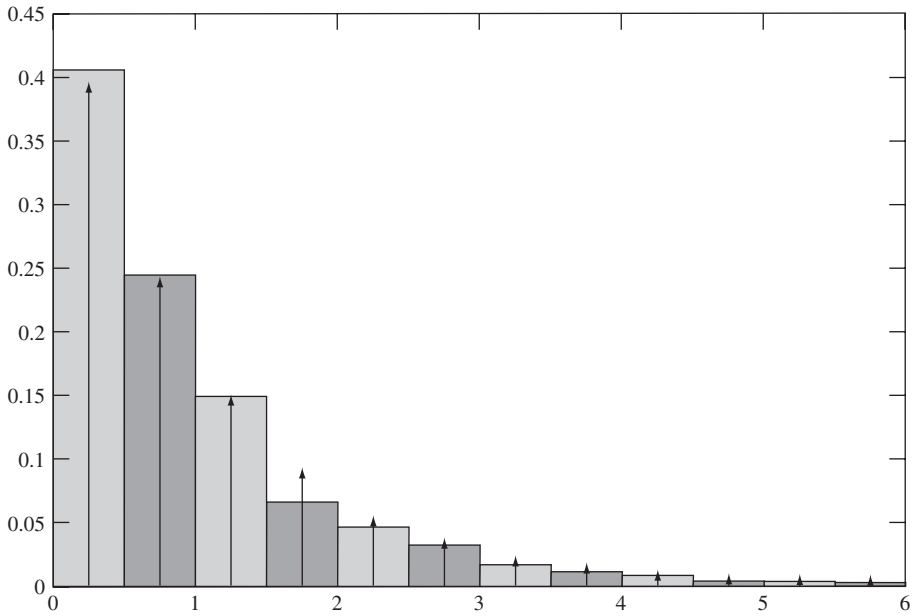
FIGURE 4.20
Histogram of 1000 exponential random variables using transformation method.
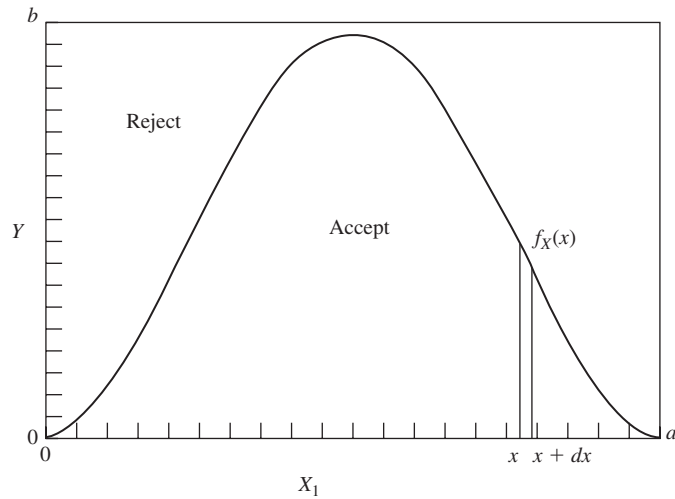
FIGURE 4.21
Rejection method for generating a random variable with pdf $f_X(x)$.

1. Generate X_1 uniform in the interval $[0, a]$.
2. Generate Y uniform in the interval $[0, b]$.
3. If $Y \leq f_X(X_1)$, then output $Z = X_1$; else, reject X_1 and return to step 1.

Note that this algorithm will perform a random number of steps before it produces the output Z.

    We now show that the output Z has the desired pdf. Steps 1 and 2 select a point at random in a rectangle of width a and height b. The probability of selecting a point in any region is simply the area of the region divided by the total area of the rectangle, ab. Thus the probability of accepting X_1 is the probability of the region below $f_X(x)$ divided by ab. But the area under any pdf is 1, so we conclude that the probability of success (i.e., acceptance) is $1/ab$. Consider now the following probability:

$$P[x < X_1 \leq x + dx | X_1 \text{ is accepted}]$$

$$= \frac{P[\{x < X_1 \leq x + dx\} \cap \{X_1 \text{ accepted}\}]}{P[X_1 \text{ accepted}]}$$

$$= \frac{\text{shaded area}/ab}{1/ab} = \frac{f_X(x)\, dx/ab}{1/ab}$$

$$= f_X(x)\, dx.$$

Therefore X_1 when accepted has the desired pdf. Thus Z has the desired pdf.

Example 4.53   Generating Beta Random Variables

Show that the beta random variables with $a' = b' = 2$ can be generated using the rejection method.

    The pdf of the beta random variable with $a' = b' = 2$ is similar to that shown in Fig. 4.21. This beta pdf is maximum at $x = 1/2$ and the maximum value is:

$$\frac{(1/2)^{2-1}(1/2)^{2-1}}{B(2,2)} = \frac{1/4}{\Gamma(2)\Gamma(2)/\Gamma(4)} = \frac{1/4}{1!1!/3!} = \frac{3}{2}.$$

Therefore we can generate this beta random variable using the rejection method with $b = 1.5$.

    The algorithm as stated above can have two problems. First, if the rectangle does not fit snugly around $f_X(x)$, the number of X_1's that need to be generated before acceptance may be excessive. Second, the above method cannot be used if $f_X(x)$ is unbounded or if its range is not finite. The general version of this algorithm overcomes both problems. Suppose we want to generate Z with pdf $f_X(x)$. Let W be a random variable with pdf $f_W(x)$ that is *easy* to generate and such that for some constant $K > 1$,

$$Kf_W(x) \geq f_X(x) \qquad \text{for all } x,$$

that is, the region under $Kf_W(x)$ contains $f_X(x)$ as shown in Fig. 4.22.

Rejection Method for Generating X:

1. Generate X_1 with pdf $f_W(x)$. Define $B(X_1) = Kf_W(X_1)$.
2. Generate Y uniform in $[0, B(X_1)]$.
3. If $Y \leq f_X(X_1)$, then output $Z = X_1$; else reject X_1 and return to step 1.

    See Problem 4.143 for a proof that Z has the desired pdf.
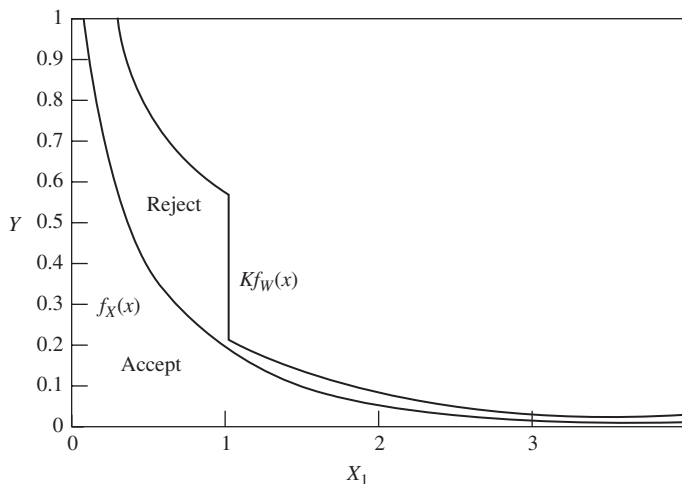
FIGURE 4.22
Rejection method for generating a random variable with gamma pdf and with
$0 < \alpha < 1$.

Example 4.54   Gamma Random Variable

We now show how the rejection method can be used to generate X with gamma pdf and parameters
$0 < \alpha < 1$ and $\lambda = 1$. A function $Kf_W(x)$ that "covers" $f_X(x)$ is easily obtained (see Fig. 4.22):

$$f_X(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} \leq Kf_W(x) = \begin{cases} \dfrac{x^{\alpha-1}}{\Gamma(\alpha)} & 0 \leq x \leq 1 \\[2mm] \dfrac{e^{-x}}{\Gamma(\alpha)} & x > 1. \end{cases}$$

The pdf $f_W(x)$ that corresponds to the function on the right-hand side is

$$f_W(x) = \begin{cases} \dfrac{\alpha e x^{\alpha-1}}{\alpha + e} & 0 \leq x \leq 1 \\[2mm] \alpha e \dfrac{e^{-x}}{\alpha + e} & x \geq 1. \end{cases}$$

The cdf of W is

$$F_W(x) = \begin{cases} \dfrac{e x^{\alpha}}{\alpha + e} & 0 \leq x \leq 1 \\[2mm] 1 - \alpha e \dfrac{e^{-x}}{\alpha + e} & x > 1. \end{cases}$$

W is easy to generate using the transformation method, with

$$F_W^{-1}(u) = \begin{cases} \left[\dfrac{(\alpha + e)u}{e} \right]^{1/\alpha} & u \leq e/(\alpha + e) \\[3mm] -\ln\left[(\alpha + e)\dfrac{(1 - u)}{\alpha e} \right] & u > e/(\alpha + e). \end{cases}$$

We can therefore use the transformation method to generate this $f_W(x)$, and then the rejection method to generate any gamma random variable X with parameters $0 < \alpha < 1$ and $\lambda = 1$. Finally we note that if we let $W = \lambda X$, then W will be gamma with parameters α and λ. The generation of gamma random variables with $\alpha > 1$ is discussed in Problem 4.142.

Example 4.55   Implementing Rejection Method for Gamma Random Variables

Given below is an Octave function definition to implement the rejection method using the above transformation.

% Generate random numbers from the gamma distribution for $0 \le \alpha \le 1$.
```
function X = gamma_rejection_method_altone(alpha)
while (true),

    X = special_inverse(alpha);           % Step 1: Generate X with pdf fx(x).
    B = special_pdf (X, alpha);           % Step 2: Generate Y uniform in [0, Kfx(X)].
    Y = rand.* B;
    if (Y <= fx_gamma_pdf (X, alpha)),    % Step 3: Accept or reject...
         break;
    end

end
```

% Helper function to generate random variables according to $Kf_Z(x)$.
```
function X = special_inverse (alpha)
u = rand;
if (u <= e./(alpha+e)),
    X = ((alpha+e).*u./e). ^ (1./alpha);
elseif (u > e./(alpha+e)),
    X = -log((alpha+e).*(1-u)./(alpha.*e));
end
```

% Return B in order to generate uniform variables in $[0, Kf_Z(X)]$.
```
function B = special_pdf (X, alpha)
if (X >=0 && X <= 1),
    B = alpha.*e.*X.^(alpha-1)./(alpha + e);
elseif (X > 1),
    B = alpha.*e.*(e. ^(-X)./(alpha + e));
end
```

% pdf of the gamma distribution.
% Could also use the built in `gamma_pdf (X, A, B)` function supplied with Octave
```
setting B = 1
function Y = fx_gamma_pdf (x, alpha)
y = (x.^ (alpha-1)).*(e.^ (-x))./(gamma(alpha));
```

Figure 4.23 shows the histogram of 1000 samples obtained using this function. The figure also shows the probability that the samples fall in the bins of the histogram.

We have presented the most common methods that are used to generate random variables. These methods are incorporated in the functions provided by programs such as MATLAB and Octave, so in practice you do not need to write programs to
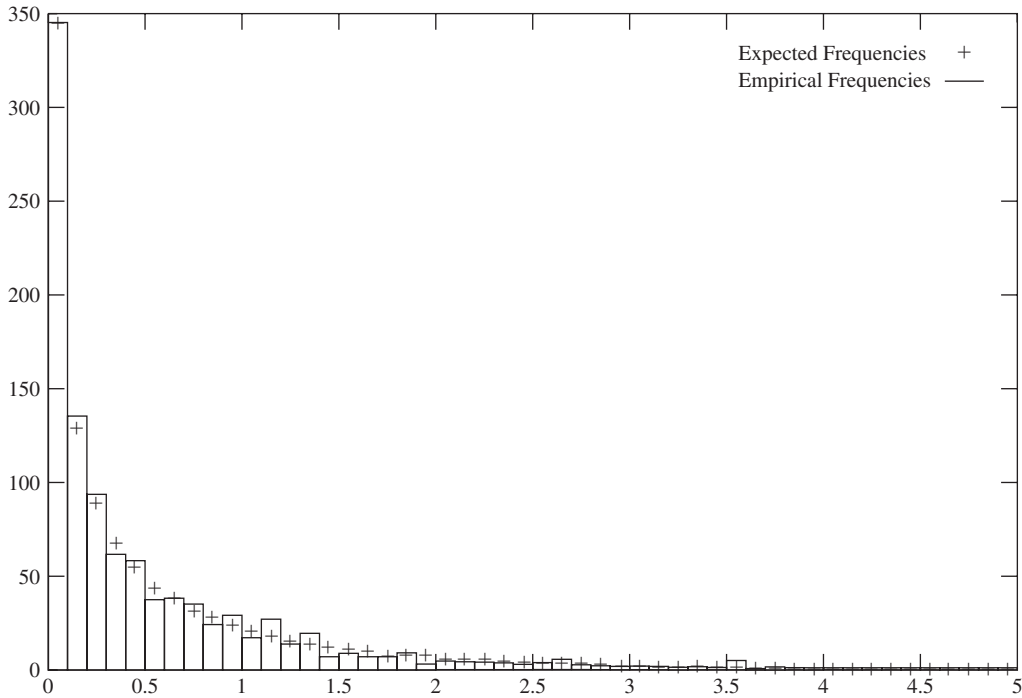
FIGURE 4.23
1000 samples of gamma random variable using rejection method.

generate the most common random variables. You simply need to invoke the appropriate functions.

Example 4.56   Generating Gamma Random Variables

Use Octave to obtain eight Gamma random variables with $\alpha = 0.25$ and $\lambda = 1$.

The Octave command and the corresponding answer are given below:

```
> gamma_rnd (0.25, 1, 1, 8)
ans =
 Columns 1 through 6:
   0.00021529   0.09331491   0.24606757   0.08665787
   0.00013400   0.23384718
 Columns 7 and 8:
   1.72940941   1.29599702
```

4.9.3   Generation of Functions of a Random Variable

Once we have a simple method of generating a random variable X, we can easily generate any random variable that is defined by $Y = g(X)$ or even $Z = h(X_1, X_2, \ldots, X_n)$, where $X_1, \ldots, X_n$ are n outputs of the random variable generator.

Example 4.57 _m_-Erlang Random Variable

Let $X_1, X_2, \ldots$ be independent, exponentially distributed random variables with parameter λ. In Chapter 7 we show that the random variable

$$Y = X_1 + X_2 + \cdots + X_m$$

has an m-Erlang pdf with parameter λ. We can therefore generate an m-Erlang random variable by first generating m exponentially distributed random variables using the transformation method, and then taking the sum. Since the m-Erlang random variable is a special case of the gamma random variable, for large m it may be preferable to use the rejection method described in Problem 4.142.

4.9.4 Generating Mixtures of Random Variables

We have seen in previous sections that sometimes a random variable consists of a mixture of several random variables. In other words, the generation of the random variable can be viewed as first selecting a random variable type according to some pmf, and then generating a random variable from the selected pdf type. This procedure can be simulated easily.

Example 4.58 Hyperexponential Random Variable

A two-stage hyperexponential random variable has pdf

$$f_X(x) = pae^{-ax} + (1 - p)be^{-bx}.$$

It is clear from the above expression that X consists of a mixture of two exponential random variables with parameters a and b, respectively. X can be generated by first performing a Bernoulli trial with probability of success p. If the outcome is a success, we then use the transformation method to generate an exponential random variable with parameter a. If the outcome is a failure, we generate an exponential random variable with parameter b instead.

---

*4.10 ENTROPY

Entropy is a measure of the uncertainty in a random experiment. In this section, we first introduce the notion of the entropy of a random variable and develop several of its fundamental properties. We then show that entropy quantifies uncertainty by the amount of information required to specify the outcome of a random experiment. Finally, we discuss the method of maximum entropy, which has found wide use in characterizing random variables when only some parameters, such as the mean or variance, are known.

4.10.1 The Entropy of a Random Variable

Let X be a discrete random variable with $S_X = \{1, 2, \ldots, K\}$ and pmf $p_k = P[X = k]$. We are interested in quantifying the uncertainty of the event $A_k = \{X = k\}$. Clearly, the uncertainty of A_k is low if the probability of A_k is close to one, and it is high if the

SUMMARY

- The cumulative distribution function $F_X(x)$ is the probability that X falls in the interval $(-\infty, x]$. The probability of any event consisting of the union of intervals can be expressed in terms of the cdf.

- A random variable is continuous if its cdf can be written as the integral of a non-negative function. A random variable is mixed if it is a mixture of a discrete and a continuous random variable.

- The probability of events involving a continuous random variable X can be expressed as integrals of the probability density function $f_X(x)$.

- If X is a random variable, then $Y = g(X)$ is also a random variable. The notion of equivalent events allows us to derive expressions for the cdf and pdf of Y in terms of the cdf and pdf of X.

- The cdf and pdf of the random variable X are sufficient to compute all probabilities involving X alone. The mean, variance, and moments of a random variable summarize some of the information about the random variable X. These parameters are useful in practice because they are easier to measure and estimate than the cdf and pdf.

- Conditional cdf's or pdf's incorporate partial knowledge about the outcome of an experiment in the calculation of probabilities of events.

- The Markov and Chebyshev inequalities allow us to bound probabilities involving X in terms of its first two moments only.

- Transforms provide an alternative but equivalent representation of the pmf and pdf. In certain types of problems it is preferable to work with the transforms rather than the pmf or pdf. The moments of a random variable can be obtained from the corresponding transform.

- The reliability of a system is the probability that it is still functioning after t hours of operation. The reliability of a system can be determined from the reliability of its subsystems.

- There are a number of methods for generating random variables with prescribed pmf's or pdf's in terms of a random variable that is uniformly distributed in the unit interval. These methods include the transformation and the rejection methods as well as methods that simulate random experiments (e.g., functions of random variables) and mixtures of random variables.

- The entropy of a random variable X is a measure of the uncertainty of X in terms of the average amount of information required to identify its value.

- The maximum entropy method is a procedure for estimating the pmf or pdf of a random variable when only partial information about X, in the form of expected values of functions of X, is available.

CHECKLIST OF IMPORTANT TERMS

Characteristic function
Chebyshev inequality
Chernoff bound
Conditional cdf, pdf
Continuous random variable
Cumulative distribution function
Differential entropy
Discrete random variable
Entropy
Equivalent event
Expected value of X
Failure rate function
Function of a random variable
Laplace transform of the pdf
Markov inequality

Maximum entropy method
Mean time to failure (MTTF)
Moment theorem
nth moment of X
Probability density function
Probability generating function
Probability mass function
Random variable
Random variable of mixed type
Rejection method
Reliability
Standard deviation of X
Transformation method
Variance of X

ANNOTATED REFERENCES

Reference [1] is the standard reference for electrical engineers for the material on random variables. Reference [2] is entirely devoted to continuous distributions. Reference [3] discusses some of the finer points regarding the concept of a random variable at a level accessible to students of this course. Reference [4] presents detailed discussions of the various methods for generating random numbers with specified distributions. Reference [5] also discusses the generation of random variables. Reference [9] is focused on signal processing. Reference [11] discusses entropy in the context of information theory.

1. A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 2002.
2. N. Johnson et al., *Continuous Univariate Distributions,* vol. 2, Wiley, New York, 1995.
3. K. L. Chung, *Elementary Probability Theory*, Springer-Verlag, New York, 1974.
4. A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, McGraw-Hill, New York, 2000.
5. S. M. Ross, *Introduction to Probability Models*, Academic Press, New York, 2003.
6. H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N.J., 1946.
7. M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, D.C., 1964. Downloadable: *www.math.sfu.ca/~cbm/aands/*.
8. R. C. Cheng, "The Generation of Gamma Variables with Nonintegral Shape Parameter," *Appl. Statist.*, 26: 71–75, 1977.
9. R. Gray and L.D. Davisson, *An Introduction to Statistical Signal Processing,* Cambridge Univ. Press, Cambridge, UK, 2005.

10. P. O. Börjesson and C. E. W. Sundberg, "Simple Approximations of the Error Function $Q(x)$ for Communications Applications," *IEEE Trans. on Communications*, March 1979, 639–643.

11. R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.

PROBLEMS

Section 4.1: The Cumulative Distribution Function

4.1. An information source produces binary pairs that we designate as $S_X = \{1, 2, 3, 4\}$ with the following pmf's:

    (i) $p_k = p_1/k$ for all k in S_X.

    (ii) $p_{k+1} = p_k/2$ for $k = 2, 3, 4$.

    (iii) $p_{k+1} = p_k/2^k$ for $k = 2, 3, 4$.

  (a) Plot the cdf of these three random variables.

  (b) Use the cdf to find the probability of the events: $\{X \leq 1\}$, $\{X < 2.5\}$, $\{0.5 < X \leq 2\}$, $\{1 < X < 4\}$.

4.2. A die is tossed. Let X be the number of full pairs of dots in the face showing up, and Y be the number of full or partial pairs of dots in the face showing up. Find and plot the cdf of X and Y.

4.3. The loose minute hand of a clock is spun hard. The coordinates (x, y) of the point where the tip of the hand comes to rest is noted. Z is defined as the sgn function of the product of x and y, where sgn(t) is 1 if $t > 0$, 0 if $t = 0$, and -1 if $t < 0$.

  (a) Find and plot the cdf of the random variable X.

  (b) Does the cdf change if the clock hand has a propensity to stop at 3, 6, 9, and 12 o'clock?

4.4. An urn contains 8 $1 bills and two $5 bills. Let X be the total amount that results when two bills are drawn from the urn without replacement, and let Y be the total amount that results when two bills are drawn from the urn *with* replacement.

  (a) Plot and compare the cdf's of the random variables.

  (b) Use the cdf to compare the probabilities of the following events in the two problems: $\{X = \$2\}$, $\{X < \$7\}$, $\{X \geq 6\}$.

4.5. Let Y be the difference between the number of heads and the number of tails in the 3 tosses of a fair coin.

  (a) Plot the cdf of the random variable Y.

  (b) Express $P[|Y| < y]$ in terms of the cdf of Y.

4.6. A dart is equally likely to land at any point inside a circular target of radius 2. Let R be the distance of the landing point from the origin.

  (a) Find the sample space S and the sample space of R, S_R.

  (b) Show the mapping from S to S_R.

  (c) The "bull's eye" is the central disk in the target of radius 0.25. Find the event A in S_R corresponding to "dart hits the bull's eye." Find the equivalent event in S and $P[A]$.

  (d) Find and plot the cdf of R.

4.7. A point is selected at random inside a square defined by $\{(x, y): 0 \leq x \leq b, 0 \leq y \leq b\}$. Assume the point is equally likely to fall anywhere in the square. Let the random variable Z be given by the minimum of the two coordinates of the point where the dart lands.

  (a) Find the sample space S and the sample space of Z, S_Z.

(b) Show the mapping from S to S_Z.

(c) Find the region in the square corresponding to the event $\{Z \leq z\}$.

(d) Find and plot the cdf of Z.

(e) Use the cdf to find: $P[Z > 0]$, $P[Z > b]$, $P[Z \leq b/2]$, $P[Z > b/4]$.

4.8. Let ζ be a point selected at random from the unit interval. Consider the random variable $X = (1 - \zeta)^{-1/2}$.

(a) Sketch X as a function of ζ.

(b) Find and plot the cdf of X.

(c) Find the probability of the events $\{X > 1\}$, $\{5 < X < 7\}$, $\{X \leq 20\}$.

4.9. The loose hand of a clock is spun hard and the outcome ζ is the angle in the range $[0, 2\pi)$ where the hand comes to rest. Consider the random variable $X(\zeta) = 2 \sin(\zeta/4)$.

(a) Sketch X as a function of ζ.

(b) Find and plot the cdf of X.

(c) Find the probability of the events $\{X > 1\}$, $\{-1/2 < X < 1/2\}$, $\{X \leq 1/\sqrt{2}\}$.

4.10. Repeat Problem 4.9 if 80% of the time the hand comes to rest anywhere in the circle, but 20% of the time the hand comes to rest at 3, 6, 9, or 12 o'clock.

4.11. The random variable X is uniformly distributed in the interval $[-1, 2]$.

(a) Find and plot the cdf of X.

(b) Use the cdf to find the probabilities of the following events: $\{X \leq 0\}$, $\{|X - 0.5| < 1\}$, and $C = \{X > -0.5\}$.

4.12. The cdf of the random variable X is given by:

$$F_X(x) = \begin{cases} 0 & x < -1 \\ 0.5 & -1 \leq x \leq 0 \\ (1 + x)/2 & 0 \leq x \leq 1 \\ 1 & x \geq 1. \end{cases}$$

(a) Plot the cdf and identify the type of random variable.

(b) Find $P[X \leq -1]$, $P[X = -1]$, $P[X < 0.5]$, $P[-0.5 < X < 0.5]$, $P[X > -1]$, $P[X \leq 2]$, $P[X > 3]$.

4.13. A random variable X has cdf:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - \dfrac{1}{4}e^{-2x} & \text{for } x \geq 0. \end{cases}$$

(a) Plot the cdf and identify the type of random variable.

(b) Find $P[X \leq 2]$, $P[X = 0]$, $P[X < 0]$, $P[2 < X < 6]$, $P[X > 10]$.

4.14. The random variable X has cdf shown in Fig. P4.1.

(a) What type of random variable is X?

(b) Find the following probabilities: $P[X < -1]$, $P[X \leq -1]$, $P[-1 < X < -0.75]$, $P[-0.5 \leq X < 0]$, $P[-0.5 \leq X \leq 0.5]$, $P[|X - 0.5| < 0.5]$.

4.15. For $\beta > 0$ and $\lambda > 0$, the Weibull random variable Y has cdf:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-(x/\lambda)^\beta} & \text{for } x \geq 0. \end{cases}$$
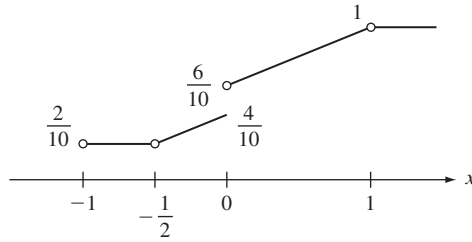
FIGURE P4.1

    (a) Plot the cdf of Y for $\beta = 0.5, 1$, and 2.

    (b) Find the probability $P[j\lambda < X < (j + 1)\lambda]$ and $P[X > j\lambda]$.

    (c) Plot $\log P[X > x]$ vs. $\log x$.

4.16. The random variable X has cdf:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.5 + c \sin^2(\pi x/2) & 0 \le x \le 1 \\ 1 & x > 1. \end{cases}$$

    (a) What values can c assume?

    (b) Plot the cdf.

    (c) Find $P[X > 0]$.

Section 4.2: The Probability Density Function

4.17. A random variable X has pdf:

$$f_X(x) = \begin{cases} c(1 - x^2) & -1 \le x \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

    (a) Find c and plot the pdf.

    (b) Plot the cdf of X.

    (c) Find $P[X = 0]$, $P[0 < X < 0.5]$, and $P[|X - 0.5| < 0.25]$.

4.18. A random variable X has pdf:

$$f_X(x) = \begin{cases} cx(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{elsewhere.} \end{cases}$$

    (a) Find c and plot the pdf.

    (b) Plot the cdf of X.

    (c) Find $P[0 < X < 0.5]$, $P[X = 1]$, $P[.25 < X < 0.5]$.

4.19. (a) In Problem 4.6, find and plot the pdf of the random variable R, the distance from the dart to the center of the target.

    (b) Use the pdf to find the probability that the dart is outside the bull's eye.

4.20. (a) Find and plot the pdf of the random variable Z in Problem 4.7.

    (b) Use the pdf to find the probability that the minimum is greater than $b/3$.

4.21. (a) Find and plot the pdf in Problem 4.8.

   (b) Use the pdf to find the probabilities of the events: $\{X > a\}$ and $\{X > 2a\}$.

4.22. (a) Find and plot the pdf in Problem 4.12.

   (b) Use the pdf to find $P[-1 \le X < 0.25]$.

4.23. (a) Find and plot the pdf in Problem 4.13.

   (b) Use the pdf to find $P[X = 0], P[X > 8]$.

4.24. (a) Find and plot the pdf of the random variable in Problem 4.14.

   (b) Use the pdf to calculate the probabilities in Problem 4.14b.

4.25. Find and plot the pdf of the Weibull random variable in Problem 4.15a.

4.26. Find the cdf of the Cauchy random variable which has pdf:

$$f_X(x) = \frac{\alpha/\pi}{x^2 + \alpha^2} \qquad -\infty < x < \infty.$$

4.27. A voltage X is uniformly distributed in the set $\{-3, -2, \ldots, 3, 4\}$.

   (a) Find the pdf and cdf of the random variable X.

   (b) Find the pdf and cdf of the random variable $Y = -2X^2 + 3$.

   (c) Find the pdf and cdf of the random variable $W = \cos(\pi X/8)$.

   (d) Find the pdf and cdf of the random variable $Z = \cos^2(\pi X/8)$.

4.28. Find the pdf and cdf of the Zipf random variable in Problem 3.70.

4.29. Let C be an event for which $P[C] > 0$. Show that $F_X(x|C)$ satisfies the eight properties of a cdf.

4.30. (a) In Problem 4.13, find $F_X(x|C)$ where $C = \{X > 0\}$.

   (b) Find $F_X(x|C)$ where $C = \{X = 0\}$.

4.31. (a) In Problem 4.10, find $F_X(x|B)$ where $B = \{$hand does not stop at 3, 6, 9, or 12 o'clock$\}$.

   (b) Find $F_X(x|B^c)$.

4.32. In Problem 4.13, find $f_X(x|B)$ and $F_X(x|B)$ where $B = \{X > 0.25\}$.

4.33. Let X be the exponential random variable.

   (a) Find and plot $F_X(x|X > t)$. How does $F_X(x|X > t)$ differ from $F_X(x)$?

   (b) Find and plot $f_X(x|X > t)$.

   (c) Show that $P[X > t + x|X > t] = P[X > x]$. Explain why this is called the memoryless property.

4.34. The Pareto random variable X has cdf:

$$F_X(x) = \begin{cases} 0 & x < x_m \\ 1 - \dfrac{x_m^\alpha}{x^\alpha} & x \ge x_m. \end{cases}$$

   (a) Find and plot the pdf of X.

   (b) Repeat Problem 4.33 parts a and b for the Pareto random variable.

   (c) What happens to $P[X > t + x|X > t]$ as t becomes large? Interpret this result.

4.35. (a) Find and plot $F_X(x|a \le X \le b)$. Compare $F_X(x|a \le X \le b)$ to $F_X(x)$.

   (b) Find and plot $f_X(x|a \le X \le b)$.

4.36. In Problem 4.6, find $F_R(r|R > 1)$ and $f_R(r|R > 1)$.

4.37. (a) In Problem 4.7, find $F_Z(z \mid b/4 \leq Z \leq b/2)$ and $f_Z(z \mid b/4 \leq Z \leq b/2)$.

(b) Find $F_Z(z \mid B)$ and $f_Z(z \mid B)$, where $B = \{x > b/2\}$.

4.38. A binary transmission system sends a "0" bit using a -1 voltage signal and a "1" bit by transmitting a $+1$. The received signal is corrupted by noise N that has a Laplacian distribution with parameter α. Assume that "0" bits and "1" bits are equiprobable.

(a) Find the pdf of the received signal $Y = X + N$, where X is the transmitted signal, given that a "0" was transmitted; that a "1" was transmitted.

(b) Suppose that the receiver decides a "0" was sent if $Y < 0$, and a "1" was sent if $Y \geq 0$. What is the probability that the receiver makes an error given that a $+1$ was transmitted? a -1 was transmitted?

(c) What is the overall probability of error?

Section 4.3: The Expected Value of X

4.39. Find the mean and variance of X in Problem 4.17.

4.40. Find the mean and variance of X in Problem 4.18.

4.41. Find the mean and variance of Y, the distance from the dart to the origin, in Problem 4.19.

4.42. Find the mean and variance of Z, the minimum of the coordinates in a square, in Problem 4.20.

4.43. Find the mean and variance of $X = (1 - \zeta)^{-1/2}$ in Problem 4.21. Find $E[X]$ using Eq. (4.28).

4.44. Find the mean and variance of X in Problems 4.12 and 4.22.

4.45. Find the mean and variance of X in Problems 4.13 and 4.23. Find $E[X]$ using Eq. (4.28).

4.46. Find the mean and variance of the Gaussian random variable by direct integration of Eqs. (4.27) and (4.34).

4.47. Prove Eqs. (4.28) and (4.29).

4.48. Find the variance of the exponential random variable.

4.49. (a) Show that the mean of the Weibull random variable in Problem 4.15 is $\Gamma(1 + 1/\beta)$ where $\Gamma(x)$ is the gamma function defined in Eq. (4.56).

(b) Find the second moment and the variance of the Weibull random variable.

4.50. Explain why the mean of the Cauchy random variable does not exist.

4.51. Show that $E[X]$ does not exist for the Pareto random variable with $\alpha = 1$ and $x_m = 1$.

4.52. Verify Eqs. (4.36), (4.37), and (4.38).

4.53. Let $Y = A \cos(\omega t) + c$ where A has mean m and variance σ^2 and ω and c are constants. Find the mean and variance of Y. Compare the results to those obtained in Example 4.15.

4.54. A limiter is shown in Fig. P4.2.
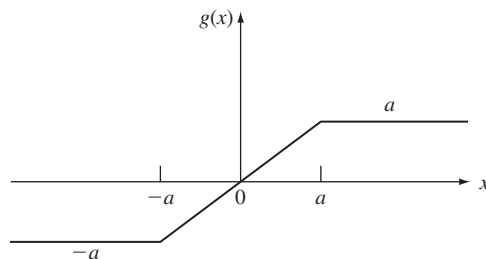
FIGURE P4.2

    (a) Find an expression for the mean and variance of $Y = g(X)$ for an arbitrary continuous random variable X.

    (b) Evaluate the mean and variance if X is a Laplacian random variable with $\lambda = a = 1$.

    (c) Repeat part (b) if X is from Problem 4.17 with $a = 1/2$.

    (d) Evaluate the mean and variance if $X = U^3$ where U is a uniform random variable in the unit interval, $[-1, 1]$ and $a = 1/2$.

4.55. A limiter with center-level clipping is shown in Fig. P4.3.

    (a) Find an expression for the mean and variance of $Y = g(X)$ for an arbitrary continuous random variable X.

    (b) Evaluate the mean and variance if X is Laplacian with $\lambda = a = 1$ and $b = 2$.

    (c) Repeat part (b) if X is from Problem 4.22, $a = 1/2, b = 3/2$.

    (d) Evaluate the mean and variance if $X = b\cos(2\pi U)$ where U is a uniform random variable in the unit interval $[-1, 1]$ and $a = 3/4, b = 1/2$.
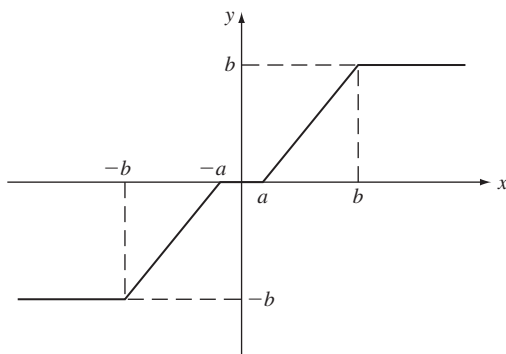
FIGURE P4.3

4.56. Let $Y = 3X + 2$.

    (a) Find the mean and variance of Y in terms of the mean and variance of X.

    (b) Evaluate the mean and variance of Y if X is Laplacian.

    (c) Evaluate the mean and variance of Y if X is an arbitrary Gaussian random variable.

    (d) Evaluate the mean and variance of Y if $X = b\cos(2\pi U)$ where U is a uniform random variable in the unit interval.

4.57. Find the nth moment of U, the uniform random variable in the unit interval. Repeat for X uniform in $[a, b]$.

4.58. Consider the quantizer in Example 4.20.

    (a) Find the conditional pdf of X given that X is in the interval $(d, 2d)$.

    (b) Find the conditional expected value and conditional variance of X given that X is in the interval $(d, 2d)$.

(c) Now suppose that when X falls in $(d, 2d)$, it is mapped onto the point c where $d < c < 2d$. Find an expression for the expected value of the mean square error: $E[(X - c)^2 | d < X < 2d]$.

(d) Find the value c that minimizes the above mean square error. Is c the midpoint of the interval? Explain why or why not by sketching possible conditional pdf shapes.

(e) Find an expression for the overall mean square error using the approach in parts c and d.

Section 4.4: Important Continuous Random Variables

4.59. Let X be a uniform random variable in the interval $[-2, 2]$. Find and plot $P[|X| > x]$.

4.60. In Example 4.20, let the input to the quantizer be a uniform random variable in the interval $[-4d, 4d]$. Show that $Z = X - Q(X)$ is uniformly distributed in $[-d/2, d/2]$.

4.61. Let X be an exponential random variable with parameter λ.

(a) For $d > 0$ and k a nonnegative integer, find $P[kd < X < (k + 1)d]$.

(b) Segment the positive real line into four equiprobable disjoint intervals.

4.62. The rth percentile, $\pi(r)$, of a random variable X is defined by $P[X \leq \pi(r)] = r/100$.

(a) Find the 90%, 95%, and 99% percentiles of the exponential random variable with parameter λ.

(b) Repeat part a for the Gaussian random variable with parameters $m = 0$ and σ^2.

4.63. Let X be a Gaussian random variable with $m = 5$ and $\sigma^2 = 16$.

(a) Find $P[X > 4]$, $P[X \geq 7]$, $P[6.72 < X < 10.16]$, $P[2 < X < 7]$, $P[6 \leq X \leq 8]$.

(b) $P[X < a] = 0.8869$, find a.

(c) $P[X > b] = 0.11131$, find b.

(d) $P[13 < X \leq c] = 0.0123$, find c.

4.64. Show that the Q-function for the Gaussian random variable satisfies $Q(-x) = 1 - Q(x)$.

4.65. Use Octave to generate Tables 4.2 and 4.3.

4.66. Let X be a Gaussian random variable with mean m and variance σ^2.

(a) Find $P[X \leq m]$.

(b) Find $P[|X - m| < k\sigma]$, for $k = 1, 2, 3, 4, 5, 6$.

(c) Find the value of k for which $Q(k) = P[X > m + k\sigma] = 10^{-j}$ for $j = 1, 2, 3, 4, 5, 6$.

4.67. A binary transmission system transmits a signal X (-1 to send a "0" bit; $+1$ to send a "1" bit). The received signal is $Y = X + N$ where noise N has a zero-mean Gaussian distribution with variance σ^2. Assume that "0" bits are three times as likely as "1" bits.

(a) Find the conditional pdf of Y given the input value: $f_Y(y | X = +1)$ and $f_Y(y | X = -1)$.

(b) The receiver decides a "0" was transmitted if the observed value of y satisfies

$$f_Y(y | X = -1)P[X = -1] > f_Y(y | X = +1)P[X = +1]$$

and it decides a "1" was transmitted otherwise. Use the results from part a to show that this decision rule is equivalent to: If $y < T$ decide "0"; if $y \geq T$ decide "1".

(c) What is the probability that the receiver makes an error given that a $+1$ was transmitted? a -1 was transmitted? Assume $\sigma^2 = 1/16$.

(d) What is the overall probability of error?

4.68. Two chips are being considered for use in a certain system. The lifetime of chip 1 is modeled by a Gaussian random variable with mean 20,000 hours and standard deviation 5000 hours. (The probability of negative lifetime is negligible.) The lifetime of chip 2 is also a Gaussian random variable but with mean 22,000 hours and standard deviation 1000 hours. Which chip is preferred if the target lifetime of the system is 20,000 hours? 24,000 hours?

4.69. Passengers arrive at a taxi stand at an airport at a rate of one passenger per minute. The taxi driver will not leave until seven passengers arrive to fill his van. Suppose that passenger interarrival times are exponential random variables, and let X be the time to fill a van. Find the probability that more than 10 minutes elapse until the van is full.

4.70. (a) Show that the gamma random variable has mean:

$$E[X] = \alpha/\lambda.$$

(b) Show that the gamma random variable has second moment, and variance given by:

$$E[X^2] = \alpha(\alpha + 1)/\lambda^2 \text{ and } \text{VAR}[X] = \alpha/\lambda^2.$$

(c) Use parts a and b to obtain the mean and variance of an m-Erlang random variable.

(d) Use parts a and b to obtain the mean and variance of a chi-square random variable.

4.71. The time X to complete a transaction in a system is a gamma random variable with mean 4 and variance 8. Use Octave to plot $P[X > x]$ as a function of x. Note: Octave uses $\beta = 1/2$.

4.72. (a) Plot the pdf of an m-Erlang random variable for $m = 1, 2, 3$ and $\lambda = 1$.

(b) Plot the chi-square pdf for $k = 1, 2, 3$.

4.73. A repair person keeps four widgets in stock. What is the probability that the widgets in stock will last 15 days if the repair person needs to replace widgets at an average rate of one widget every three days, where the time between widget failures is an exponential random variable?

4.74. (a) Find the cdf of the m-Erlang random variable by integration of the pdf. *Hint:* Use integration by parts.

(b) Show that the derivative of the cdf given by Eq. (4.58) gives the pdf of an m-Erlang random variable.

4.75. Plot the pdf of a beta random variable with: $a = b = 1/4, 1, 4, 8; a = 5, b = 1; a = 1, b = 3; a = 2, b = 5$.

Section 4.5: Functions of a Random Variable

4.76. Let X be a Gaussian random variable with mean 2 and variance 4. The reward in a system is given by $Y = (X)^+$. Find the pdf of Y.

4.77. The amplitude of a radio signal X is a Rayleigh random variable with pdf:

$$f_X(x) = \frac{x}{\alpha^2} e^{-x^2/2\alpha^2} \qquad x > 0, \quad \alpha > 0.$$

(a) Find the pdf of $Z = (X - r)^+$.

(b) Find the pdf of $Z = X^2$.

4.78. A wire has length X, an exponential random variable with mean 5π cm. The wire is cut to make rings of diameter 1 cm. Find the probability for the number of complete rings produced by each length of wire.

4.79. A signal that has amplitudes with a Gaussian pdf with zero mean and unit variance is applied to the quantizer in Example 4.27.

    (a) Pick d so that the probability that X falls outside the range of the quantizer is 1%.

    (b) Find the probability of the output levels of the quantizer.

4.80. The signal X is amplified and shifted as follows: $Y = 2X + 3$, where X is the random variable in Problem 4.12. Find the cdf and pdf of Y.

4.81. The net profit in a transaction is given by $Y = 2 - 4X$ where X is the random variable in Problem 4.13. Find the cdf and pdf of Y.

4.82. Find the cdf and pdf of the output of the limiter in Problem 4.54 parts b, c, and d.

4.83. Find the cdf and pdf of the output of the limiter with center-level clipping in Problem 4.55 parts b, c, and d.

4.84. Find the cdf and pdf of $Y = 3X + 2$ in Problem 4.56 parts b, c, and d.

4.85. The exam grades in a certain class have a Gaussian pdf with mean m and standard deviation σ. Find the constants a and b so that the random variable $y = aX + b$ has a Gaussian pdf with mean m' and standard deviation σ'.

4.86. Let $X = U^n$ where n is a positive integer and U is a uniform random variable in the unit interval. Find the cdf and pdf of X.

4.87. Repeat Problem 4.86 if U is uniform in the interval $[-1, 1]$.

4.88. Let $Y = |X|$ be the output of a full-wave rectifier with input voltage X.

    (a) Find the cdf of Y by finding the equivalent event of $\{Y \leq y\}$. Find the pdf of Y by differentiation of the cdf.

    (b) Find the pdf of Y by finding the equivalent event of $\{y < Y \leq y + dy\}$. Does the answer agree with part a?

    (c) What is the pdf of Y if the $f_X(x)$ is an even function of x?

4.89. Find and plot the cdf of Y in Example 4.34.

4.90. A voltage X is a Gaussian random variable with mean 1 and variance 2. Find the pdf of the power dissipated by an R-ohm resistor $P = RX^2$.

4.91. Let $Y = e^X$.

    (a) Find the cdf and pdf of Y in terms of the cdf and pdf of X.

    (b) Find the pdf of Y when X is a Gaussian random variable. In this case Y is said to be a lognormal random variable. Plot the pdf and cdf of Y when X is zero-mean with variance 1/8; repeat with variance 8.

4.92. Let a radius be given by the random variable X in Problem 4.18.

    (a) Find the pdf of the area covered by a disc with radius X.

    (b) Find the pdf of the volume of a sphere with radius X.

    (c) Find the pdf of the volume of a sphere in R^n:

$$Y = \begin{cases} (2\pi)^{(n-1)/2} X^n/(2 \times 4 \times \cdots \times n) & \text{for } n \text{ even} \\ 2(2\pi)^{(n-1)/2} X^n/(1 \times 3 \times \cdots \times n) & \text{for } n \text{ odd.} \end{cases}$$

4.93. In the quantizer in Example 4.20, let $Z = X - q(X)$. Find the pdf of Z if X is a Laplacian random variable with parameter $\alpha = d/2$.

4.94. Let $Y = \alpha \tan \pi X$, where X is uniformly distributed in the interval $(-1, 1)$.

    (a) Show that Y is a Cauchy random variable.

    (b) Find the pdf of $Y = 1/X$.

4.95. Let X be a Weibull random variable in Problem 4.15. Let $Y = (X/\lambda)^\beta$. Find the cdf and pdf of Y.

4.96. Find the pdf of $X = -\ln(1 - U)$, where U is a uniform random variable in $(0, 1)$.

Section 4.6: The Markov and Chebyshev Inequalities

4.97. Compare the Markov inequality and the exact probability for the event $\{X > c\}$ as a function of c for:

    (a) X is a uniform random variable in the interval $[0, b]$.

    (b) X is an exponential random variable with parameter λ.

    (c) X is a Pareto random variable with $\alpha > 1$.

    (d) X is a Rayleigh random variable.

4.98. Compare the Markov inequality and the exact probability for the event $\{X > c\}$ as a function of c for:

    (a) X is a uniform random variable in $\{1, 2, \ldots, L\}$.

    (b) X is a geometric random variable.

    (c) X is a Zipf random variable with $L = 10$; $L = 100$.

    (d) X is a binomial random variable with $n = 10$, $p = 0.5$; $n = 50$, $p = 0.5$.

4.99. Compare the Chebyshev inequality and the exact probability for the event $\{|X - m| > c\}$ as a function of c for:

    (a) X is a uniform random variable in the interval $[-b, b]$.

    (b) X is a Laplacian random variable with parameter α.

    (c) X is a zero-mean Gaussian random variable.

    (d) X is a binomial random variable with $n = 10$, $p = 0.5$; $n = 50$, $p = 0.5$.

4.100. Let X be the number of successes in n Bernoulli trials where the probability of success is p. Let $Y = X/n$ be the average number of successes per trial. Apply the Chebyshev inequality to the event $\{|Y - p| > a\}$. What happens as $n \to \infty$?

4.101. Suppose that light bulbs have exponentially distributed lifetimes with unknown mean $E[X]$. Suppose we measure the lifetime of n light bulbs, and we estimate the mean $E[X]$ by the arithmetic average Y of the measurements. Apply the Chebyshev inequality to the event $\{|Y - E[X]| > a\}$. What happens as $n \to \infty$? *Hint:* Use the m-Erlang random variable.

Section 4.7: Transform Methods

4.102. (a) Find the characteristic function of the uniform random variable in $[-b, b]$.

    (b) Find the mean and variance of X by applying the moment theorem.

4.103. (a) Find the characteristic function of the Laplacian random variable.

    (b) Find the mean and variance of X by applying the moment theorem.

4.104. Let $\Phi_X(\omega)$ be the characteristic function of an exponential random variable. What random variable does $\Phi_X^n(\omega)$ correspond to?

4.105. Find the mean and variance of the Gaussian random variable by applying the moment theorem to the characteristic function given in Table 4.1.

4.106. Find the characteristic function of $Y = aX + b$ where X is a Gaussian random variable. *Hint:* Use Eq. (4.79).

4.107. Show that the characteristic function for the Cauchy random variable is $e^{-|\omega|}$.

4.108. Find the Chernoff bound for the exponential random variable with $\lambda = 1$. Compare the bound to the exact value for $P[X > 5]$.

4.109. **(a)** Find the probability generating function of the geometric random variable.

 (b) Find the mean and variance of the geometric random variable from its pgf.

4.110. **(a)** Find the pgf for the binomial random variable X with parameters n and p.

 (b) Find the mean and variance of X from the pgf.

4.111. Let $G_X(z)$ be the pgf for a binomial random variable with parameters n and p, and let $G_Y(z)$ be the pgf for a binomial random variable with parameters m and p. Consider the function $G_X(z) G_Y(z)$. Is this a valid pgf? If so, to what random variable does it correspond?

4.112. Let $G_N(z)$ be the pgf for a Poisson random variable with parameter α, and let $G_M(z)$ be the pgf for a Poisson random variable with parameters β. Consider the function $G_N(z) G_M(z)$. Is this a valid pgf? If so, to what random variable does it correspond?

4.113. Let N be a Poisson random variable with parameter $\alpha = 1$. Compare the Chernoff bound and the exact value for $P[X \geq 5]$.

4.114. **(a)** Find the pgf $G_U(z)$ for the discrete uniform random variable U.

 (b) Find the mean and variance from the pgf.

 (c) Consider $G_U(z)^2$. Does this function correspond to a pgf? If so, find the mean of the corresponding random variable.

4.115. **(a)** Find $P[X = r]$ for the negative binomial random variable from the pgf in Table 3.1.

 (b) Find the mean of X.

4.116. Derive Eq. (4.89).

4.117. Obtain the nth moment of a gamma random variable from the Laplace transform of its pdf.

4.118. Let X be the mixture of two exponential random variables (see Example 4.58). Find the Laplace transform of the pdf of X.

4.119. The Laplace transform of the pdf of a random variable X is given by:

$$X^*(s) = \frac{a}{s+a}\frac{b}{s+b}.$$

Find the pdf of X. *Hint:* Use a partial fraction expansion of $X^*(s)$.

4.120. Find a relationship between the Laplace transform of a gamma random variable pdf with parameters α and λ and the Laplace transform of a gamma random variable with parameters $\alpha - 1$ and λ. What does this imply if X is an m-Erlang random variable?

4.121. **(a)** Find the Chernoff bound for $P[X > t]$ for the gamma random variable.

 (b) Compare the bound to the exact value of $P[X \geq 9]$ for an $m = 3, \lambda = 1$ Erlang random variable.

Section 4.8: Basic Reliability Calculations

4.122. The lifetime T of a device has pdf

$$f_T(t) = \begin{cases} 1/10T_0 & 0 < t < T_0 \\ 0.9\lambda e^{-\lambda(t-T_0)} & t \geq T_0 \\ 0 & t < T_0. \end{cases}$$

- **(a)** Find the reliability and MTTF of the device.
- **(b)** Find the failure rate function.
- **(c)** How many hours of operation can be considered to achieve 99% reliability?

4.123. The lifetime T of a device has pdf

$$f_T(t) = \begin{cases} 1/T_0 & a \leq t \leq a + T_0 \\ 0 & \text{elsewhere.} \end{cases}$$

- **(a)** Find the reliability and MTTF of the device.
- **(b)** Find the failure rate function.
- **(c)** How many hours of operation can be considered to achieve 99% reliability?

4.124. The lifetime T of a device is a Rayleigh random variable.
- **(a)** Find the reliability of the device.
- **(b)** Find the failure rate function. Does $r(t)$ increase with time?
- **(c)** Find the reliability of two devices that are in series.
- **(d)** Find the reliability of two devices that are in parallel.

4.125. The lifetime T of a device is a Weibull random variable.
- **(a)** Plot the failure rates for $\alpha = 1$ and $\beta = 0.5$; for $\alpha = 1$ and $\beta = 2$.
- **(b)** Plot the reliability functions in part a.
- **(c)** Plot the reliability of two devices that are in series.
- **(d)** Plot the reliability of two devices that are in parallel.

4.126. A system starts with m devices, 1 active and $m - 1$ on standby. Each device has an exponential lifetime. When a device fails it is immediately replaced with another device (if one is still available).
- **(a)** Find the reliability of the system.
- **(b)** Find the failure rate function.

4.127. Find the failure rate function of the memory chips discussed in Example 2.28. Plot $\ln(r(t))$ versus αt.

4.128. A device comes from two sources. Devices from source 1 have mean m and exponentially distributed lifetimes. Devices from source 2 have mean m and Pareto-distributed lifetimes with $\alpha > 1$. Assume a fraction p is from source 1 and a fraction $1 - p$ from source 2.
- **(a)** Find the reliability of an arbitrarily selected device.
- **(b)** Find the failure rate function.

4.129. A device has the failure rate function:

$$r(t) = \begin{cases} 1 + 9(1 - t) & 0 \le t < 1 \\ 1 & 1 \le t < 10 \\ 1 + 10(t - 10) & t \ge 10. \end{cases}$$

Find the reliability function and the pdf of the device.

4.130. A system has three identical components and the system is functioning if two or more components are functioning.

    (a) Find the reliability and MTTF of the system if the component lifetimes are exponential random variables with mean 1.

    (b) Find the reliability of the system if one of the components has mean 2.

4.131. Repeat Problem 4.130 if the component lifetimes are Weibull distributed with $\beta = 3$.

4.132. A system consists of two processors and three peripheral units. The system is functioning as long as one processor and two peripherals are functioning.

    (a) Find the system reliability and MTTF if the processor lifetimes are exponential random variables with mean 5 and the peripheral lifetimes are Rayleigh random variables with mean 10.

    (b) Find the system reliability and MTTF if the processor lifetimes are exponential random variables with mean 10 and the peripheral lifetimes are exponential random variables with mean 5.

4.133. An operation is carried out by a subsystem consisting of three units that operate in a series configuration.

    (a) The units have exponentially distributed lifetimes with mean 1. How many subsystems should be operated in parallel to achieve a reliability of 99% in T hours of operation?

    (b) Repeat part a with Rayleigh-distributed lifetimes.

    (c) Repeat part a with Weibull-distributed lifetimes with $\beta = 3$.

Section 4.9: Computer Methods for Generating Random Variables

4.134. Octave provides function calls to evaluate the pdf and cdf of important continuous random variables. For example, the functions `\normal_cdf(x, m, var)` and `normal_pdf(x, m, var)` compute the cdf and pdf, respectively, at `x` for a Gaussian random variable with mean `m` and variance `var`.

    (a) Plot the conditional pdfs in Example 4.11 if $v = \pm 2$ and the noise is zero-mean and unit variance.

    (b) Compare the cdf of the Gaussian random variable with the Chernoff bound obtained in Example 4.44.

4.135. Plot the pdf and cdf of the gamma random variable for the following cases.

    (a) $\lambda = 1$ and $\alpha = 1, 2, 4$.

    (b) $\lambda = 1/2$ and $\alpha = 1/2, 1, 3/2, 5/2$.

4.136. The random variable X has the triangular pdf shown in Fig. P4.4.

    (a) Find the transformation needed to generate X.

    (b) Use Octave to generate 100 samples of X. Compare the empirical pdf of the samples with the desired pdf.
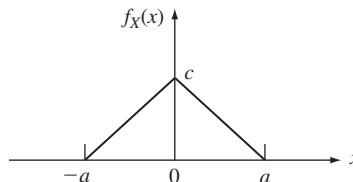
FIGURE P4.4

4.137. For each of the following random variables: Find the transformation needed to generate the random variable X; use Octave to generate 1000 samples of X; Plot the sequence of outcomes; compare the empirical pdf of the samples with the desired pdf.

    (a) Laplacian random variable with $\alpha = 1$.

    (b) Pareto random variable with $\alpha = 1.5, 2, 2.5$.

    (c) Weibull random variable with $\beta = 0.5, 2, 3$ and $\lambda = 1$.

4.138. A random variable Y of mixed type has pdf

$$f_Y(x) = p\delta(x) + (1 - p)f_Y(x),$$

where X is a Laplacian random variable and p is a number between zero and one. Find the transformation required to generate Y.

4.139. Specify the transformation method needed to generate the geometric random variable with parameter $p = 1/2$. Find the average number of comparisons needed in the search to determine each outcome.

4.140. Specify the transformation method needed to generate the Poisson random variable with small parameter α. Compute the average number of comparisons needed in the search.

4.141. The following rejection method can be used to generate Gaussian random variables:

    1. Generate U_1, a uniform random variable in the unit interval.

    2. Let $X_1 = -\ln(U_1)$.

    3. Generate U_2, a uniform random variable in the unit interval. If $U_2 \leq \exp\{-(X_1 - 1)^2/2\}$, accept X_1. Otherwise, reject X_1 and go to step 1.

    4. Generate a random sign ($+$ or $-$) with equal probability. Output X equal to X_1 with the resulting sign.

    (a) Show that if X_1 is accepted, then its pdf corresponds to the pdf of the absolute value of a Gaussian random variable with mean 0 and variance 1.

    (b) Show that X is a Gaussian random variable with mean 0 and variance 1.

4.142. Cheng (1977) has shown that the function $Kf_Z(x)$ bounds the pdf of a gamma random variable with $\alpha > 1$, where

$$f_Z(x) = \frac{\lambda\alpha^\lambda x^{\lambda-1}}{(\alpha^\lambda + x^\lambda)^2} \quad \text{and} \quad K = (2\alpha - 1)^{1/2}.$$

Find the cdf of $f_Z(x)$ and the corresponding transformation needed to generate Z.

4.143. (a) Show that in the modified rejection method, the probability of accepting X_1 is $1/K$. *Hint:* Use conditional probability.

   (b) Show that Z has the desired pdf.

4.144. Two methods for generating binomial random variables are: (1) Generate n Bernoulli random variables and add the outcomes; (2) Divide the unit interval according to binomial probabilities. Compare the methods under the following conditions:

   (a) $p = 1/2, n = 5, 25, 50$;

   (b) $p = 0.1, n = 5, 25, 50$.

   (c) Use Octave to implement the two methods by generating 1000 binomially distributed samples.

4.145. Let the number of event occurrences in a time interval be a Poisson random variable. In Section 3.4, it was found that the time between events for a Poisson random variable is an exponentially distributed random variable.

   (a) Explain how one can generate Poisson random variables from a sequence of exponentially distributed random variables.

   (b) How does this method compare with the one presented in Problem 4.140?

   (c) Use Octave to implement the two methods when $\alpha = 3$, $\alpha = 25$, and $\alpha = 100$.

4.146. Write a program to generate the gamma pdf with $\alpha > 1$ using the rejection method discussed in Problem 4.142. Use this method to generate m-Erlang random variables with $m = 2, 10$ and $\lambda = 1$ and compare the method to the straightforward generation of m exponential random variables as discussed in Example 4.57.

*Section 4.10: Entropy

4.147. Let X be the outcome of the toss of a fair die.

   (a) Find the entropy of X.

   (b) Suppose you are told that X is even. What is the reduction in entropy?

4.148. A biased coin is tossed three times.

   (a) Find the entropy of the outcome if the sequence of heads and tails is noted.

   (b) Find the entropy of the outcome if the number of heads is noted.

   (c) Explain the difference between the entropies in parts a and b.

4.149. Let X be the number of tails until the first heads in a sequence of tosses of a biased coin.

   (a) Find the entropy of X given that $X \geq k$.

   (b) Find the entropy of X given that $X \leq k$.

4.150. One of two coins is selected at random: Coin A has $P[\text{heads}] = 1/10$ and coin B has $P[\text{heads}] = 9/10$.

   (a) Suppose the coin is tossed once. Find the entropy of the outcome.

   (b) Suppose the coin is tossed twice and the sequence of heads and tails is observed. Find the entropy of the outcome.

4.151. Suppose that the randomly selected coin in Problem 4.150 is tossed until the first occurrence of heads. Suppose that heads occurs in the kth toss. Find the entropy regarding the identity of the coin.

4.152. A communication channel accepts input I from the set $\{0, 1, 2, 3, 4, 5, 6\}$. The channel output is $X = I + N \bmod 7$, where N is equally likely to be $+1$ or -1.

   (a) Find the entropy of I if all inputs are equiprobable.

   (b) Find the entropy of I given that $X = 4$.

4.153. Let X be a discrete random variable with entropy H_X.

    (a) Find the entropy of $Y = 2X$.

    (b) Find the entropy of any invertible transformation of X.

4.154. Let (X, Y) be the pair of outcomes from two independent tosses of a die.

    (a) Find the entropy of X.

    (b) Find the entropy of the pair (X, Y).

    (c) Find the entropy in n independent tosses of a die. Explain why entropy is additive in this case.

4.155. Let X be the outcome of the toss of a die, and let Y be a randomly selected integer less than or equal to X.

    (a) Find the entropy of Y.

    (b) Find the entropy of the pair (X, Y) and denote it by $H(X, Y)$.

    (c) Find the entropy of Y given $X = k$ and denote it by $g(k) = H(Y \mid X = k)$. Find $E[g(X)] = E[H(Y \mid X)]$.

    (d) Show that $H(X, Y) = H_X + E[H(Y \mid X)]$. Explain the meaning of this equation.

4.156. Let X take on values from $\{1, 2, \ldots, K\}$. Suppose that $P[X = K] = p$, and let H_Y be the entropy of X given that X is not equal to K. Show that $H_X = -p \ln p - (1 - p) \ln(1 - p) + (1 - p)H_Y$.

4.157. Let X be a uniform random variable in Example 4.62. Find and plot the entropy of Q as a function of the variance of the error $X - Q(X)$. *Hint:* Express the variance of the error in terms of d and substitute into the expression for the entropy of Q.

4.158. A communication channel accepts as input either 000 or 111. The channel transmits each binary input correctly with probability $1 - p$ and erroneously with probability p. Find the entropy of the input given that the output is 000; given that the output is 010.

4.159. Let X be a uniform random variable in the interval $[-a, a]$. Suppose we are told that the X is positive. Use the approach in Example 4.62 to find the reduction in entropy. Show that this is equal to the difference of the differential entropy of X and the differential entropy of X given $\{X > 0\}$.

4.160. Let X be uniform in $[a, b]$, and let $Y = 2X$. Compare the differential entropies of X and Y. How does this result differ from the result in Problem 4.153?

4.161. Find the pmf for the random variable X for which the sequence of questions in Fig. 4.26(a) is optimum.

4.162. Let the random variable X have $S_X = \{1, 2, 3, 4, 5, 6\}$ and pmf (3/8, 3/8, 1/8, 1/16, 1/32, 1/32). Find the entropy of X. What is the best code you can find for X?

4.163. Seven cards are drawn from a deck of 52 distinct cards. How many bits are required to represent all possible outcomes?

4.164. Find the optimum encoding for the geometric random variable with $p = 1/2$.

4.165. An urn experiment has 10 equiprobable distinct outcomes. Find the performance of the best tree code for encoding (a) a single outcome of the experiment; (b) a sequence of n outcomes of the experiment.

4.166. A binary information source produces n outputs. Suppose we are told that there are k 1's in these n outputs.

    (a) What is the best code to indicate which pattern of k 1's and $n - k$ 0's occurred?

    (b) How many bits are required to specify the value of k using a code with a fixed number of bits?

4.167. The random variable X takes on values from the set $\{1, 2, 3, 4\}$. Find the maximum entropy pmf for X given that $E[X] = 2$.

4.168. The random variable X is nonnegative. Find the maximum entropy pdf for X given that $E[X] = 10$.

4.169. Find the maximum entropy pdf of X given that $E[X^2] = c$.

4.170. Suppose we are given two parameters of the random variable X, $E[g_1(X)] = c_1$ and $E[g_2(X)] = c_2$.

    (a) Show that the maximum entropy pdf for X has the form

$$f_X(x) = Ce^{-\lambda_1 g_1(x) - \lambda_2 g_2(x)}.$$

    (b) Find the entropy of X.

4.171. Find the maximum entropy pdf of X given that $E[X] = m$ and $\mathrm{VAR}[X] = \sigma^2$.

Problems Requiring Cumulative Knowledge

4.172. Three types of customers arrive at a service station. The time required to service type 1 customers is an exponential random variable with mean 2. Type 2 customers have a Pareto distribution with $\alpha = 3$ and $x_m = 1$. Type 3 customers require a constant service time of 2 seconds. Suppose that the proportion of type 1, 2, and 3 customers is 1/2, 1/8, and 3/8, respectively. Find the probability that an arbitrary customer requires more than 15 seconds of service time. Compare the above probability to the bound provided by the Markov inequality.

4.173. The lifetime X of a light bulb is a random variable with

$$P[X > t] = 2/(2 + t) \text{ for } t > 0.$$

Suppose three new light bulbs are installed at time $t = 0$. At time $t = 1$ all three light bulbs are still working. Find the probability that at least one light bulb is still working at time $t = 9$.

4.174. The random variable X is uniformly distributed in the interval $[0, a]$. Suppose a is unknown, so we estimate a by the maximum value observed in n independent repetitions of the experiment; that is, we estimate a by $Y = \max\{X_1, X_2, \ldots, X_n\}$.

    (a) Find $P[Y \leq y]$.

    (b) Find the mean and variance of Y, and explain why Y is a good estimate for a when N is large.

4.175. The sample X of a signal is a Gaussian random variable with $m = 0$ and $\sigma^2 = 1$. Suppose that X is quantized by a nonuniform quantizer consisting of four intervals: $(-\infty, -a], (-a, 0], (0, a], \text{ and } (a, \infty)$.

    (a) Find the value of a so that X is equally likely to fall in each of the four intervals.

    (b) Find the representation point $x_i = q(X)$ for X in $(0, a]$ that minimizes the mean-squared error, that is,

$$\int_0^a (x - x_1)^2 f_X(x) \, dx \text{ is minimized.}$$

    Hint: Differentiate the above expression with respect to x_i. Find the representation points for the other intervals.

    (c) Evaluate the mean-squared error of the quantizer $E[(X - q(X)^2]$.

4.176. The output Y of a binary communication system is a unit-variance Gaussian random with mean zero when the input is "0" and mean one when the input is "one". Assume the input is 1 with probability p.

  (a) Find $P[\text{input is } 1 \mid y < Y < y + h]$ and $P[\text{input is } 0 \mid y < Y < y + h]$.

  (b) The receiver uses the following decision rule:

  If $P[\text{input is } 1 \mid y < Y < y + h] > P[\text{input is } 0 \mid y < Y < y + h]$, decide input was 1; otherwise, decide input was 0.

  Show that this decision rule leads to the following threshold rule:

  If $Y > T$, decide input was 1; otherwise, decide input was 0.

  (c) What is the probability of error for the above decision rule?

Pairs of Random Variables

Many random experiments involve several random variables. In some experiments a number of different quantities are measured. For example, the voltage signals at several points in a circuit at some specific time may be of interest. Other experiments involve the repeated measurement of a certain quantity such as the repeated measurement ("sampling") of the amplitude of an audio or video signal that varies with time. In Chapter 4 we developed techniques for calculating the probabilities of events involving a single random variable *in isolation*. In this chapter, we extend the concepts already introduced to two random variables:

- We use the joint pmf, cdf, and pdf to calculate the probabilities of events that involve the *joint* behavior of two random variables;
- We use expected value to define joint moments that summarize the behavior of two random variables;
- We determine when two random variables are independent, and we quantify their degree of "correlation" when they are not independent;
- We obtain conditional probabilities involving a pair of random variables.

In a sense we have already covered all the fundamental concepts of probability and random variables, and we are "simply" elaborating on the case of two or more random variables. Nevertheless, there are significant analytical techniques that need to be learned, e.g., double summations of pmf's and double integration of pdf's, so we first discuss the case of two random variables in detail because we can draw on our geometric intuition. Chapter 6 considers the general case of vector random variables. Throughout these two chapters you should be mindful of the forest (fundamental concepts) and the trees (specific techniques)!

5.1 TWO RANDOM VARIABLES

The notion of a random variable as a mapping is easily generalized to the case where two quantities are of interest. Consider a random experiment with sample space S and event class $\mathcal{F}$. We are interested in a function that assigns a pair of real numbers
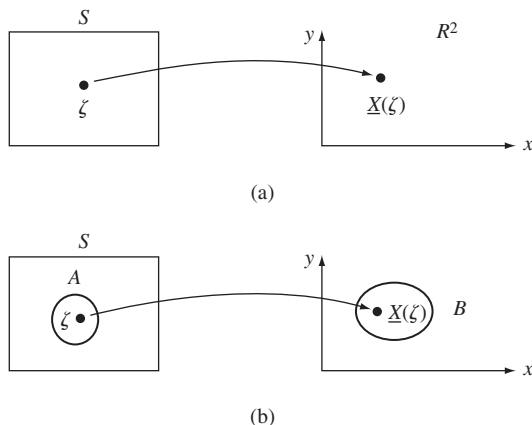
FIGURE 5.1
(a) A function assigns a pair of real numbers to each outcome in S. (b) Equivalent events for two random variables.

$\mathbf{X}(\zeta) = (X(\zeta), Y(\zeta))$ to each outcome ζ in S. Basically we are dealing with a vector function that maps S into R^2, the real plane, as shown in Fig. 5.1(a). We are ultimately interested in events involving the pair (X, Y).

Example 5.1

Let a random experiment consist of selecting a student's name from an urn. Let ζ denote the outcome of this experiment, and define the following two functions:

$$H(\zeta) = \text{height of student } \zeta \text{ in centimeters}$$

$$W(\zeta) = \text{weight of student } \zeta \text{ in kilograms}$$

$(H(\zeta), W(\zeta))$ assigns a pair of numbers to each ζ in S.

We are interested in events involving the pair (H, W). For example, the event $B = \{H \leq 183, W \leq 82\}$ represents students with height less that 183 cm (6 feet) and weight less than 82 kg (180 lb).

Example 5.2

A Web page provides the user with a choice either to watch a brief ad or to move directly to the requested page. Let ζ be the patterns of user arrivals in T seconds, e.g., number of arrivals, and listing of arrival times and types. Let $N_1(\zeta)$ be the number of times the Web page is directly requested and let $N_2(\zeta)$ be the number of times that the ad is chosen. $(N_1(\zeta), N_2(\zeta))$ assigns a pair of nonnegative integers to each ζ in S. Suppose that a type 1 request brings 0.001¢ in revenue and a type 2 request brings in 1¢. Find the event "revenue in T seconds is less than \$100."

The total revenue in T seconds is $0.001\ N_1 + 1 N_2$, and so the event of interest is $B = \{0.001\ N_1 + 1 N_2 < 10,000\}$.

Example 5.3

Let the outcome ζ in a random experiment be the length of a randomly selected message. Suppose that messages are broken into packets of maximum length M bytes. Let Q be the number of full packets in a message and let R be the number of bytes left over. $(Q(\zeta), R(\zeta))$ assigns a pair of numbers to each ζ in S. Q takes on values in the range $0, 1, 2, \ldots$, and R takes on values in the range $0, 1, \ldots, M - 1$. An event of interest may be $B = \{R < M/2\}$, "the last packet is less than half full."

Example 5.4

Let the outcome of a random experiment result in a pair $\underline{\zeta} = (\zeta_1, \zeta_2)$ that results from two independent spins of a wheel. Each spin of the wheel results in a number in the interval $(0, 2\pi]$. Define the pair of numbers (X, Y) in the plane as follows:

$$X(\underline{\zeta}) = \left(2 \ln \frac{2\pi}{\zeta_1} \right)^{1/2} \cos \zeta_2 \qquad Y(\underline{\zeta}) = \left(2 \ln \frac{2\pi}{\zeta_1} \right)^{1/2} \sin \zeta_2.$$

The vector function $(X(\underline{\zeta}), Y(\underline{\zeta}))$ assigns a pair of numbers in the plane to each $\underline{\zeta}$ in S. The square root term corresponds to a radius and to ζ_2 an angle.

  We will see that (X, Y) models the noise voltages encountered in digital communication systems. An event of interest here may be $B = \{X^2 + Y^2 < r^2\}$, "total noise power is less than r^2."

  The events involving a pair of random variables (X, Y) are specified by conditions that we are interested in and can be represented by regions in the plane. Figure 5.2 shows three examples of events:

$$A = \{X + Y \leq 10\}$$
$$B = \{\min(X, Y) \leq 5\}$$
$$C = \{X^2 + Y^2 \leq 100\}.$$

Event A divides the plane into two regions according to a straight line. Note that the event in Example 5.2 is of this type. Event C identifies a disk centered at the origin and
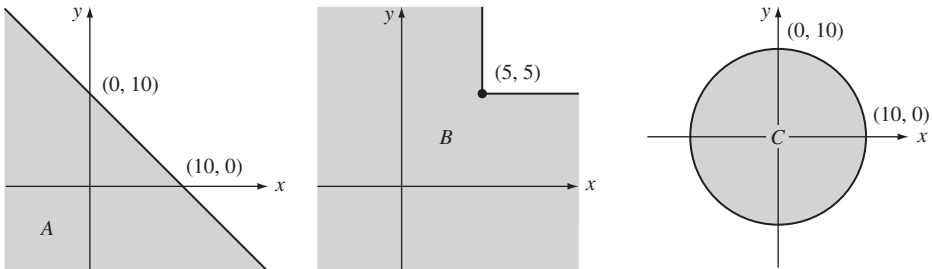

FIGURE 5.2
Examples of two-dimensional events.

it corresponds to the event in Example 5.4. Event B is found by noting that $\{\min(X, Y) \leq 5\} = \{X \leq 5\} \cup \{Y \leq 5\}$, that is, the minimum of X and Y is less than or equal to 5 if either X and/or Y is less than or equal to 5.

To determine the probability that the pair $\mathbf{X} = (X, Y)$ is in some region B in the plane, we proceed as in Chapter 3 to find the equivalent event for B in the underlying sample space S:

$$A = \mathbf{X}^{-1}(B) = \{\zeta: (X(\zeta), Y(\zeta)) \text{ in } B\}. \tag{5.1a}$$

The relationship between $A = \mathbf{X}^{-1}(B)$ and B is shown in Fig. 5.1(b). If A is in $\mathcal{F}$, then it has a probability assigned to it, and we obtain:

$$P[X \text{ in } B] = P[A] = P[\{\zeta: (X(\zeta), Y(\zeta)) \text{ in } B\}]. \tag{5.1b}$$

The approach is identical to what we followed in the case of a single random variable. The only difference is that we are considering the *joint behavior of X and Y* that is induced by the underlying random experiment.

A scattergram can be used to deduce the joint behavior of two random variables. A scattergram plot simply places a dot at every observation pair (x, y) that results from performing the experiment that generates (X, Y). Figure 5.3 shows the scattergram for 200 observations of four different pairs of random variables. The pairs in Fig. 5.3(a) appear to be uniformly distributed in the unit square. The pairs in Fig. 5.3(b) are clearly confined to a disc of unit radius and appear to be more concentrated near the origin. The pairs in Fig. 5.3(c) are concentrated near the origin, and appear to have circular symmetry, but are not bounded to an enclosed region. The pairs in Fig. 5.3(d) again are concentrated near the origin and appear to have a clear linear relationship of some sort, that is, larger values of x tend to have linearly proportional increasing values of y. We later introduce various functions and moments to characterize the behavior of pairs of random variables illustrated in these examples.

The joint probability mass function, joint cumulative distribution function, and joint probability density function provide approaches to specifying the probability law that governs the behavior of the pair (X, Y). Our general approach is as follows. We first focus on events that correspond to rectangles in the plane:

$$B = \{X \text{ in } A_1\} \cap \{Y \text{ in } A_2\} \tag{5.2}$$

where A_k is a one-dimensional event (i.e., subset of the real line). We say that these events are of **product form**. The event B occurs when both $\{X \text{ in } A_1\}$ and $\{Y \text{ in } A_2\}$ occur jointly. Figure 5.4 shows some two-dimensional product-form events. We use Eq. (5.1b) to find the probability of product-form events:

$$P[B] = P[\{X \text{ in } A_1\} \cap \{Y \text{ in } A_2\}] \triangleq P[X \text{ in } A_1, Y \text{ in } A_n]. \tag{5.3}$$

By defining A appropriately we then obtain the joint pmf, joint cdf, and joint pdf of (X, Y).

5.2   PAIRS OF DISCRETE RANDOM VARIABLES

Let the vector random variable $\mathbf{X} = (X, Y)$ assume values from some countable set $S_{X,Y} = \{(x_j, y_k), j = 1, 2, \ldots, k = 1, 2, \ldots\}$. The **joint probability mass function** of $\mathbf{X}$ specifies the probabilities of the event $\{X = x\} \cap \{Y = y\}$:
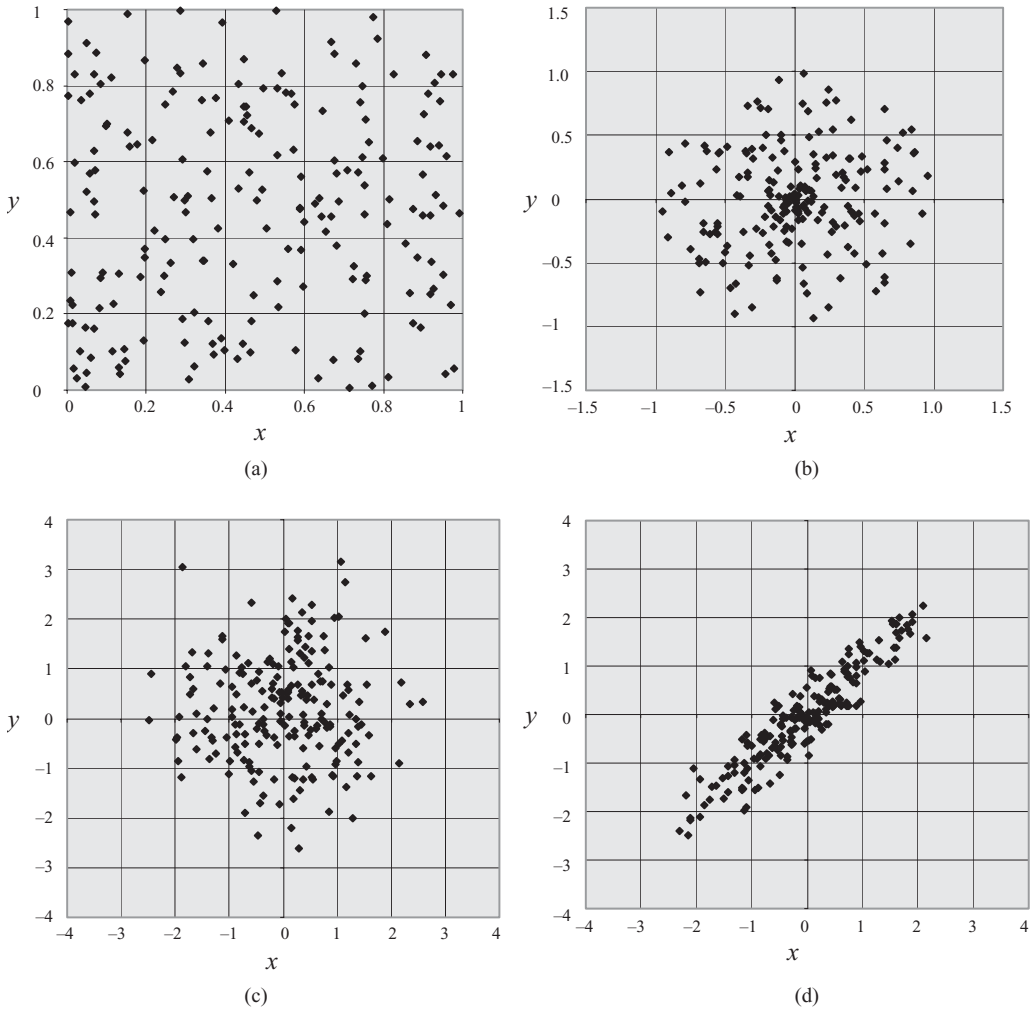
FIGURE 5.3

A scattergram for 200 observations of four different pairs of random variables.
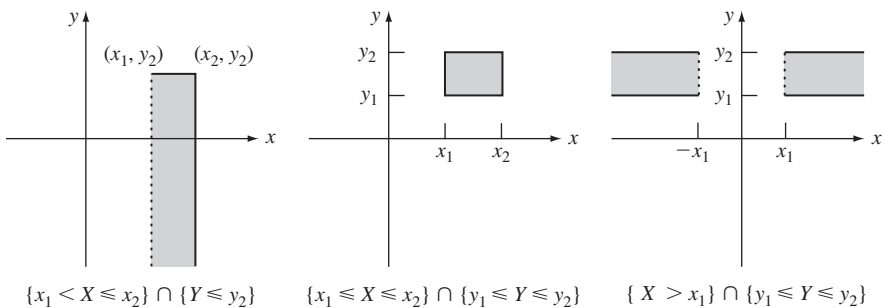
FIGURE 5.4

Some two-dimensional product-form events.

$$p_{X,Y}(x, y) = P[\{X = x\} \cap \{Y = y\}]$$
$$\triangleq P[X = x, Y = y] \quad \text{for } (x, y) \in R^2. \quad (5.4a)$$

The values of the pmf on the set $S_{X,Y}$ provide the essential information:

$$p_{X,Y}(x_j, y_k) = P[\{X = x_j\} \cap \{Y = y_k\}]$$
$$\triangleq P[X = x_j, Y = y_k] \quad (x_j, y_k) \in S_{X,Y}. \quad (5.4b)$$

There are several ways of showing the pmf graphically: (1) For small sample spaces we can present the pmf in the form of a table as shown in Fig. 5.5(a). (2) We can present the pmf using arrows of height $p_{X,Y}(x_j, y_k)$ placed at the points $\{(x_j, y_k)\}$ in the plane, as shown in Fig. 5.5(b), but this can be difficult to draw. (3) We can place dots at the points $\{(x_j, y_k)\}$ and label these with the corresponding pmf value as shown in Fig. 5.5(c).

The probability of any event B is the sum of the pmf over the outcomes in B:

$$P[\mathbf{X} \text{ in } B] = \sum_{(x_j, y_k) \text{ in } B} \sum p_{X,Y}(x_j, y_k). \quad (5.5)$$

Frequently it is helpful to sketch the region that contains the points in B as shown, for example, in Fig. 5.6. When the event B is the entire sample space $S_{X,Y}$, we have:

$$\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} p_{X,Y}(x_j, y_k) = 1. \quad (5.6)$$

Example 5.5

A packet switch has two input ports and two output ports. At a given time slot a packet arrives at each input port with probability 1/2, and is equally likely to be destined to output port 1 or 2. Let X and Y be the number of packets destined for output ports 1 and 2, respectively. Find the pmf of X and Y, and show the pmf graphically.

The outcome I_j for an input port j can take the following values: "n", no packet arrival (with probability 1/2); "a1", packet arrival destined for output port 1 (with probability 1/4); "a2", packet arrival destined for output port 2 (with probability 1/4). The underlying sample space S consists of the pair of input outcomes $\zeta = (I_1, I_2)$. The mapping for (X, Y) is shown in the table below:

ζ	(n, n)	(n, a1)	(n, a2)	(a1, n)	(a1, a1)	(a1, a2)	(a2, n)	(a2, a1)	(a2, a2)
X, Y	(0,0)	(1,0)	(0,1)	(1,0)	(2,0)	(1,1)	(0,1)	(1,1)	(0,2)

The pmf of (X, Y) is then:

$$p_{X,Y}(0, 0) = P[\zeta = (n, n)] = \frac{1}{2}\frac{1}{2} = \frac{1}{4},$$

$$p_{X,Y}(0, 1) = P[\zeta \in \{(n, a2), (a2, n)\}] = 2 * \frac{1}{8} = \frac{1}{4},$$
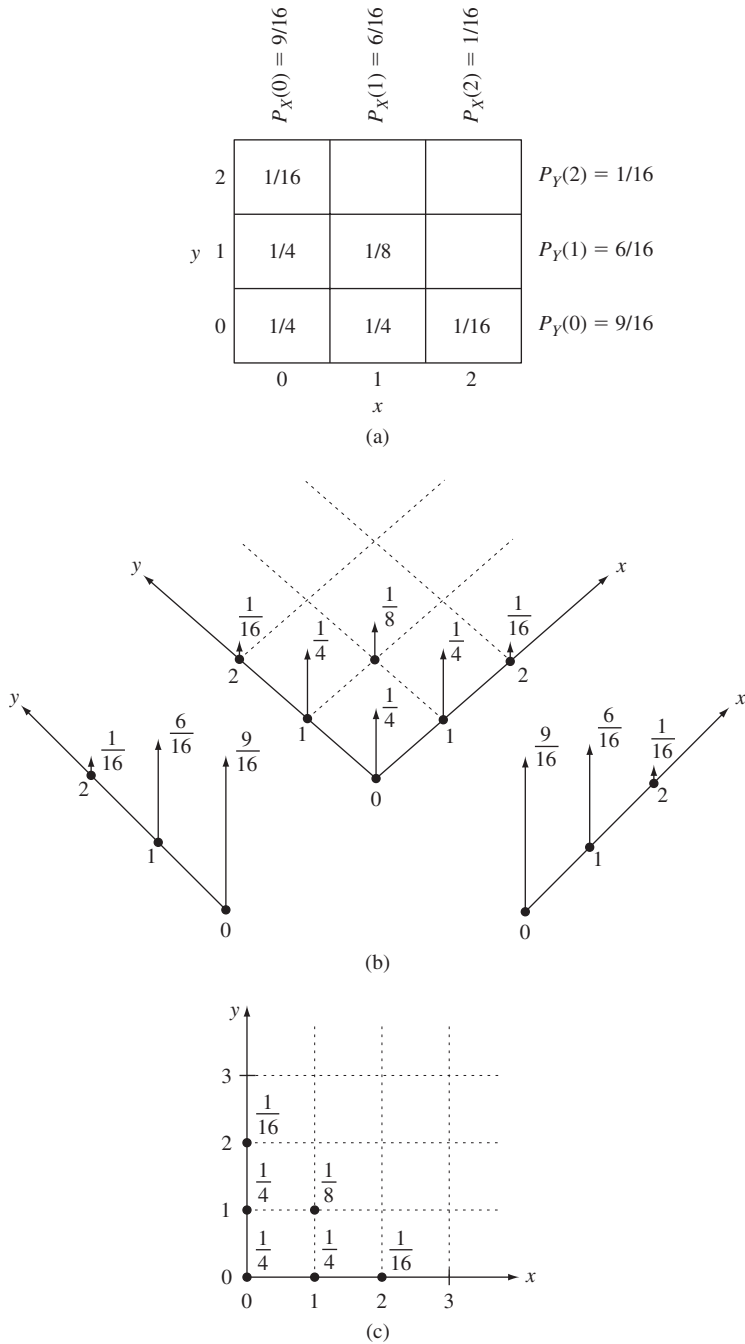
FIGURE 5.5
Graphical representations of pmf's: (a) in table format; (b) use of arrows to show height;
(c) labeled dots corresponding to pmf value.
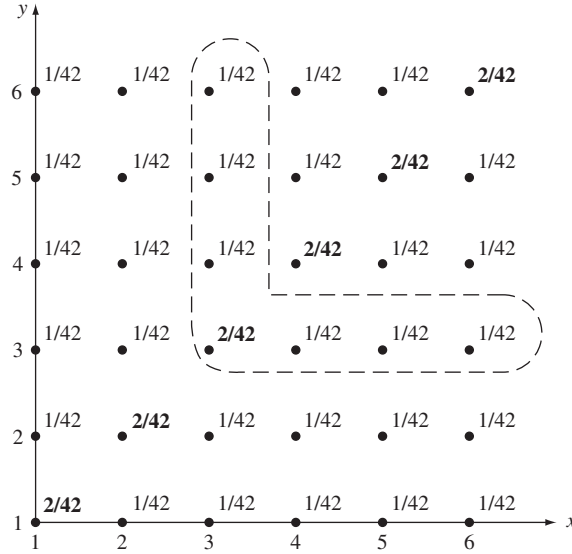
FIGURE 5.6
Showing the pmf via a sketch containing the points in *B*.

$$p_{X,Y}(1, 0) = P[\zeta \in \{(n, a1), (a1, n)\}] = \frac{1}{4},$$

$$p_{X,Y}(1, 1) = P[\zeta \in \{(a1, a2), (a2, a1)\}] = \frac{1}{8},$$

$$p_{X,Y}(0, 2) = P[\zeta = (a2, a2)] = \frac{1}{16},$$

$$p_{X,Y}(2, 0) = P[\zeta = (a1, a1)] = \frac{1}{16}.$$

Figure 5.5(a) shows the pmf in tabular form where the number of rows and columns accommodate the range of *X* and *Y* respectively. Each entry in the table gives the pmf value for the corresponding *x* and *y*. Figure 5.5(b) shows the pmf using arrows in the plane. An arrow of height $p_{X,Y}(j, k)$ is placed at each of the points in $S_{X,Y} = \{(0, 0), (0, 1), (1, 0), (1, 1), (0, 2), (2, 0)\}$. Figure 5.5(c) shows the pmf using labeled dots in the plane. A dot with label $p_{X,Y}(j, k)$ is placed at each of the points in $S_{X,Y}$.

Example 5.6

A random experiment consists of tossing two "loaded" dice and noting the pair of numbers (X, Y) facing up. The joint pmf $p_{X,Y}(j, k)$ for $j = 1, \ldots, 6$ and $k = 1, \ldots, 6$ is given by the two-dimensional table shown in Fig. 5.6. The (j, k) entry in the table contains the value $p_{X,Y}(j, k)$. Find the $P[\min(X, Y) = 3]$.

Figure 5.6 shows the region that corresponds to the set $\{\min(x, y) = 3\}$. The probability of this event is given by:

$$P[\min(X, Y) = 3] = p_{X,Y}(6, 3) + p_{X,Y}(5, 3) + p_{X,Y}(4, 3)$$
$$+ p_{X,Y}(3, 3) + p_{X,Y}(3, 4) + p_{X,Y}(3, 5) + p_{X,Y}(3, 6)$$
$$= 6\left(\frac{1}{42}\right) + \frac{2}{42} = \frac{8}{42}.$$

5.2.1   Marginal Probability Mass Function

The joint pmf of **X** provides the information about the joint behavior of X and Y. We are also interested in the probabilities of events involving each of the random variables in isolation. These can be found in terms of the **marginal probability mass functions**:

$$p_X(x_j) = P[X = x_j]$$
$$= P[X = x_j, Y = \text{anything}]$$
$$= P[\{X = x_j \text{ and } Y = y_1\} \cup \{X = x_j \text{ and } Y = y_2\} \cup \ldots]$$
$$= \sum_{k=1}^{\infty} p_{X,Y}(x_j, y_k), \tag{5.7a}$$

and similarly,

$$p_Y(y_k) = P[Y = y_k]$$
$$= \sum_{j=1}^{\infty} p_{X,Y}(x_j, y_k). \tag{5.7b}$$

The marginal pmf's satisfy all the properties of one-dimensional pmf's, and they supply the information required to compute the probability of events involving the corresponding random variable.

The probability $p_{X,Y}(x_j, y_k)$ can be interpreted as the long-term relative frequency of the joint event $\{X = X_j\} \cap \{Y = Y_k\}$ in a sequence of repetitions of the random experiment. Equation (5.7a) corresponds to the fact that the relative frequency of the event $\{X = X_j\}$ is found by adding the relative frequencies of all outcome pairs in which X_j appears. In general, it is impossible to deduce the relative frequencies of pairs of values X and Y from the relative frequencies of X and Y in isolation. The same is true for pmf's: In general, knowledge of the marginal pmf's is insufficient to specify the joint pmf.

Example 5.7

Find the marginal pmf for the output ports (X, Y) in Example 5.2.

Figure 5.5(a) shows that the marginal pmf is found by adding entries along a row or column in the table. For example, by adding along the $x = 1$ column we have:

$$p_X(1) = P[X = 1] = p_{X,Y}(1, 0) + p_{X,Y}(1, 1) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}.$$

Similarly, by adding along the $y = 0$ row:

$$p_Y(0) = P[Y = 0] = p_{X,Y}(0, 0) + p_{X,Y}(1, 0) + p_{X,Y}(2, 0) = \frac{1}{4} + \frac{1}{4} + \frac{1}{16} = \frac{9}{16}.$$

Figure 5.5(b) shows the marginal pmf using arrows on the real line.

Example 5.8

Find the marginal pmf's in the loaded dice experiment in Example 5.2.

The probability that $X = 1$ is found by summing over the first row:

$$P[X = 1] = \frac{2}{42} + \frac{1}{42} + \cdots + \frac{1}{42} = \frac{1}{6}.$$

Similarly, we find that $P[X = j] = 1/6$ for $j = 2, \ldots, 6$. The probability that $Y = k$ is found by summing over the kth column. We then find that $P[Y = k] = 1/6$ for $k = 1, 2, \ldots, 6$. Thus each die, in isolation, appears to be fair in the sense that each face is equiprobable. If we knew only these marginal pmf's we would have no idea that the dice are loaded.

Example 5.9

In Example 5.3, let the number of bytes N in a message have a geometric distribution with parameter $1 - p$ and range $S_N = \{0, 1, 2, \ldots\}$. Find the joint pmf and the marginal pmf's of Q and R.

If a message has N bytes, then the number of full packets is the quotient Q in the division of N by M, and the number of remaining bytes is the remainder R. The probability of the pair $\{(q, r)\}$ is given by

$$P[Q = q, R = r] = P[N = qM + r] = (1 - p)p^{qM+r}.$$

The marginal pmf of Q is

$$P[Q = q] = P[N \text{ in} \{qM, qM + 1, \ldots, qM + (M - 1)\}]$$

$$= \sum_{k=0}^{(M-1)} (1 - p)p^{qM+k}$$

$$= (1 - p)p^{qM}\frac{1 - p^M}{1 - p} = (1 - p^M)(p^M)^q \qquad q = 0, 1, 2, \ldots$$

The marginal pmf of Q is geometric with parameter p^M. The marginal pmf of R is:

$$P[R = r] = P[N \text{ in} \{r, M + r, 2M + r, \ldots\}]$$

$$= \sum_{q=0}^{\infty} (1 - p)p^{qM+r} = \frac{(1 - p)}{1 - p^M}p^r \quad r = 0, 1, \ldots, M - 1.$$

R has a truncated geometric pmf. As an exercise, you should verify that all the above marginal pmf's add to 1.

5.3   THE JOINT CDF OF X AND Y

In Chapter 3 we saw that semi-infinite intervals of the form $(-\infty, x]$ are a basic building block from which other one-dimensional events can be built. By defining the cdf $F_X(x)$ as the probability of $(-\infty, x]$, we were then able to express the probabilities of other events in terms of the cdf. In this section we repeat the above development for two-dimensional random variables.
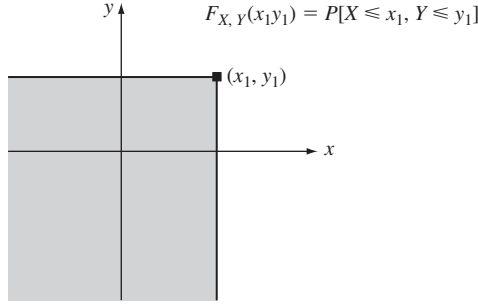
FIGURE 5.7
The joint cumulative distribution function is defined as
the probability of the semi-infinite rectangle defined by
the point (x_1, y_1).

A basic building block for events involving two-dimensional random variables is
the semi-infinite rectangle defined by $\{(x, y): x \leq x_1 \text{ and } y \leq y_1\}$, as shown in Fig. 5.7.
We also use the more compact notation $\{x \leq x_1, y \leq y_1\}$ to refer to this region. The
joint cumulative distribution function of X and Y is defined as the probability of the
event $\{X \leq x_1\} \cap \{Y \leq y_1\}$:

$$F_{X,Y}(x_1, y_1) = P[X \leq x_1, Y \leq y_1]. \tag{5.8}$$

In terms of relative frequency, $F_{X,Y}(x_1, y_1)$ represents the long-term proportion
of time in which the outcome of the random experiment yields a point X that falls in
the rectangular region shown in Fig. 5.7. In terms of probability "mass," $F_{X,Y}(x_1, y_1)$
represents the amount of mass contained in the rectangular region.

The joint cdf satisfies the following properties.

(i) The joint cdf is a nondecreasing function of x and y:

$$F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2) \qquad \text{if } x_1 \leq x_2 \text{ and } y_1 \leq y_2, \tag{5.9a}$$

(ii) $F_{X,Y}(x_1, -\infty) = 0, \qquad F_{X,Y}(-\infty, y_1) = 0, \qquad F_{X,Y}(\infty, \infty) = 1. \tag{5.9b}$

(iii) We obtain the **marginal cumulative distribution functions** by removing the
constraint on one of the variables. The marginal cdf's are the probabilities of
the regions shown in Fig. 5.8:

$$F_X(x_1) = F_{X,Y}(x_1, \infty) \quad \text{and} \quad F_Y(y_1) = F_{X,Y}(\infty, y_1). \tag{5.9c}$$

(iv) The joint cdf is continuous from the "north" and from the "east," that is,

$$\lim_{x \to a^+} F_{X,Y}(x, y) = F_{X,Y}(a, y) \quad \text{and} \quad \lim_{y \to b^+} F_{X,Y}(x, y) = F_{X,Y}(x, b). \tag{5.9d}$$

(v) The probability of the rectangle $\{x_1 < x \leq x_2, y_1 < y \leq y_2\}$ is given by:

$$P[x_1 < X \leq x_2, y_1 < Y \leq y_2] =$$
$$F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1). \tag{5.9e}$$

$$F_X(x_1) = P[X \leq x_1, Y < \infty] \qquad F_Y(y_1) = P[X < \infty, Y \leq y_1]$$
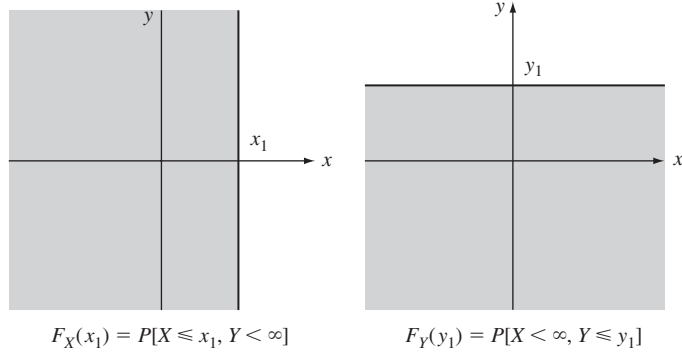
FIGURE 5.8
The marginal cdf's are the probabilities of these half-planes.

Property (i) follows by noting that the semi-infinite rectangle defined by (x_1, y_1) is contained in that defined by (x_2, y_2) and applying Corollary 7. Properties (ii) to (iv) are obtained by limiting arguments. For example, the sequence $\{x \leq x_1 \text{ and } y \leq -n\}$ is decreasing and approaches the empty set $\varnothing$, so

$$F_{X,Y}(x_1, -\infty) = \lim_{n \to \infty} F_{X,Y}(x_1, -n) = P[\varnothing] = 0.$$

For property (iii) we take the sequence $\{x \leq x_1 \text{ and } y \leq n\}$ which increases to $\{x \leq x_1\}$, so

$$\lim_{n \to \infty} F_{X,Y}(x_1, n) = P[X \leq x_1] = F_X(x_1).$$

For property (v) note in Fig. 5.9(a) that $B = \{x_1 < x \leq x_2, y \leq y_1\} = \{X \leq x_2, Y \leq y_1\} - \{X \leq x_1, Y \leq y_1\}$, so $P[B] = P[x_1 < X \leq x_2, Y \leq y_1] = F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_1)$. In Fig. 5.9(b), note that $F_{X,Y}(x_2, y_2) = P[A] + P[B] + F_{X,Y}(x_1, y_2)$. Property (v) follows by solving for $P[A]$ and substituting the expression for $P[B]$.

(a)   (b)
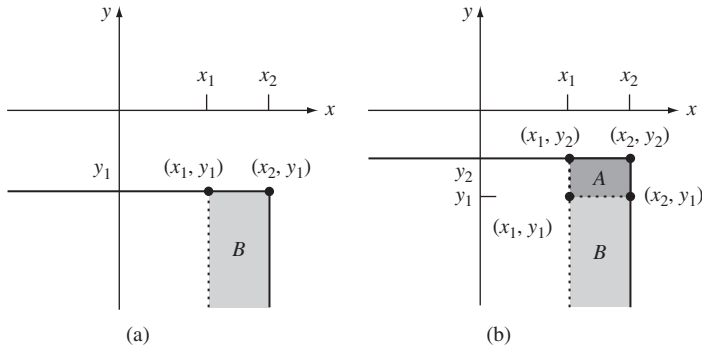
FIGURE 5.9
The joint cdf can be used to determine the probability of various events.
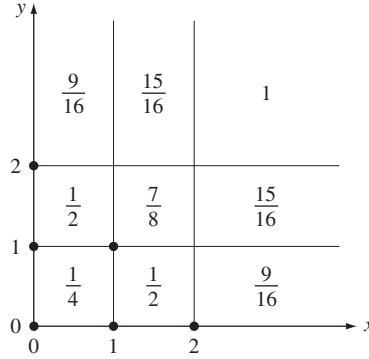
FIGURE 5.10
Joint cdf for packet switch example.

Example 5.10

Plot the joint cdf of X and Y from Example 5.6. Find the marginal cdf of X.
  To find the cdf of **X**, we identify the regions in the plane according to which points in $S_{X,Y}$ are included in the rectangular region defined by (x, y). For example,

- The regions outside the first quadrant do not include any of the points, so $F_{X,Y}(x, y) = 0$.
- The region $\{0 \le x < 1, 0 \le y < 1\}$ contains the point $(0,0)$, so $F_{X,Y}(x, y) = 1/4$.

Figure 5.10 shows the cdf after all possible regions are examined.
  We need to consider several cases to find $F_X(x)$. For $x < 0$, we have $F_X(x) = 0$. For $0 \le x < 1$, we have $F_X(x) = F_{X,Y}(x, \infty) = 9/16$. For $1 \le x < 2$, we have $F_X(x) = F_{X,Y}(x, \infty) = 15/16$. Finally, for $x \ge 1$, we have $F_X(x) = F_{X,Y}(x, \infty) = 1$. Therefore $F_X(x)$ is a staircase function and X is a discrete random variable with $p_X(0) = 9/16$, $p_X(1) = 6/16$, and $p_X(2) = 1/16$.

Example 5.11

The joint cdf for the pair of random variables $\mathbf{X} = (X, Y)$ is given by

$$F_{X,Y}(x, y) = \begin{cases} 0 & x < 0 \text{ or } y < 0 \\ xy & 0 \le x \le 1, 0 \le y \le 1 \\ x & 0 \le x \le 1, y > 1 \\ y & 0 \le y \le 1, x > 1 \\ 1 & x \ge 1, y \ge 1. \end{cases} \tag{5.10}$$

Plot the joint cdf and find the marginal cdf of X.
  Figure 5.11 shows a plot of the joint cdf of X and Y. $F_{X,Y}(x, y)$ is continuous for all points in the plane. $F_{X,Y}(x, y) = 1$ for all $x \ge 1$ and $y \ge 1$, which implies that X and Y each assume values less than or equal to one.
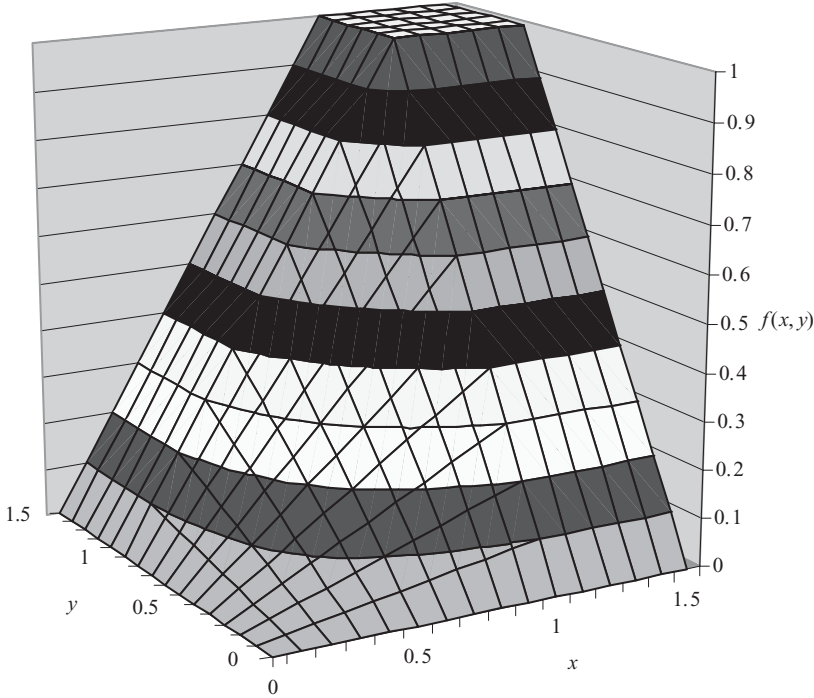
FIGURE 5.11
Joint cdf for two uniform random variables.

The marginal cdf of X is:

$$F_X(x) = F_{X,Y}(x, \infty) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x \geq 1. \end{cases}$$

X is uniformly distributed in the unit interval.

Example 5.12

The joint cdf for the vector of random variable $\mathbf{X} = (X, Y)$ is given by

$$F_{X,Y}(x, y) = \begin{cases} (1 - e^{-\alpha x})(1 - e^{-\beta y}) & x \geq 0, y \geq 0 \\ 0 & \text{elsewhere.} \end{cases}$$

Find the marginal cdf's.

The marginal cdf's are obtained by letting one of the variables approach infinity:

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x, y) = 1 - e^{-\alpha x} \quad x \geq 0$$

$$F_Y(y) = \lim_{x \to \infty} F_{X,Y}(x, y) = 1 - e^{-\beta y} \quad y \geq 0.$$

X and Y individually have exponential distributions with parameters α and β, respectively.

Example 5.13

Find the probability of the events $A = \{X \leq 1, Y \leq 1\}$, $B = \{X > x, Y > y\}$, where $x > 0$ and $y > 0$, and $D = \{1 < X \leq 2, 2 < Y \leq 5\}$ in Example 5.12.

The probability of A is given directly by the cdf:

$$P[A] = P[X \leq 1, Y \leq 1] = F_{X,Y}(1, 1) = (1 - e^{-\alpha})(1 - e^{-\beta}).$$

The probability of B requires more work. By DeMorgan's rule:

$$B^c = (\{X > x\} \cap \{Y > y\})^c = \{X \leq x\} \cup \{Y \leq y\}.$$

Corollary 5 in Section 2.2 gives the probability of the union of two events:

$$\begin{aligned} P[B^c] &= P[X \leq x] + P[Y \leq y] - P[X \leq x, Y \leq y] \\ &= (1 - e^{-\alpha x}) + (1 - e^{-\beta y}) - (1 - e^{-\alpha x})(1 - e^{-\beta y}) \\ &= 1 - e^{-\alpha x}e^{-\beta y}. \end{aligned}$$

Finally we obtain the probability of B:

$$P[B] = 1 - P[B^c] = e^{-\alpha x}e^{-\beta y}.$$

You should sketch the region B on the plane and identify the events involved in the calculation of the probability of B^c.

The probability of event D is found by applying property (vi) of the joint cdf:

$$\begin{aligned} P[1 < X &\leq 2, 2 < Y \leq 5] \\ &= F_{X,Y}(2, 5) - F_{X,Y}(2, 2) - F_{X,Y}(1, 5) + F_{X,Y}(1, 2) \\ &= (1 - e^{-2\alpha})(1 - e^{-5\beta}) - (1 - e^{-2\alpha})(1 - e^{-2\beta}) \\ &\quad -(1 - e^{-\alpha})(1 - e^{-5\beta}) + (1 - e^{-\alpha})(1 - e^{-2\beta}). \end{aligned}$$

5.3.1   Random Variables That Differ in Type

In some problems it is necessary to work with joint random variables that differ in type, that is, one is discrete and the other is continuous. Usually it is rather clumsy to work with the joint cdf, and so it is preferable to work with either $P[X = k, Y \leq y]$ or $P[X = k, y_1 < Y \leq y_2]$. These probabilities are sufficient to compute the joint cdf should we have to.

Example 5.14   Communication Channel with Discrete Input and Continuous Output

The input X to a communication channel is $+1$ volt or -1 volt with equal probability. The output Y of the channel is the input plus a noise voltage N that is uniformly distributed in the interval from -2 volts to $+2$ volts. Find $P[X = +1, Y \leq 0]$.

This problem lends itself to the use of conditional probability:

$$P[X = +1, Y \leq y] = P[Y \leq y \mid X = +1]P[X = +1],$$

where $P[X = +1] = 1/2$. When the input $X = 1$, the output Y is uniformly distributed in the interval $[-1, 3]$; therefore

$$P[Y \leq y \mid X = +1] = \frac{y + 1}{4} \quad \text{for } -1 \leq y \leq 3.$$

Thus $P[X = +1, Y \leq 0] = P[Y \leq 0 \mid X = +1]P[X = +1] = (1/2)(1/4) = 1/8$.

5.4   THE JOINT PDF OF TWO CONTINUOUS RANDOM VARIABLES

The joint cdf allows us to compute the probability of events that correspond to "rectangular" shapes in the plane. To compute the probability of events corresponding to regions other than rectangles, we note that any reasonable shape (i.e., disk, polygon, or half-plane) can be approximated by the union of disjoint infinitesimal rectangles, $B_{j,k}$. For example, Fig. 5.12 shows how the events $A = \{X + Y \leq 1\}$ and $B = \{X^2 + X^2 \leq 1\}$ are approximated by rectangles of infinitesimal width. The probability of such events can therefore be approximated by the sum of the probabilities of infinitesimal rectangles, and if the cdf is sufficiently smooth, the probability of each rectangle can be expressed in terms of a density function:

$$P[B] \approx \sum_j \sum_k P[B_{j,k}] = \sum \sum_{(x_j, y_k) \in B} f_{X,Y}(x_j, y_k)\, \Delta x \Delta y.$$

As Δx and Δy approach zero, the above equation becomes an integral of a probability density function over the region B.

We say that the random variables **X and Y are jointly continuous** if the probabilities of events involving (X, Y) can be expressed as an integral of a probability density function. In other words, there is a nonnegative function $f_{X,Y}(x, y)$, called the **joint**
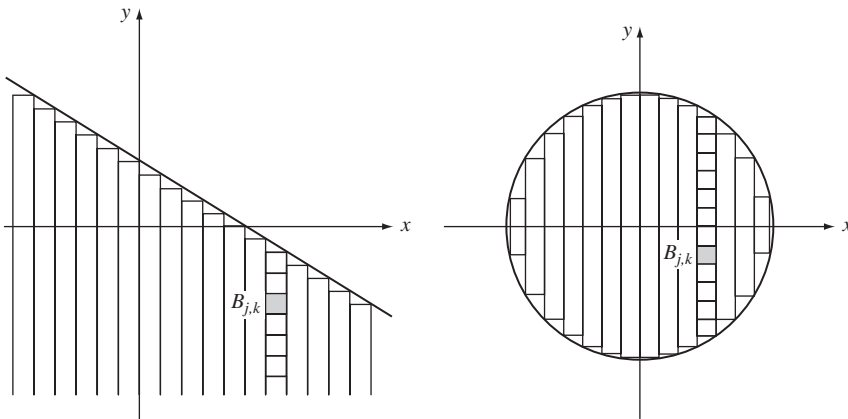
FIGURE 5.12
Some two-dimensional non-product form events.
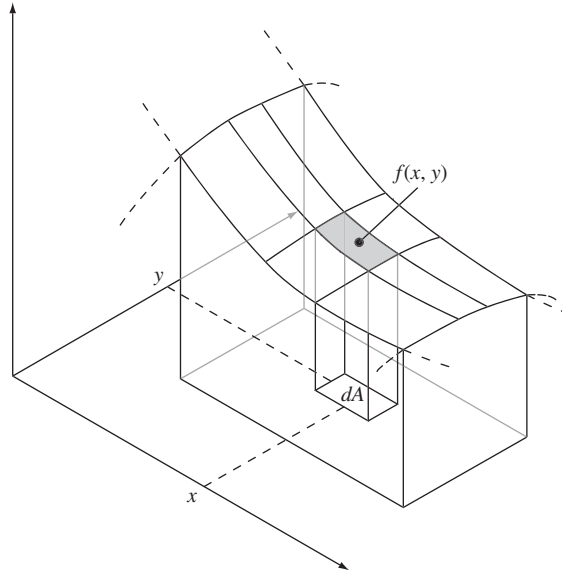
FIGURE 5.13
The probability of A is the integral of $f_{X,Y}(x, y)$ over the region
defined by A.

probability density function, that is defined on the real plane such that for every event
B, a subset of the plane,

$$P[\mathbf{X} \text{ in } B] = \int_B \int f_{X,Y}(x', y') \, dx' \, dy', \tag{5.11}$$

as shown in Fig. 5.13. Note the similarity to Eq. (5.5) for discrete random variables.
When B is the entire plane, the integral must equal one:

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x', y') \, dx' \, dy'. \tag{5.12}$$

Equations (5.11) and (5.12) again suggest that the probability "mass" of an event is
found by integrating the density of probability mass over the region corresponding to
the event.

    The joint cdf can be obtained in terms of the joint pdf of jointly continuous ran-
dom variables by integrating over the semi-infinite rectangle defined by (x, y):

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x', y') \, dx' \, dy'. \tag{5.13}$$

It then follows that *if X and Y* are jointly continuous random variables, then the pdf
can be obtained from the cdf by differentiation:

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \, \partial y}. \tag{5.14}$$

Note that if X and Y are not jointly continuous, then it is possible that the above partial derivative does not exist. In particular, if the $F_{X,Y}(x, y)$ is discontinuous or if its partial derivatives are discontinuous, then the joint pdf as defined by Eq. (5.14) will not exist.

The probability of a rectangular region is obtained by letting $B = \{(x, y): a_1 < x \le b_1$ and $a_2 < y \le b_2\}$ in Eq. (5.11):

$$P[a_1 < X \le b_1, a_2 < Y \le b_2] = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{X,Y}(x', y')\, dx'\, dy'. \qquad (5.15)$$

It then follows that the probability of an infinitesimal rectangle is the product of the pdf and the area of the rectangle:

$$P[x < X \le x + dx, y < Y \le y + dy] = \int_x^{x+dx} \int_y^{y+dy} f_{X,Y}(x', y')\, dx'\, dy'$$

$$\simeq f_{X,Y}(x, y)\, dx\, dy. \qquad (5.16)$$

Equation (5.16) can be interpreted as stating that the joint pdf specifies the probability of the product-form events

$$\{x < X \le x + dx\} \cap \{y < Y \le y + dy\}.$$

The **marginal pdf's** $f_X(x)$ and $f_Y(y)$ are obtained by taking the derivative of the corresponding marginal cdf's, $F_X(x) = F_{X,Y}(x, \infty)$ and $F_Y(y) = F_{X,Y}(\infty, y)$. Thus

$$f_X(x) = \frac{d}{dx} \int_{-\infty}^{x} \left\{ \int_{-\infty}^{\infty} f_{X,Y}(x', y')\, dy' \right\} dx'$$

$$= \int_{-\infty}^{\infty} f_{X,Y}(x, y')\, dy'. \qquad (5.17a)$$

Similarly,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x', y)\, dx'. \qquad (5.17b)$$

Thus the marginal pdf's are obtained by integrating out the variables that are not of interest.

Note that $f_X(x)\, dx \simeq P[x < X \le x + dx, Y < \infty]$ is the probability of the infinitesimal strip shown in Fig. 5.14(a). This reminds us of the interpretation of the marginal pmf's as the probabilities of columns and rows in the case of discrete random variables. It is not surprising then that Eqs. (5.17a) and (5.17b) for the marginal pdf's and Eqs. (5.7a) and (5.7b) for the marginal pmf's are identical except for the fact that one contains an integral and the other a summation. As in the case of pmf's, we note that, in general, the joint pdf cannot be obtained from the marginal pdf's.

$$f_X(x)dx \cong P[x < X \leqslant x + dx, Y < \infty] \qquad f_Y(y)dy \cong P[X < \infty, y < Y \leqslant y + dy]$$

(a)   (b)

FIGURE 5.14
Interpretation of marginal pdf's.

Example 5.15   Jointly Uniform Random Variables

A randomly selected point (X, Y) in the unit square has the uniform joint pdf given by

$$f_{X,Y}(x, y) = \begin{cases} 1 & 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

The scattergram in Fig. 5.3(a) corresponds to this pair of random variables. Find the joint cdf of X and Y.

    The cdf is found by evaluating Eq. (5.13). You must be careful with the limits of the integral: The limits should define the region consisting of the intersection of the semi-infinite rectangle defined by (x, y) and the region where the pdf is nonzero. There are five cases in this problem, corresponding to the five regions shown in Fig. 5.15.

1.   If $x < 0$ or $y < 0$, the pdf is zero and Eq. (5.14) implies

$$F_{X,Y}(x, y) = 0.$$

2.   If (x, y) is inside the unit interval,

$$F_{X,Y}(x, y) = \int_0^x \int_0^y 1 \, dx' \, dy' = xy.$$

3.   If $0 \leq x \leq 1$ and $y > 1$,

$$F_{X,Y}(x, y) = \int_0^x \int_0^1 1 \, dx' \, dy' = x.$$

4.   Similarly, if $x > 1$ and $0 \leq y \leq 1$,

$$F_{X,Y}(x, y) = y.$$

FIGURE 5.15
Regions that need to be considered separately in computing cdf
in Example 5.15.

5. Finally, if $x > 1$ and $y > 1$,

$$F_{X,Y}(x, y) = \int_0^1 \int_0^1 1 \, dx' \, dy' = 1.$$

We see that this is the joint cdf of Example 5.11.

Example 5.16

Find the normalization constant c and the marginal pdf's for the following joint pdf:

$$f_{X,Y}(x, y) = \begin{cases} ce^{-x}e^{-y} & 0 \le y \le x < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

The pdf is nonzero in the shaded region shown in Fig. 5.16(a). The constant c is found from the normalization condition specified by Eq. (5.12):

$$1 = \int_0^\infty \int_0^x ce^{-x}e^{-y} \, dy \, dx = \int_0^\infty ce^{-x}(1 - e^{-x}) \, dx = \frac{c}{2}.$$

Therefore $c = 2$. The marginal pdf's are found by evaluating Eqs. (5.17a) and (5.17b):

$$f_X(x) = \int_0^\infty f_{X,Y}(x, y) \, dy = \int_0^x 2e^{-x}e^{-y} \, dy = 2e^{-x}(1 - e^{-x}) \qquad 0 \le x < \infty$$

and

$$f_Y(y) = \int_0^\infty f_{X,Y}(x, y) \, dx = \int_y^\infty 2e^{-x}e^{-y} \, dx = 2e^{-2y} \qquad 0 \le y < \infty.$$

You should fill in the steps in the evaluation of the integrals as well as verify that the marginal pdf's integrate to 1.

(a)   (b)
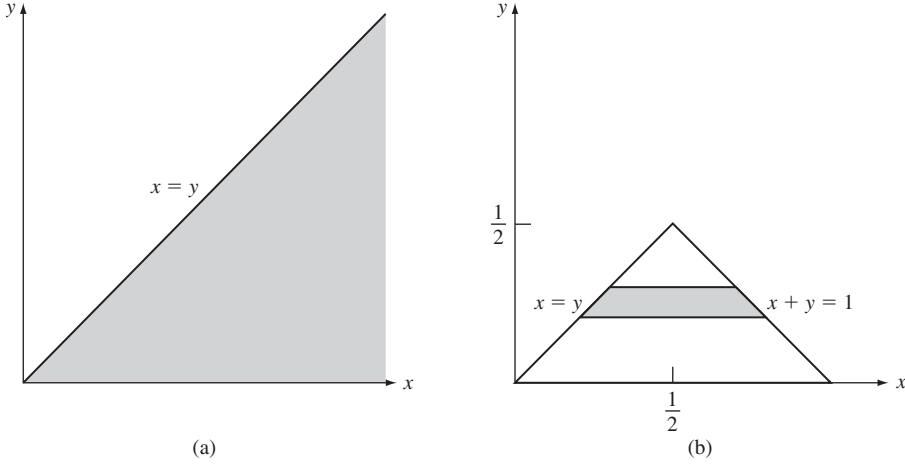
FIGURE 5.16
The random variables X and Y in Examples 5.16 and 5.17 have a pdf that is nonzero only in the shaded region shown in part (a).

Example 5.17

Find $P[X + Y \leq 1]$ in Example 5.16.

Figure 5.16(b) shows the intersection of the event $\{X + Y \leq 1\}$ and the region where the pdf is nonzero. We obtain the probability of the event by "adding" (actually integrating) infinitesimal rectangles of width dy as indicated in the figure:

$$P[X + Y \leq 1] = \int_0^{.5} \int_y^{1-y} 2e^{-x}e^{-y}\, dx\, dy = \int_0^{.5} 2e^{-y}[e^{-y} - e^{-(1-y)}]\, dy$$

$$= 1 - 2e^{-1}.$$

Example 5.18   Jointly Gaussian Random Variables

The joint pdf of X and Y, shown in Fig. 5.17, is

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1 - \rho^2}} e^{-(x^2 - 2\rho xy + y^2)/2(1-\rho^2)} \qquad -\infty < x, y < \infty. \tag{5.18}$$

We say that X and Y are jointly Gaussian.[1] Find the marginal pdf's.

The marginal pdf of X is found by integrating $f_{X,Y}(x, y)$ over y:

$$f_X(x) = \frac{e^{-x^2/2(1-\rho^2)}}{2\pi\sqrt{1 - \rho^2}} \int_{-\infty}^{\infty} e^{-(y^2 - 2\rho xy)/2(1-\rho^2)}\, dy.$$

[1]This is an important special case of jointly Gaussian random variables. The general case is discussed in Section 5.9.
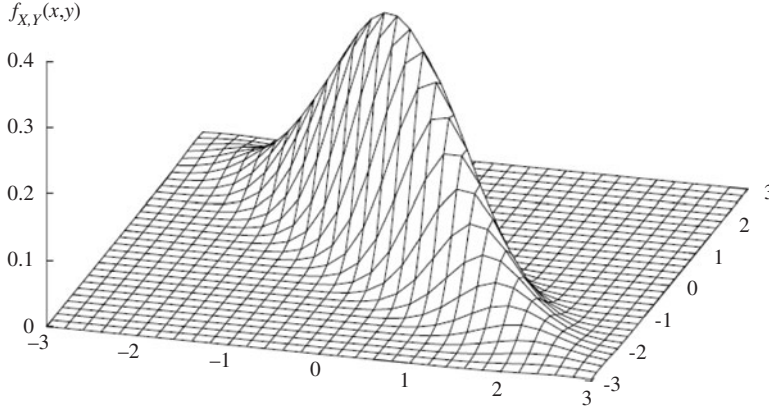
FIGURE 5.17
Joint pdf of two jointly Gaussian random variables.

We complete the square of the argument of the exponent by adding and subtracting $\rho^2 x^2$, that is, $y^2 - 2\rho xy + \rho^2 x^2 - \rho^2 x^2 = (y - \rho x)^2 - \rho^2 x^2$. Therefore

$$f_X(x) = \frac{e^{-x^2/2(1-\rho^2)}}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-[(y-\rho x)^2 - \rho^2 x^2]/2(1-\rho^2)} \, dy$$

$$= \frac{e^{-x^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-(y-\rho x)^2/2(1-\rho^2)}}{\sqrt{2\pi(1-\rho^2)}} \, dy$$

$$= \frac{e^{-x^2/2}}{\sqrt{2\pi}},$$

where we have noted that the last integral equals one since its integrand is a Gaussian pdf with mean ρx and variance $1 - \rho^2$. The marginal pdf of X is therefore a one-dimensional Gaussian pdf with mean 0 and variance 1. From the symmetry of $f_{X,Y}(x, y)$ in x and y, we conclude that the marginal pdf of Y is also a one-dimensional Gaussian pdf with zero mean and unit variance.

5.5 INDEPENDENCE OF TWO RANDOM VARIABLES

X and Y are independent random variables if *any* event A_1 defined in terms of X is independent of *any* event A_2 defined in terms of Y; that is,

$$P[X \text{ in } A_1, Y \text{ in } A_2] = P[X \text{ in } A_1]P[Y \text{ in } A_2]. \tag{5.19}$$

In this section we present a simple set of conditions for determining when X and Y are independent.

Suppose that X and Y are a pair of discrete random variables, and suppose we are interested in the probability of the event $A = A_1 \cap A_2$, where A_1 involves only X and A_2 involves only Y. In particular, if X and Y are independent, then A_1 and A_2 are independent events. If we let $A_1 = \{X = x_j\}$ and $A_2 = \{Y = y_k\}$, then the

independence of X and Y implies that

$$
\begin{aligned}
p_{X,Y}(x_j, y_k) &= P[X = x_j, Y = y_k] \\
&= P[X = x_j]P[Y = y_k] \\
&= p_X(x_j)p_Y(y_k) \qquad \text{for all } x_j \text{ and } y_k.
\end{aligned} \tag{5.20}
$$

Therefore, *if X and Y are independent discrete random variables, then the joint pmf is equal to the product of the marginal pmf's.*

Now suppose that we don't know if X and Y are independent, but we do know that the pmf satisfies Eq. (5.20). Let $A = A_1 \cap A_2$ be a product-form event as above, then

$$
\begin{aligned}
P[A] &= \sum_{x_j \text{ in } A_1} \sum_{y_k \text{ in } A_2} p_{X,Y}(x_j, y_k) \\
&= \sum_{x_j \text{ in } A_1} \sum_{y_k \text{ in } A_2} p_X(x_j)p_Y(y_k) \\
&= \sum_{x_j \text{ in } A_1} p_X(x_j) \sum_{y_k \text{ in } A_2} p_Y(y_k) \\
&= P[A_1]P[A_2],
\end{aligned} \tag{5.21}
$$

which implies that A_1 and A_2 are independent events. Therefore, *if the joint pmf of X and Y equals the product of the marginal pmf's, then X and Y are independent.* We have just proved that the statement "X and Y are independent" is equivalent to the statement "the joint pmf is equal to the product of the marginal pmf's." In mathematical language, we say, the "*discrete random variables X and Y are independent if and only if the joint pmf is equal to the product of the marginal pmf's for all x_j, y_k.*"

Example 5.19

Is the pmf in Example 5.6 consistent with an experiment that consists of the independent tosses of two fair dice?

The probability of each face in a toss of a fair die is 1/6. If two fair dice are tossed and if the tosses are independent, then the probability of any pair of faces, say *j* and *k,* is:

$$
P[X = j, Y = k] = P[X = j]P[Y = k] = \frac{1}{36}.
$$

Thus all possible pairs of outcomes should be equiprobable. This is not the case for the joint pmf given in Example 5.6. Therefore the tosses in Example 5.6 are not independent.

Example 5.20

Are Q and R in Example 5.9 independent? From Example 5.9 we have

$$
\begin{aligned}
P[Q = q]P[R = r] &= (1 - p^M)(p^M)^q \frac{(1 - p)}{1 - p^M} p^r \\
&= (1 - p)p^{Mq+r}
\end{aligned}
$$

$$= P[Q = q, R = r] \qquad \text{for all } q = 0, 1, \ldots$$
$$r = 0, \ldots, M - 1.$$

Therefore Q and R are independent.

In general, it can be shown that the random variables X and Y are independent if and only if their joint cdf is equal to the product of its marginal cdf's:

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \qquad \text{for all } x \text{ and } y. \tag{5.22}$$

Similarly, if X and Y are jointly continuous, then X and Y are independent if and only if their joint pdf is equal to the product of the marginal pdf's:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \qquad \text{for all } x \text{ and } y. \tag{5.23}$$

Equation (5.23) is obtained from Eq. (5.22) by differentiation. Conversely, Eq. (5.22) is obtained from Eq. (5.23) by integration.

Example 5.21

Are the random variables X and Y in Example 5.16 independent?

Note that $f_X(x)$ and $f_Y(y)$ are nonzero for all $x > 0$ and all $y > 0$. Hence $f_X(x)f_Y(y)$ is nonzero in the entire positive quadrant. However $f_{X,Y}(x, y)$ is nonzero only in the region $y < x$ inside the positive quadrant. Hence Eq. (5.23) does not hold for all x, y and the random variables are not independent. You should note that in this example the joint pdf appears to factor, but nevertheless it is not the product of the marginal pdf's.

Example 5.22

Are the random variables X and Y in Example 5.18 independent? The product of the marginal pdf's of X and Y in Example 5.18 is

$$f_X(x)f_Y(y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2} \qquad -\infty < x, y < \infty.$$

By comparing to Eq. (5.18) we see that the product of the marginals is equal to the joint pdf if and only if $\rho = 0$. Therefore the jointly Gaussian random variables X and Y are independent if and only if $\rho = 0$. We see in a later section that ρ is the *correlation coefficient* between X and Y.

Example 5.23

Are the random variables X and Y independent in Example 5.12? If we multiply the marginal cdf's found in Example 5.12 we find

$$F_X(x)F_Y(y) = (1 - e^{-\alpha x})(1 - e^{-\beta y}) = F_{X,Y}(x, y) \qquad \text{for all } x \text{ and } y.$$

Therefore Eq. (5.22) is satisfied so X and Y are independent.

If X and Y are independent random variables, then the random variables defined by any pair of functions $g(X)$ and $h(Y)$ are also independent. To show this, consider the

one-dimensional events A and B. Let A' be the set of all values of x such that if x is in A' then $g(x)$ is in A, and let B' be the set of all values of y such that if y is in B' then $h(y)$ is in B. (In Chapter 3 we called A' and B' the equivalent events of A and B.) Then

$$
\begin{aligned}
P[g(X) \text{ in } A, h(Y) \text{ in } B] &= P[X \text{ in } A', Y \text{ in } B'] \\
&= P[X \text{ in } A']P[Y \text{ in } B'] \\
&= P[g(X) \text{ in } A]P[h(Y) \text{ in } B]. \qquad (5.24)
\end{aligned}
$$

The first and third equalities follow from the fact that A and A' and B and B' are equivalent events. The second equality follows from the independence of X and Y. Thus $g(X)$ and $h(Y)$ are independent random variables.

5.6   JOINT MOMENTS AND EXPECTED VALUES OF A FUNCTION OF TWO RANDOM VARIABLES

The expected value of X identifies the center of mass of the distribution of X. The variance, which is defined as the expected value of $(X - m)^2$, provides a measure of the spread of the distribution. In the case of two random variables we are interested in how X and Y vary together. In particular, we are interested in whether the variation of X and Y are correlated. For example, if X increases does Y tend to increase or to decrease? The joint moments of X and Y, which are defined as expected values of functions of X and Y, provide this information.

5.6.1   Expected Value of a Function of Two Random Variables

The problem of finding the expected value of a function of two or more random variables is similar to that of finding the expected value of a function of a single random variable. It can be shown that the expected value of $Z = g(X, Y)$ can be found using the following expressions:

$$
E[Z] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) \, dx \, dy & X, Y \text{ jointly continuous} \\[2em] \displaystyle\sum_{i} \sum_{n} g(x_i, y_n) p_{X,Y}(x_i, y_n) & X, Y \text{ discrete.} \end{cases} \qquad (5.25)
$$

Example 5.24   Sum of Random Variables

Let $Z = X + Y$. Find $E[Z]$.

$$
\begin{aligned}
E[Z] &= E[X + Y] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x' + y') f_{X,Y}(x', y') \, dx' \, dy' \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x' f_{X,Y}(x', y') \, dy' \, dx' + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y' \, f_{X,Y}(x', y') \, dx' \, dy' \\
&= \int_{-\infty}^{\infty} x' f_X(x') \, dx' + \int_{-\infty}^{\infty} y' f_Y(y') \, dy' = E[X] + E[Y]. \qquad (5.26)
\end{aligned}
$$

Thus the expected value of the sum of two random variables is equal to the sum of the individual expected values. Note that X and Y need not be independent.

The result in Example 5.24 and a simple induction argument show that *the expected value of a sum of n random variables is equal to the sum of the expected values:*

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]. \tag{5.27}$$

Note that the random variables do not have to be independent.

Example 5.25   Product of Functions of Independent Random Variables

Suppose that X and Y are independent random variables, and let $g(X, Y) = g_1(X)g_2(Y)$. Find $E[g(X, Y)] = E[g_1(X)g_2(Y)]$.

$$E[g_1(X)g_2(Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x')g_2(y')f_X(x')f_Y(y') \, dx' \, dy'$$

$$= \left\{ \int_{-\infty}^{\infty} g_1(x')f_X(x') \, dx' \right\} \left\{ \int_{-\infty}^{\infty} g_2(y')f_Y(y') \, dy' \right\}$$

$$= E[g_1(X)]E[g_2(Y)].$$

5.6.2   Joint Moments, Correlation, and Covariance

The joint moments of two random variables X and Y summarize information about their joint behavior. The jkth **joint moment of X and Y** is defined by

$$E[X^j Y^k] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^j y^k f_{X,Y}(x, y) \, dx \, dy & X, Y \text{ jointly continuous} \\[2ex] \displaystyle\sum_i \sum_n x_i^j y_n^k p_{X,Y}(x_i, y_n) & X, Y \text{ discrete}. \end{cases} \tag{5.28}$$

If $j = 0$, we obtain the moments of Y, and if $k = 0$, we obtain the moments of X. In electrical engineering, it is customary to call the $j = 1 \ k = 1$ moment, $E[XY]$, the **correlation of X and Y**. If $E[XY] = 0$, then we say that **X and Y are orthogonal**.

The jkth **central moment of X and Y** is defined as the joint moment of the centered random variables, $X - E[X]$ and $Y - E[Y]$:

$$E[(X - E[X])^j (Y - E[Y])^k].$$

Note that $j = 2 \ k = 0$ gives VAR(X) and $j = 0 \ k = 2$ gives VAR(Y).

The **covariance of X and Y** is defined as the $j = k = 1$ central moment:

$$\text{COV}(X, Y) = E[(X - E[X])(Y - E[Y])]. \tag{5.29}$$

The following form for COV(X, Y) is sometimes more convenient to work with:

$$\text{COV}(X, Y) = E[XY - XE[Y] - YE[X] + E[X]E[Y]]$$

$$= E[XY] - 2E[X]E[Y] + E[X]E[Y]$$

$$= E[XY] - E[X]E[Y]. \tag{5.30}$$

Note that $\text{COV}(X, Y) = E[XY]$ if either of the random variables has mean zero.

Example 5.26   Covariance of Independent Random Variables

Let X and Y be independent random variables. Find their covariance.

$$\text{COV}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

$$= E[X - E[X]]E[Y - E[Y]]$$

$$= 0,$$

where the second equality follows from the fact that X and Y are independent, and the third equality follows from $E[X - E[X]] = E[X] - E[X] = 0$. Therefore *pairs of independent random variables have covariance zero.*

Let's see how the covariance measures the correlation between X and Y. The covariance measures the deviation from $m_X = E[X]$ and $m_Y = E[Y]$. If a positive value of $(X - m_X)$ tends to be accompanied by a positive values of $(Y - m_Y)$, and negative $(X - m_X)$ tend to be accompanied by negative $(Y - m_Y)$; then $(X - m_X)(Y - m_Y)$ will tend to be a positive value, and its expected value, $\text{COV}(X, Y)$, will be positive. This is the case for the scattergram in Fig. 5.3(d) where the observed points tend to cluster along a line of positive slope. On the other hand, if $(X - m_X)$ and $(Y - m_Y)$ tend to have opposite signs, then $\text{COV}(X, Y)$ will be negative. A scattergram for this case would have observation points cluster along a line of negative slope. Finally if $(X - m_X)$ and $(Y - m_Y)$ sometimes have the same sign and sometimes have opposite signs, then $\text{COV}(X, Y)$ will be close to zero. The three scattergrams in Figs. 5.3(a), (b), and (c) fall into this category.

Multiplying either X or Y by a large number will increase the covariance, so we need to normalize the covariance to measure the correlation in an absolute scale. The **correlation coefficient of X and Y** is defined by

$$\rho_{X,Y} = \frac{\text{COV}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y}, \tag{5.31}$$

where $\sigma_X = \sqrt{\text{VAR}(X)}$ and $\sigma_Y = \sqrt{\text{VAR}(Y)}$ are the standard deviations of X and Y, respectively.

The correlation coefficient is a number that is at most 1 in magnitude:

$$-1 \leq \rho_{X,Y} \leq 1. \tag{5.32}$$

To show Eq. (5.32), we begin with an inequality that results from the fact that the expected value of the square of a random variable is nonnegative:

$$0 \leq E\left\{\left(\frac{X - E[X]}{\sigma_X} \pm \frac{Y - E[Y]}{\sigma_Y}\right)^2\right\}$$

$$= 1 \pm 2\rho_{X,Y} + 1$$

$$= 2(1 \pm \rho_{X,Y}).$$

The last equation implies Eq. (5.32).

The extreme values of $\rho_{X,Y}$ are achieved when X and Y are related linearly, $Y = aX + b$; $\rho_{X,Y} = 1$ if $a > 0$ and $\rho_{X,Y} = -1$ if $a < 0$. In Section 6.5 we show that $\rho_{X,Y}$ can be viewed as a statistical measure of the extent to which Y can be predicted by a linear function of X.

X and Y are said to be **uncorrelated** if $\rho_{X,Y} = 0$. If X and Y are independent, then $\text{COV}(X, Y) = 0$, so $\rho_{X,Y} = 0$. Thus *if X and Y are independent, then X and Y are uncorrelated.* In Example 5.22, we saw that if X and Y are jointly Gaussian and $\rho_{X,Y} = 0$, *then X and Y are independent random variables.* Example 5.27 shows that this is not always true for non-Gaussian random variables: It is possible for X and Y to be uncorrelated but not independent.

Example 5.27   Uncorrelated but Dependent Random Variables

Let Θ be uniformly distributed in the interval $(0, 2\pi)$. Let

$$X = \cos \Theta \qquad \text{and} \qquad Y = \sin \Theta.$$

The point (X, Y) then corresponds to the point on the unit circle specified by the angle Θ, as shown in Fig. 5.18. In Example 4.36, we saw that the marginal pdf's of X and Y are arcsine pdf's, which are nonzero in the interval $(-1, 1)$. The product of the marginals is nonzero in the square defined by $-1 \le x \le 1$ and $-1 \le y \le 1$, so if X and Y were independent the point (X, Y) would assume all values in this square. This is not the case, so X and Y are dependent.

We now show that X and Y are uncorrelated:

$$E[XY] = E[\sin \Theta \cos \Theta] = \frac{1}{2\pi} \int_0^{2\pi} \sin \phi \cos \phi \, d\phi$$

$$= \frac{1}{4\pi} \int_0^{2\pi} \sin 2\phi \, d\phi = 0.$$

Since $E[X] = E[Y] = 0$, Eq. (5.30) then implies that X and Y are uncorrelated.

Example 5.28

Let X and Y be the random variables discussed in Example 5.16. Find $E[XY]$, $\text{COV}(X, Y)$, and $\rho_{X,Y}$.

Equations (5.30) and (5.31) require that we find the mean, variance, and correlation of X and Y. From the marginal pdf's of X and Y obtained in Example 5.16, we find that $E[X] = 3/2$ and $\text{VAR}[X] = 5/4$, and that $E[Y] = 1/2$ and $\text{VAR}[Y] = 1/4$. The correlation of X and Y is

$$E[XY] = \int_0^\infty \int_0^x xy2e^{-x}e^{-y} \, dy \, dx$$

$$= \int_0^\infty 2xe^{-x}(1 - e^{-x} - xe^{-x}) \, dx = 1.$$
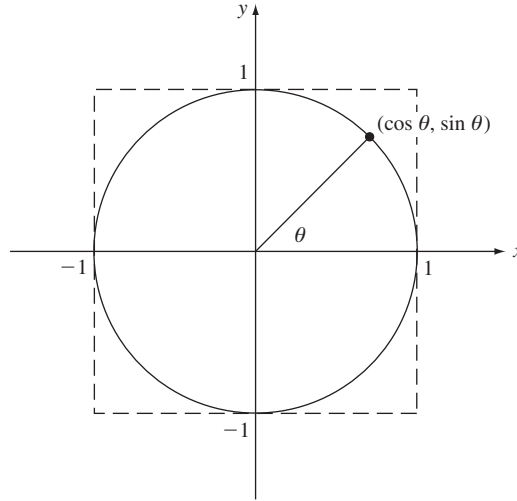
FIGURE 5.18
(X, Y) is a point selected at random *on* the unit circle. X and Y
are uncorrelated but not independent.

Thus the correlation coefficient is given by

$$\rho_{X,Y} = \frac{1 - \dfrac{3}{2}\dfrac{1}{2}}{\sqrt{\dfrac{5}{4}}\sqrt{\dfrac{1}{4}}} = \frac{1}{\sqrt{5}}.$$

5.7   CONDITIONAL PROBABILITY AND CONDITIONAL EXPECTATION

Many random variables of practical interest are not independent: The output Y of a communication channel must depend on the input X in order to convey information; consecutive samples of a waveform that varies slowly are likely to be close in value and hence are not independent. In this section we are interested in computing the probability of events concerning the random variable Y given that we know $X = x$. We are also interested in the expected value of Y given $X = x$. We show that the notions of conditional probability and conditional expectation are extremely useful tools in solving problems, even in situations where we are only concerned with one of the random variables.

5.7.1   Conditional Probability

The definition of conditional probability in Section 2.4 allows us to compute the probability that Y is in A given that we know that $X = x$:

$$P[Y \text{ in } A \,|\, X = x] = \frac{P[Y \text{ in } A, X = x]}{P[X = x]} \quad \text{for } P[X = x] > 0. \tag{5.33}$$

Case 1: *X* Is a Discrete Random Variable

For X and Y discrete random variables, the **conditional pmf of Y given X = x** is defined by:

$$p_Y(y\,|\,x) = P[Y = y\,|\,X = x] = \frac{P[X = x, Y = y]}{P[X = x]} = \frac{p_{X,Y}(x, y)}{p_X(x)} \qquad (5.34)$$

for *x* such that $P[X = x] > 0$. We define $p_Y(y\,|\,x) = 0$ for *x* such that $P[X = x] = 0$. Note that $p_Y(y\,|\,x)$ is a function of *y* over the real line, and that $p_Y(y\,|\,x) > 0$ only for *y* in a discrete set $\{y_1, y_2, \dots\}$.

The conditional pmf satisfies all the properties of a pmf, that is, it assigns non-negative values to every *y* and these values add to 1. Note from Eq. (5.34) that $p_Y(y\,|\,x_k)$ is simply the cross section of $p_{X,Y}(x_k, y)$ along the $X = x_k$ column in Fig. 5.6, but normalized by the probability $p_X(x_k)$.

The probability of an event *A* given $X = x_k$ is found by adding the pmf values of the outcomes in *A*:

$$P[Y \text{ in } A\,|\,X = x_k] = \sum_{y_j \text{in } A} p_Y(y_j\,|\,x_k). \qquad (5.35)$$

If *X* and *Y* are independent, then using Eq (5.20)

$$p_Y(y_j\,|\,x_k) = \frac{P[X = x_k, Y = y_j]}{P[X = x_k]} = P[Y = y_j] = p_Y(y_j). \qquad (5.36)$$

In other words, knowledge that $X = x_k$ does not affect the probability of events *A* involving *Y*.

Equation (5.34) implies that the joint pmf $p_{X,Y}(x, y)$ can be expressed as the product of a conditional pmf and a marginal pmf:

$$p_{X,Y}(x_k, y_j) = p_Y(y_j\,|\,x_k)p_X(x_k) \text{ and } p_{X,Y}(x_k, y_j) = p_X(x_k\,|\,y_j)p_Y(y_j). \qquad (5.37)$$

This expression is very useful when we can view the pair (X, Y) as being generated sequentially, e.g., first *X*, and then *Y* given $X = x$. We find the probability that *Y* is in *A* as follows:

$$P[Y \text{ in } A] = \sum_{\text{all } x_k} \sum_{y_j \text{ in } A} p_{X,Y}(x_k, y_j)$$

$$= \sum_{\text{all } x_k} \sum_{y_j \text{ in } A} p_Y(y_j\,|\,x_k)p_X(x_k)$$

$$= \sum_{\text{all } x_k} p_X(x_k) \sum_{y_j \text{ in } A} p_Y(y_j\,|\,x_k)$$

$$= \sum_{\text{all } x_k} P[Y \text{ in } A\,|\,X = x_k]p_X(x_k). \qquad (5.38)$$

Equation (5.38) is simply a restatement of the theorem on total probability discussed in Chapter 2. In other words, to compute $P[Y \text{ in } A]$ we can first compute $P[Y \text{ in } A\,|\,X = x_k]$ and then "average" over X_k.

Example 5.29   Loaded Dice

Find $p_Y(y|5)$ in the loaded dice experiment considered in Examples 5.6 and 5.8.

In Example 5.8 we found that $p_X(5) = 1/6$. Therefore:

$$p_Y(y|5) = \frac{p_{X,Y}(5, y)}{p_X(5)} \text{ and so } p_Y(5|5) = 2/7 \text{ and}$$

$$p_Y(1|5) = p_Y(2|5) = p_Y(3|5) = p_Y(4|5) = p_Y(6|5) = 1/7.$$

Clearly this die is loaded.

Example 5.30   Number of Defects in a Region; Random Splitting of Poisson Counts

The total number of defects X on a chip is a Poisson random variable with mean α. Each defect has a probability p of falling in a specific region R and the location of each defect is independent of the locations of other defects. Find the pmf of the number of defects Y that fall in the region R.

We can imagine performing a Bernoulli trial each time a defect occurs with a "success" occurring when the defect falls in the region R. If the total number of defects is $X = k$, then Y is a binomial random variable with parameters k and p:

$$p_Y(j|k) = \begin{cases} 0 & j > k \\ \binom{k}{j} p^j (1 - p)^{k-j} & 0 \le j \le k. \end{cases}$$

From Eq. (5.38) and noting that $k \ge j$, we have

$$p_Y(j) = \sum_{k=0}^{\infty} p_Y(j|k) p_X(k) = \sum_{k=j}^{\infty} \frac{k!}{j!(k-j)!} p^j (1-p)^{k-j} \frac{\alpha^k}{k!} e^{-\alpha}$$

$$= \frac{(\alpha p)^j e^{-\alpha}}{j!} \sum_{k=j}^{\infty} \frac{\{(1-p)\alpha\}^{k-j}}{(k-j)!}$$

$$= \frac{(\alpha p)^j e^{-\alpha}}{j!} e^{(1-p)\alpha} = \frac{(\alpha p)^j}{j!} e^{-\alpha p}.$$

Thus Y is a Poisson random variable with mean αp.

Suppose Y is a continuous random variable. Eq. (5.33) can be used to define the **conditional cdf of Y given $X = x_k$:**

$$F_Y(y|x_k) = \frac{P[Y \le y, X = x_k]}{P[X = x_k]}, \qquad \text{for } P[X = x_k] > 0. \tag{5.39}$$

It is easy to show that $F_Y(y|x_k)$ satisfies all the properties of a cdf. The **conditional pdf of Y given $X = x_k$**, if the derivative exists, is given by

$$f_Y(y|x_k) = \frac{d}{dy} F_Y(y|x_k). \tag{5.40}$$

If X and Y are independent, $P[Y \leq y, X = X_k] = P[Y \leq y]P[X = X_k]$ so $F_Y(y \mid x) = F_Y(y)$ and $f_Y(y \mid x) = f_Y(y)$. The probability of event A given $X = x_k$ is obtained by integrating the conditional pdf:

$$P[Y \text{ in } A \mid X = x_k] = \int_{y \text{ in } A} f_Y(y \mid x_k) \, dy. \tag{5.41}$$

We obtain $P[Y \text{ in } A]$ using Eq. (5.38).

Example 5.31   Binary Communications System

The input X to a communication channel assumes the values +1 or −1 with probabilities 1/3 and 2/3. The output Y of the channel is given by $Y = X + N$, where N is a zero-mean, unit variance Gaussian random variable. Find the conditional pdf of Y given $X = +1$, and given $X = -1$. Find $P[X = +1 \mid Y > 0]$.

The conditional cdf of Y given $X = +1$ is:

$$F_Y(y \mid +1) = P[Y \leq y \mid X = +1] = P[N + 1 \leq y]$$

$$= P[N \leq y - 1] = \int_{-\infty}^{y-1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx$$

where we noted that if $X = +1$, then $Y = N + 1$ and Y depends only on N. Thus, if $X = +1$, then Y is a Gaussian random variable with mean 1 and unit variance. Similarly, if $X = -1$, then Y is Gaussian with mean −1 and unit variance.

The probabilities that $Y > 0$ given $X = +1$ and $X = -1$ is:

$$P[Y > 0 \mid X = +1] = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-(x-1)^2/2} \, dx = \int_{-1}^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt = 1 - Q(1) = 0.841.$$

$$P[Y > 0 \mid X = -1] = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-(x+1)^2/2} \, dx = \int_1^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt = Q(1) = 0.159.$$

Applying Eq. (5.38), we obtain:

$$P[Y > 0] = P[Y > 0 \mid X = +1]\frac{1}{3} + P[Y > 0 \mid X = -1]\frac{2}{3} = 0.386.$$

From Bayes' theorem we find:

$$P[X = +1 \mid Y > 0] = \frac{P[Y > 0 \mid X = +1]P[X = +1]}{P[Y > 0]} = \frac{(1 - Q(1))/3}{(1 + Q(1))/3} = 0.726.$$

We conclude that if $Y > 0$, then $X = +1$ is more likely than $X = -1$. Therefore the receiver should decide that the input is $X = +1$ when it observes $Y > 0$.

In the previous example, we made an interesting step that is worth elaborating on because it comes up quite frequently: $P[Y \leq y \mid X = +1] = P[N + 1 \leq y]$, where $Y = X + N$. Let's take a closer look:

$$P[Y \le z \,|\, X = x] = \frac{P[\{X + N \le z\} \cap \{X = x\}]}{P[X = x]} = \frac{P[\{x + N \le z\} \cap \{X = x\}]}{P[X = x]}$$

$$= P[x + N \le z \,|\, X = x] = P[N \le z - x \,|\, X = x].$$

In the first line, the events $\{X + N \le z\}$ and $\{x + N \le z\}$ are quite different. The first involves the two random variables X and N, whereas the second only involves N and consequently is much simpler. We can then apply an expression such as Eq. (5.38) to obtain $P[Y \le z]$. The step we made in the example, however, is even more interesting. Since X and N are independent random variables, we can take the expression one step further:

$$P[Y \le z \,|\, X = x] = P[N \le z - x \,|\, X = x] = P[N \le z - x].$$

The independence of X and N allows us to dispense with the conditioning on x altogether!

Case 2: *X* Is a Continuous Random Variable

If X is a continuous random variable, then $P[X = x] = 0$ so Eq. (5.33) is undefined for all x. If X and Y have a joint pdf that is continuous and nonzero over some region of the plane, we define the **conditional cdf of Y given $X = x$** by the following limiting procedure:

$$F_Y(y \,|\, x) = \lim_{h \to 0} F_Y(y \,|\, x < X \le x + h). \tag{5.42}$$

The conditional cdf on the right side of Eq. (5.42) is:

$$F_Y(y \,|\, x < X \le x + h) = \frac{P[Y \le y, x < X \le x + h]}{P[x < X \le x + h]}$$

$$= \frac{\displaystyle\int_{-\infty}^{y} \int_{x}^{x+h} f_{X,Y}(x', y') \, dx' \, dy'}{\displaystyle\int_{x}^{x+h} f_X(x') \, dx'} = \frac{\displaystyle\int_{-\infty}^{y} f_{X,Y}(x, y') \, dy' h}{f_X(x) h}. \tag{5.43}$$

As we let h approach zero, Eqs. (5.42) and (5.43) imply that

$$F_Y(y \,|\, x) = \frac{\displaystyle\int_{-\infty}^{y} f_{X,Y}(x, y') \, dy'}{f_X(x)}. \tag{5.44}$$

The **conditional pdf of Y given $X = x$** is then:

$$f_Y(y \,|\, x) = \frac{d}{dy} F_Y(y \,|\, x) = \frac{f_{X,Y}(x, y)}{f_X(x)}. \tag{5.45}$$
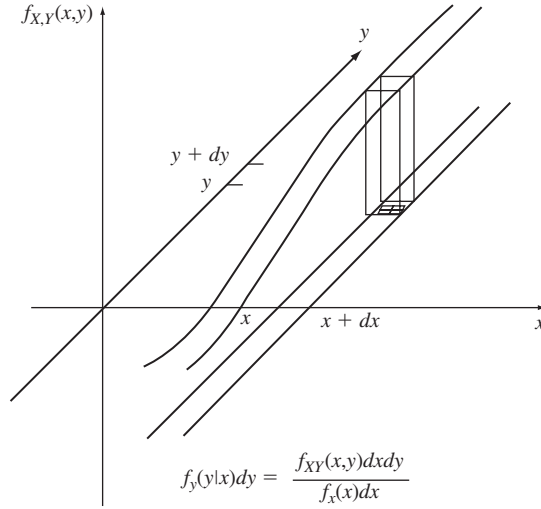
FIGURE 5.19
Interpretation of conditional pdf.

It is easy to show that $f_Y(y \mid x)$ satisfies the properties of a pdf. We can interpret $f_Y(y \mid x)\,dy$ as the probability that Y is in the infinitesimal strip defined by $(y, y + dy)$ given that X is in the infinitesimal strip defined by $(x, x + dx)$, as shown in Fig. 5.19.

The probability of event A given $X = x$ is obtained as follows:

$$P[Y \text{ in } A \mid X = x] = \int_{y \text{ in } A} f_Y(y \mid x)\,dy. \tag{5.46}$$

There is a strong resemblance between Eq. (5.34) for the discrete case and Eq. (5.45) for the continuous case. Indeed many of the same properties hold. For example, we obtain the multiplication rule from Eq. (5.45):

$$f_{X,Y}(x, y) = f_Y(y \mid x)f_X(x) \text{ and } f_{X,Y}(x, y) = f_X(x \mid y)f_Y(y). \tag{5.47}$$

If X and Y are independent, then $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ and $f_Y(y \mid x) = f_Y(y)$, $f_X(x \mid y) = f_X(x)$, $F_Y(y \mid x) = F_Y(y)$, and $F_X(x \mid y) = F_X(x)$.

By combining Eqs. (5.46) and (5.47), we can show that:

$$P[Y \text{ in } A] = \int_{-\infty}^{\infty} P[Y \text{ in } A \mid X = x]f_X(x)\,dx. \tag{5.48}$$

You can think of Eq. (5.48) as the "continuous" version of the theorem on total probability. The following examples show the usefulness of the above results in calculating the probabilities of complicated events.

Example 5.32

Let X and Y be the random variables in Example 5.8. Find $f_X(x \mid y)$ and $f_Y(y \mid x)$.

Using the marginal pdf's obtained in Example 5.8, we have

$$f_X(y \mid x) = \frac{2e^{-x}e^{-y}}{2e^{-2y}} = e^{-(x-y)} \qquad \text{for } x \geq y$$

$$f_Y(y \mid x) = \frac{2e^{-x}e^{-y}}{2e^{-x}(1 - e^{-x})} = \frac{e^{-y}}{1 - e^{-x}} \qquad \text{for } 0 < y < x.$$

The conditional pdf of X is an exponential pdf shifted by y to the right. The conditional pdf of Y is an exponential pdf that has been truncated to the interval $[0, x]$.

Example 5.33   Number of Arrivals During a Customer's Service Time

The number N of customers that arrive at a service station during a time t is a Poisson random variable with parameter βt. The time T required to service each customer is an exponential random variable with parameter α. Find the pmf for the number N that arrive during the service time T of a specific customer. Assume that the customer arrivals are independent of the customer service time.

Equation (5.48) holds even if Y is a discrete random variable, thus

$$P[N = k] = \int_0^\infty P[N = k \mid T = t] f_T(t) \, dt$$

$$= \int_0^\infty \frac{(\beta t)^k}{k!} e^{-\beta t} \alpha e^{-\alpha t} \, dt$$

$$= \frac{\alpha \beta^k}{k!} \int_0^\infty t^k e^{-(\alpha + \beta)t} \, dt.$$

Let $r = (\alpha + \beta)t$, then

$$P[N = k] = \frac{\alpha \beta^k}{k!(\alpha + \beta)^{k+1}} \int_0^\infty r^k e^{-r} \, dr$$

$$= \frac{\alpha \beta^k}{(\alpha + \beta)^{k+1}} = \left(\frac{\alpha}{(\alpha + \beta)} \right) \left(\frac{\beta}{(\alpha + \beta)} \right)^k,$$

where we have used the fact that the last integral is a gamma function and is equal to $k!$. Thus N is a geometric random variable with probability of "success" $\alpha/(\alpha + \beta)$. Each time a customer arrives we can imagine that a new Bernoulli trial begins where "success" occurs if the customer's service time is completed before the next arrival.

Example 5.34

X is selected at random from the unit interval; Y is then selected at random from the interval $(0, X)$. Find the cdf of Y.

When $X = x$, Y is uniformly distributed in $(0, x)$ so the conditional cdf given $X = x$ is

$$P[Y \leq y \mid X = k] = \begin{cases} y/x & 0 \leq y \leq x \\ 1 & x < y. \end{cases}$$

Equation (5.48) and the above conditional cdf yield:

$$F_Y(y) = P[Y \leq y] = \int_0^1 P[Y \leq y \mid X = x] f_X(x) \, dx =$$

$$= \int_0^y 1 \, dx' + \int_y^1 \frac{y}{x'} \, dx' = y - y \ln y.$$

The corresponding pdf is obtained by taking the derivative of the cdf:

$$f_Y(y) = -\ln y \quad 0 \leq y \leq 1.$$

Example 5.35   Maximum A Posteriori Receiver

For the communications system in Example 5.31, find the probability that the input was $X = +1$ given that the output of the channel is $Y = y$.

This is a tricky version of Bayes' rule. Condition on the event $\{y < Y \leq y + \Delta\}$ instead of $\{Y = y\}$:

$$P[X = +1 \mid y < Y < y + \Delta] = \frac{P[y < Y < y + \Delta \mid X = +1] P[X = +1]}{P[y < Y < y + \Delta]}$$

$$= \frac{f_Y(y \mid +1) \Delta (1/3)}{f_Y(y \mid +1) \Delta (1/3) + f_Y(y \mid -1) \Delta (2/3)}$$

$$= \frac{\dfrac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2} (1/3)}{\dfrac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2} (1/3) + \dfrac{1}{\sqrt{2\pi}} e^{-(y+1)^2/2} (2/3)}$$

$$= \frac{e^{-(y-1)^2/2}}{e^{-(y-1)^2/2} + 2e^{-(y+1)^2/2}} = \frac{1}{1 + 2e^{-2y}}.$$

The above expression is equal to 1/2 when $y_T = 0.3466$. For $y > y_T$, $X = +1$ is more likely, and for $y < y_T$, $X = -1$ is more likely. A receiver that selects the input X that is more likely given $Y = y$ is called a *maximum a posteriori receiver*.

5.7.2   Conditional Expectation

The **conditional expectation of Y given $X = x$** is defined by

$$E[Y \mid x] = \int_{-\infty}^{\infty} y f_Y(y \mid x) \, dy. \tag{5.49a}$$

In the special case where X and Y are both discrete random variables we have:

$$E[Y \,|\, x_k] = \sum_{y_j} y_j p_Y(y_j \,|\, x_k). \tag{5.49b}$$

Clearly, $E[Y \,|\, x]$ is simply the center of mass associated with the conditional pdf or pmf.

The conditional expectation $E[Y \,|\, x]$ can be viewed as defining a function of x: $g(x) = E[Y \,|\, x]$. It therefore makes sense to talk about the random variable $g(X) = E[Y \,|\, X]$. We can imagine that a random experiment is performed and a value for X is obtained, say $X = x_0$, and then the value $g(x_0) = E[Y \,|\, x_0]$ is produced. We are interested in $E[g(X)] = E[E[Y \,|\, X]]$. In particular, we now show that

$$E[Y] = E[E[Y \,|\, X]], \tag{5.50}$$

where the right-hand side is

$$E[E[Y \,|\, X]] = \int_{-\infty}^{\infty} E[Y \,|\, x] f_X(x)\, dx \qquad X \text{ continuous} \tag{5.51a}$$

$$E[E[Y \,|\, X]] = \sum_{x_k} E[Y \,|\, x_k] p_X(x_k) \qquad X \text{ discrete.} \tag{5.51b}$$

We prove Eq. (5.50) for the case where X and Y are jointly continuous random variables, then

$$
\begin{aligned}
E[E[Y \,|\, X]] &= \int_{-\infty}^{\infty} E[Y \,|\, x] f_X(x)\, dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_Y(y \,|\, x)\, dy\, f_X(x)\, dx \\
&= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} y f_Y(y)\, dy = E[Y].
\end{aligned}
$$

The above result also holds for the expected value of a function of Y:

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

In particular, the kth moment of Y is given by

$$E[Y^k] = E[E[Y^k \,|\, X]].$$

Example 5.36   Average Number of Defects in a Region

Find the mean of Y in Example 5.30 using conditional expectation.

$$E[Y] = \sum_{k=0}^{\infty} E[Y \,|\, X = k] P[X = k] = \sum_{k=0}^{\infty} k p P[X = k] = p E[X] = p\alpha.$$

The second equality uses the fact that $E[Y \mid X = k] = kp$ since Y is binomial with parameters k and p. Note that the second to the last equality holds for *any* pmf of X. The fact that X is Poisson with mean α is not used until the last equality.

Example 5.37   Binary Communications Channel

Find the mean of the output Y in the communications channel in Example 5.31.

Since Y is a Gaussian random variable with mean $+1$ when $X = +1$, and -1 when $X = -1$, the conditional expected values of Y given X are:

$$E[Y \mid +1] = 1 \quad \text{and} \quad E[Y \mid -1] = -1.$$

Equation (5.38b) implies

$$E[Y] = \sum_{k=0}^{\infty} E[Y \mid X = k]P[X = k] = +1(1/3) - 1(2/3) = -1/3.$$

The mean is negative because the $X = -1$ inputs occur twice as often as $X = +1$.

Example 5.38   Average Number of Arrivals in a Service Time

Find the mean and variance of the number of customer arrivals N during the service time T of a specific customer in Example (5.33).

N is a Poisson random variable with parameter βt when $T = t$ is given, so the first two conditional moments are:

$$E[N \mid T = t] = \beta t \qquad E[N^2 \mid T = t] = (\beta t) + (\beta t)^2.$$

The first two moments of N are obtained from Eq. (5.50):

$$E[N] = \int_0^{\infty} E[N \mid T = t]f_T(t)\, dt = \int_0^{\infty} \beta t f_T(t)\, dt = \beta E[T]$$

$$E[N^2] = \int_0^{\infty} E[N^2 \mid T = t]f_T(t)\, dt = \int_0^{\infty} \{\beta t + \beta^2 t^2\}f_T(t)\, dt$$

$$= \beta E[T] + \beta^2 E[T^2].$$

The variance of N is then

$$\begin{aligned}
\text{VAR}[N] &= E[N^2] - (E[N])^2 \\
&= \beta^2 E[T^2] + \beta E[T] - \beta^2(E[T])^2 \\
&= \beta^2\, \text{VAR}[T] + \beta E[T].
\end{aligned}$$

Note that if T is not random (i.e., $E[T] = $ constant and $\text{VAR}[T] = 0$) then the mean and variance of N are those of a Poisson random variable with parameter $\beta E[T]$. When T is random, the mean of N remains the same but the variance of N increases by the term $\beta^2\, \text{VAR}[T]$, that is, the variability of T causes greater variability in N. Up to this point, we have intentionally avoided using the fact that T has an exponential distribution to emphasize that the above results hold

for *any* service time distribution $f_T(t)$. If *T is exponential with parameter* α, then $E[T] = 1/\alpha$ and $\text{VAR}[T] = 1/\alpha^2$, so

$$E[N] = \frac{\beta}{\alpha} \quad \text{and} \quad \text{VAR}[N] = \frac{\beta^2}{\alpha^2} + \frac{\beta}{\alpha}.$$

5.8   FUNCTIONS OF TWO RANDOM VARIABLES

Quite often we are interested in one or more functions of the random variables associated with some experiment. For example, if we make repeated measurements of the same random quantity, we might be interested in the maximum and minimum value in the set, as well as the sample mean and sample variance. In this section we present methods of determining the probabilities of events involving functions of two random variables.

5.8.1   One Function of Two Random Variables

Let the random variable Z be defined as a function of two random variables:

$$Z = g(X, Y). \tag{5.52}$$

The cdf of Z is found by first finding the equivalent event of $\{Z \leq z\}$, that is, the set $R_z = \{\mathbf{x} = (x, y) \text{ such that } g(\mathbf{x}) \leq z\}$, then

$$F_z(z) = P[\mathbf{X} \text{ in } R_z] = \iint\limits_{(x, y) \in R_z} f_{X,Y}(x', y') \, dx' \, dy'. \tag{5.53}$$

The pdf of Z is then found by taking the derivative of $F_z(z)$.

Example 5.39   Sum of Two Random Variables

Let $Z = X + Y$. Find $F_Z(z)$ and $f_Z(z)$ in terms of the joint pdf of X and Y.

The cdf of Z is found by integrating the joint pdf of X and Y over the region of the plane corresponding to the event $\{Z \leq z\}$, as shown in Fig. 5.20.
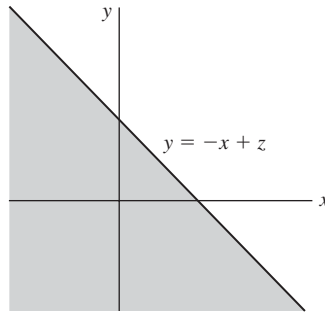
FIGURE 5.20
$P[Z \leq z] = P[X + Y \leq z]$.

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x'} f_{X,Y}(x', y')\, dy'\, dx'.$$

The pdf of Z is

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x', z - x')\, dx'. \tag{5.54}$$

Thus the pdf for the sum of two random variables is given by a *superposition* integral.

If X and Y are independent random variables, then by Eq. (5.23) the pdf is given by the *convolution integral* of the marginal pdf's of X and Y:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x') f_Y(z - x')\, dx'. \tag{5.55}$$

In Chapter 7 we show how transform methods are used to evaluate convolution integrals such as Eq. (5.55).

Example 5.40   Sum of Nonindependent Gaussian Random Variables

Find the pdf of the sum $Z = X + Y$ of two zero-mean, unit-variance Gaussian random variables with correlation coefficient $\rho = -1/2$.

The joint pdf for this pair of random variables was given in Example 5.18. The pdf of Z is obtained by substituting the pdf for the joint Gaussian random variables into the superposition integral found in Example 5.39:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{X,Y}(x', z - x')\, dx' \\[2mm] &= \frac{1}{2\pi(1 - \rho^2)^{1/2}} \int_{-\infty}^{\infty} e^{-[x'^2 - 2\rho x'(z-x') + (z-x')^2]/2(1-\rho^2)}\, dx' \\[2mm] &= \frac{1}{2\pi(3/4)^{1/2}} \int_{-\infty}^{\infty} e^{-(x'^2 - x'z + z^2)/2(3/4)}\, dx'. \end{aligned}$$

After completing the square of the argument in the exponent we obtain

$$f_Z(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}.$$

Thus the sum of these two nonindependent Gaussian random variables is also a zero-mean, unit-variance Gaussian random variable.

Example 5.41   A System with Standby Redundancy

A system with standby redundancy has a single key component in operation and a duplicate of that component in standby mode. When the first component fails, the second component is put into operation. Find the pdf of the lifetime of the standby system if the components have independent exponentially distributed lifetimes with the same mean.

Let T_1 and T_2 be the lifetimes of the two components, then the system lifetime is $T = T_1 + T_2$, and the pdf of T is given by Eq. (5.55). The terms in the integrand are

$$f_{T_1}(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$f_{T_2}(z - x) = \begin{cases} \lambda e^{-\lambda(z-x)} & z - x \geq 0 \\ 0 & x > z. \end{cases}$$

Note that the first equation sets the lower limit of integration to 0 and the second equation sets the upper limit to z. Equation (5.55) becomes

$$f_T(z) = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)} \, dx$$

$$= \lambda^2 e^{-\lambda z} \int_0^z dx = \lambda^2 z e^{-\lambda z}.$$

Thus T is an Erlang random variable with parameter $m = 2$.

The conditional pdf can be used to find the pdf of a function of several random variables. Let $Z = g(X, Y)$, and suppose we are given that $Y = y$, then $Z = g(X, y)$ is a function of one random variable. Therefore we can use the methods developed in Section 4.5 for single random variables to find the pdf of Z *given $Y = y$*: $f_Z(z \mid Y = y)$. The pdf of Z is then found from

$$f_Z(z) = \int_{-\infty}^{\infty} f_Z(z \mid y') f_Y(y') \, dy'.$$

Example 5.42

Let $Z = X/Y$. Find the pdf of Z if X and Y are independent and both exponentially distributed with mean one.

Assume $Y = y$, then $Z = X/y$ is simply a scaled version of X. Therefore from Example 4.31

$$f_Z(z \mid y) = |y| f_X(yz \mid y).$$

The pdf of Z is therefore

$$f_Z(z) = \int_{-\infty}^{\infty} |y'| f_X(y'z \mid y') f_Y(y') \, dy' = \int_{-\infty}^{\infty} |y'| f_{X,Y}(y'z, y') \, dy'.$$

We now use the fact that X and Y are independent and exponentially distributed with mean one:

$$f_Z(z) = \int_0^{\infty} y' f_X(y'z) f_Y(y') \, dy' \qquad z > 0$$

$$= \int_0^{\infty} y' e^{-y'z} e^{-y'} \, dy'$$

$$= \frac{1}{(1 + z)^2} \qquad z > 0.$$

5.8.2 Transformations of Two Random Variables

Let X and Y be random variables associated with some experiment, and let the random variables Z_1 and Z_2 be defined by two functions of $\mathbf{X} = (X, Y)$:

$$Z_1 = g_1(\mathbf{X}) \quad \text{and} \quad Z_2 = g_2(\mathbf{X}).$$

We now consider the problem of finding the joint cdf and pdf of Z_1 and Z_2.

The joint cdf of Z_1 and Z_2 at the point $\mathbf{z} = (z_1, z_2)$ is equal to the probability of the region of $\mathbf{x}$ where $g_k(\mathbf{x}) \le z_k$ for $k = 1, 2$:

$$F_{z_1, z_2}(z_1, z_2) = P[g_1(\mathbf{X}) \le z_1, g_2(\mathbf{X}) \le z_2]. \tag{5.56a}$$

If X, Y have a joint pdf, then

$$F_{z_1, z_2}(z_1, z_2) = \iint\limits_{\mathbf{x'}:\, \mathbf{g_k(x')} \le \mathbf{z_k}} f_{X,Y}(x', y')\, dx'\, dy'. \tag{5.56b}$$

Example 5.43

Let the random variables W and Z be defined by

$$W = \min(X, Y) \quad \text{and} \quad Z = \max(X, Y).$$

Find the joint cdf of W and Z in terms of the joint cdf of X and Y.

Equation (5.56a) implies that

$$F_{W, Z}(w\ z) = P[\{\min(X, Y) \le w\} \cap \{\max(X, Y) \le z\}].$$

The region corresponding to this event is shown in Fig. 5.21. From the figure it is clear that if $z > w$, the above probability is the probability of the semi-infinite rectangle defined by the
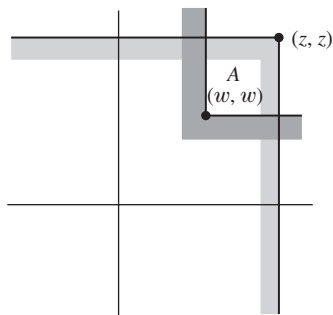
FIGURE 5.21
$\{\min(X, Y) \le w = \{X \le w\} \cup \{Y \le w\}$ and
$\{\max(X, Y) \le z = \{X \le z\} \cap \{Y \le z\}$.

point (z, z) minus the square region denoted by A. Thus if $z > w$,

$$F_{W, Z}(w, z) = F_{X,Y}(z, z) - P[A]$$

$$= F_{X,Y}(z, z)$$

$$- \{F_{X,Y}(z, z) - F_{X,Y}(w, z) - F_{X,Y}(z, w) + F_{X,Y}(w, w)\}$$

$$= F_{X,Y}(w, z) + F_{X,Y}(z, w) - F_{X,Y}(w, w).$$

If $z < w$ then

$$F_{W,Z}(w, z) = F_{X,Y}(z, z).$$

Example 5.44   Radius and Angle of Independent Gaussian Random Variables

Let X and Y be zero-mean, unit-variance independent Gaussian random variables. Find the joint cdf and pdf of R and Θ, the radius and angle of the point (X, Y):

$$R = (X^2 + Y^2)^{1/2} \qquad \Theta = \tan^{-1}(Y/X).$$

The joint cdf of R and Θ is:

$$F_{R, \Theta}(r_0, \theta_0) = P[R \le r_0, \Theta \le \theta_0] = \iint\limits_{(x, y) \in R_{(r_0, \theta_0)}} \frac{e^{-(x^2+y^2)/2}}{2\pi} \, dx \, dy$$

where

$$R_{(r_0, \theta_0)} = \{(x, y) : \sqrt{x^2 + y^2} \le r_0, 0 < \tan^{-1}(Y/X) \le \theta_0\}.$$

The region R_{r_0, θ_0} is the pie-shaped region in Fig. 5.22. We change variables from Cartesian to polar coordinates to obtain:

$$F_{R,\Theta}(r_0, \theta_0) = P[R \le r_0, \Theta \le \theta_0] = \int_0^{r_0} \int_0^{\theta_0} \frac{e^{-r^2/2}}{2\pi} r \, dr \, d\theta$$

$$= \frac{\theta_0}{2\pi}(1 - e^{-r_0^2/2}), \quad 0 < \theta_0 < 2\pi \quad 0 < r_0 < \infty. \tag{5.57}$$
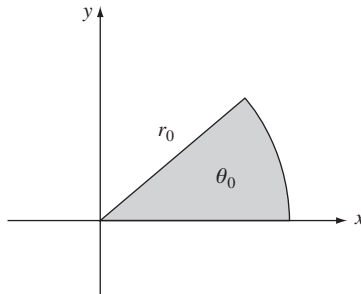
FIGURE 5.22
Region of integration R_{r_0, θ_0} in Example 5.44.

R and Θ are independent random variables, where R has a Rayleigh distribution and Θ is uniformly distributed in $(0, 2\pi)$. The joint pdf is obtained by taking partial derivatives with respect to r and θ:

$$f_{R,\Theta}(r, \theta) = \frac{\partial^2}{\partial r \partial \theta} \frac{\theta}{2\pi} (1 - e^{-r^2/2})$$

$$= \frac{1}{2\pi} (re^{-r^2/2}), \quad 0 < \theta < 2\pi \quad 0 < r < \infty.$$

This transformation maps every point in the plane from Cartesian coordinates to polar coordinates. We can also go *backwards* from polar to Cartesian coordinates. First we generate independent Rayleigh R and uniform Θ random variables. We then transform R and Θ into Cartesian coordinates to obtain an independent pair of zero-mean, unit-variance Gaussians. Neat!

5.8.3   pdf of Linear Transformations

The joint pdf of $\mathbf{Z}$ can be found directly in terms of the joint pdf of $\mathbf{X}$ by finding the equivalent events of infinitesimal rectangles. We consider the **linear transformation** of two random variables:

$$\begin{aligned} V &= aX + bY \\ W &= cX + eY \end{aligned} \quad \text{or} \quad \begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} a & b \\ c & e \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

Denote the above matrix by A. We will assume that A has an inverse, that is, it has determinant $|ae - bc| \neq 0$, so each point (v, w) has a unique corresponding point (x, y) obtained from

$$\begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} v \\ w \end{bmatrix}. \tag{5.58}$$

Consider the infinitesimal rectangle shown in Fig. 5.23. The points in this rectangle are mapped into the parallelogram shown in the figure. The infinitesimal rectangle and the parallelogram are equivalent events, so their probabilities must be equal. Thus

$$f_{X,Y}(x, y)dx\,dy \simeq f_{V,W}(v, w)\,dP$$

where dP is the area of the parallelogram. The joint pdf of V and W is thus given by

$$f_{V,W}(v, w) = \frac{f_{X,Y}(x, y)}{\left| \dfrac{dP}{dx\,dy} \right|}, \tag{5.59}$$

where x and y are related to (v, w) by Eq. (5.58). Equation (5.59) states that the joint pdf of V and W at (v, w) is the pdf of X and Y at the corresponding point (x, y), but rescaled by the "stretch factor" $dP/dx\,dy$. It can be shown that $dP = (|ae - bc|)\,dx\,dy$, so the "stretch factor" is

$$\left| \frac{dP}{dx\,dy} \right| = \frac{|ae - bc|(dx\,dy)}{(dx\,dy)} = |ae - bc| = |A|,$$
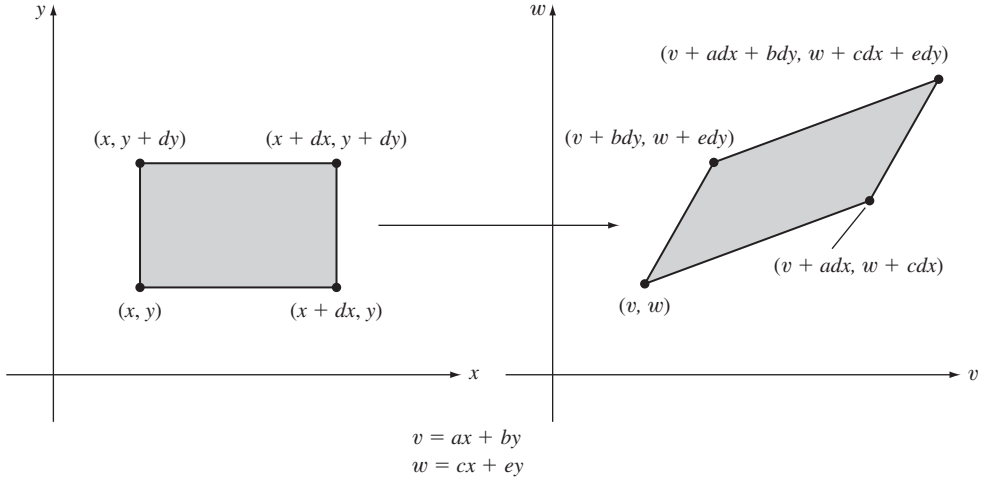
FIGURE 5.23
Image of an infinitesimal rectangle under a linear transformation.

where $|A|$ is the determinant of A.

The above result can be written compactly using matrix notation. Let the vector **Z** be

$$\mathbf{Z} = A\mathbf{X},$$

where A is an $n \times n$ invertible matrix. The joint pdf of **Z** is then

$$f_{\mathbf{z}}(\mathbf{z}) = \frac{f_{\mathbf{x}}(A^{-1}\mathbf{z})}{|A|} \tag{5.60}$$

Example 5.45   Linear Transformation of Jointly Gaussian Random Variables

Let X and Y be the jointly Gaussian random variables introduced in Example 5.18. Let V and W be obtained from (X, Y) by

$$\begin{bmatrix} V \\ W \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = A \begin{bmatrix} X \\ Y \end{bmatrix}.$$

Find the joint pdf of V and W.

The determinant of the matrix is $|A| = 1$, and the inverse mapping is given by

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} V \\ W \end{bmatrix},$$

so $X = (V - W)/\sqrt{2}$ and $Y = (V + W)/\sqrt{2}$. Therefore the pdf of V and W is

$$f_{V,W}(v, w) = f_{X,Y}\left(\frac{v - w}{\sqrt{2}}, \frac{v + w}{\sqrt{2}} \right),$$

where

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-(x^2-2\rho xy+y^2)/2(1-\rho^2)}.$$

By substituting for x and y, the argument of the exponent becomes

$$\frac{(v-w)^2/2 - 2\rho(v-w)(v+w)/2 + (v+w)^2/2}{2(1-\rho^2)}$$

$$= \frac{v^2}{2(1+\rho)} + \frac{w^2}{2(1-\rho)}.$$

Thus

$$f_{V,W}(v, w) = \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{-\{[v^2/2(1+\rho)]+[w^2/2(1-\rho)]\}}.$$

It can be seen that the transformed variables V and W are independent, zero-mean Gaussian random variables with variance $1 + \rho$ and $1 - \rho$, respectively. Figure 5.24 shows contours of equal value of the joint pdf of (X, Y). It can be seen that the pdf has elliptical symmetry about the origin with principal axes at $45°$ with respect to the axes of the plane. In Section 5.9 we show that the above linear transformation corresponds to a rotation of the coordinate system so that the axes of the plane are aligned with the axes of the ellipse.

5.9 PAIRS OF JOINTLY GAUSSIAN RANDOM VARIABLES

The jointly Gaussian random variables appear in numerous applications in electrical engineering. They are frequently used to model signals in signal processing applications, and they are the most important model used in communication systems that involve dealing with signals in the presence of noise. They also play a central role in many statistical methods.

The random variables X and Y are said to be **jointly Gaussian** if their joint pdf has the form

$$f_{X,Y}(x, y) = \frac{\exp\left\{\frac{-1}{2(1-\rho_{X,Y}^2)}\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{y-m_2}{\sigma_2}\right) + \left(\frac{y-m_2}{\sigma_2}\right)^2\right]\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{X,Y}^2}}$$

(5.61a)

for $-\infty < x < \infty$ and $-\infty < y < \infty$.

The pdf is centered at the point (m_1, m_2), and it has a bell shape that depends on the values of σ_1, σ_2, and $\rho_{X,Y}$ as shown in Fig. 5.25. As shown in the figure, the pdf is constant for values x and y for which the argument of the exponent is constant:

$$\left[\left(\frac{x-m_1}{\sigma_1}\right)^2 - 2\rho_{X,Y}\left(\frac{x-m_1}{\sigma_1}\right)\left(\frac{y-m_2}{\sigma_2}\right) + \left(\frac{y-m_2}{\sigma_2}\right)^2\right] = \text{constant}. \quad (5.61b)$$
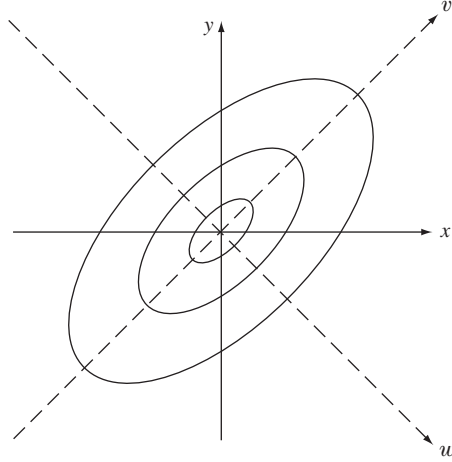
FIGURE 5.24
Contours of equal value of joint Gaussian pdf
discussed in Example 5.45.
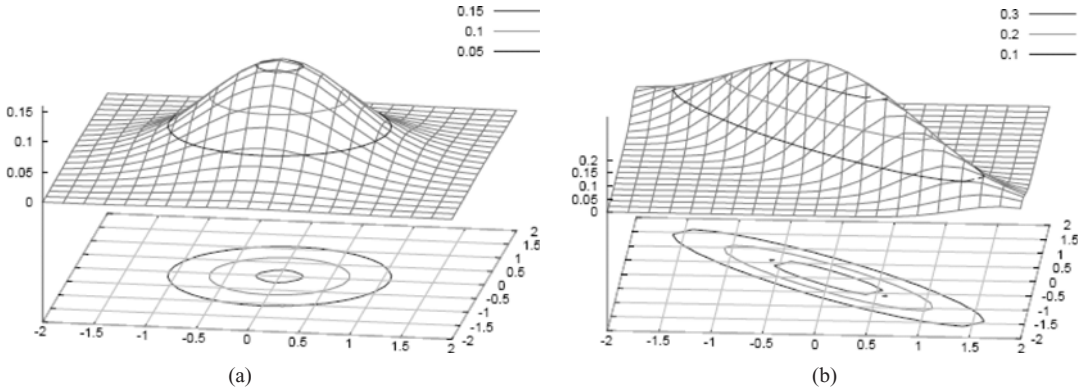
FIGURE 5.25
Jointly Gaussian pdf (a) $\rho = 0$ (b) $\rho = -0.9$.

Figure 5.26 shows the orientation of these elliptical contours for various values of σ_1, σ_2, and $\rho_{X,Y}$. When $\rho_{X,Y} = 0$, that is, when X and Y are independent, the equal-pdf contour is an ellipse with principal axes aligned with the x- and y-axes. When $\rho_{X,Y} \neq 0$, the major axis of the ellipse is oriented along the angle [Edwards and Penney, pp. 570–571]

$$\theta = \tfrac{1}{2} \arctan^{-1} \tan\left(\frac{2\rho_{X,Y}\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2} \right). \tag{5.62}$$

Note that the angle is 45° when the variances are equal.
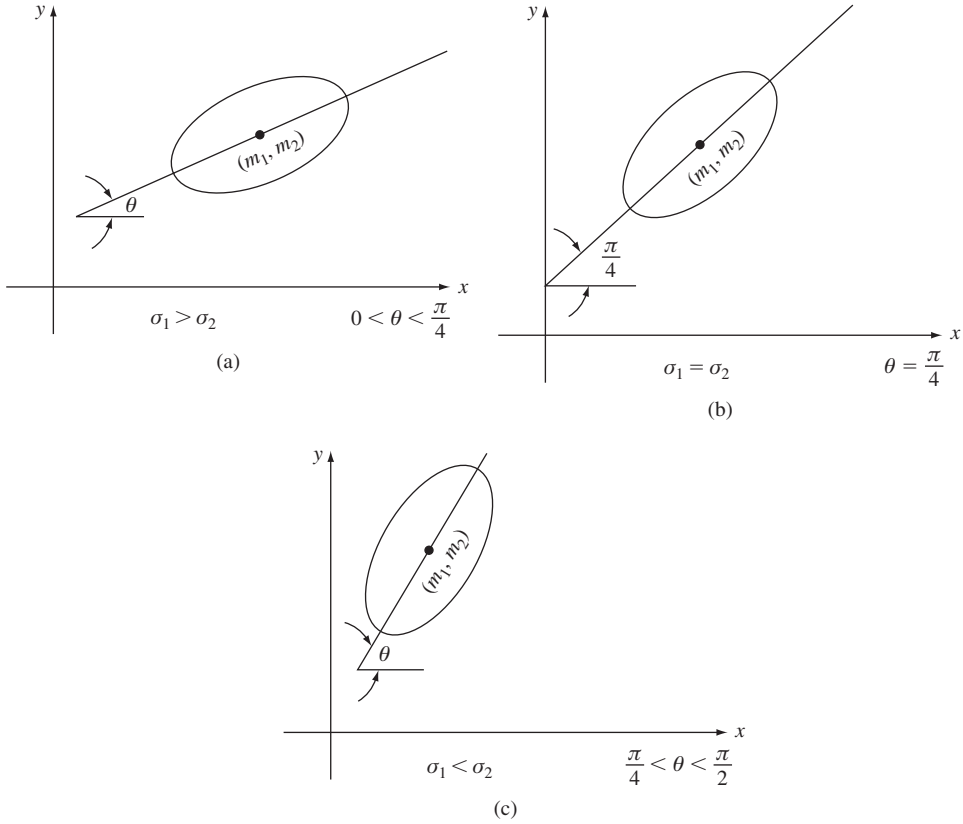
FIGURE 5.26
Orientation of contours of equal value of joint Gaussian pdf for $\rho_{X,Y} > 0$.

The marginal pdf of X is found by integrating $f_{X,Y}(x, y)$ over all y. The integration is carried out by completing the square in the exponent as was done in Example 5.18. The result is that the marginal pdf of X is

$$f_X(x) = \frac{e^{-(x-m_1)^2/2\sigma_1^2}}{\sqrt{2\pi}\sigma_1}, \tag{5.63}$$

that is, X is a Gaussian random variable with mean m_1 and variance σ_1^2. Similarly, the marginal pdf for Y is found to be Gaussian with pdf mean m_2 and variance σ_2^2.

The conditional pdf's $f_X(x\,|\,y)$ and $f_Y(y\,|\,x)$ give us information about the interrelation between X and Y. The conditional pdf of X given $Y = y$ is

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

$$= \frac{\exp\left\{\frac{-1}{2(1-\rho_{X,Y}^2)\sigma_1^2}\left[x - \rho_{X,Y}\frac{\sigma_1}{\sigma_2}(y - m_2) - m_1\right]^2\right\}}{\sqrt{2\pi\sigma_1^2(1 - \rho_{X,Y}^2)}}. \tag{5.64}$$

Equation (5.64) shows that the conditional pdf of X given $Y = y$ is also Gaussian but with conditional mean $m_1 + \rho_{X,Y}(\sigma_1/\sigma_2)(y - m_2)$ and conditional variance $\sigma_1^2(1 - \rho_{X,Y}^2)$. Note that when $\rho_{X,Y} = 0$, the conditional pdf of X given $Y = y$ equals the marginal pdf of X. This is consistent with the fact that X and Y are independent when $\rho_{X,Y} = 0$. On the other hand, as $|\rho_{X,Y}| \to 1$ the variance of X about the conditional mean approaches zero, so the conditional pdf approaches a delta function at the conditional mean. Thus when $|\rho_{X,Y}| = 1$, the conditional variance is zero and X is equal to the conditional mean with probability one. We note that similarly $f_Y(y \mid x)$ is Gaussian with conditional mean $m_2 + \rho_{X,Y}$ $(\sigma_2/\sigma_1)(x - m_1)$ and conditional variance $\sigma_2^2(1-\rho_{X,Y}^2)$.

We now show that the $\rho_{X,Y}$ in Eq. (5.61a) is indeed the correlation coefficient between X and Y. The covariance between X and Y is defined by

$$\text{COV}(X, Y) = E[(X - m_1)(Y - m_2)]$$
$$= E[E[(X - m_1)(Y - m_2) \mid Y]].$$

Now the conditional expectation of $(X - m_1)(Y - m_2)$ given $Y = y$ is

$$E[(X - m_1)(Y - m_2) \mid Y = y] = (y - m_2)E[X - m_1 \mid Y = y]$$
$$= (y - m_2)(E[X \mid Y = y] - m_1)$$
$$= (y - m_2)\left(\rho_{X,Y}\frac{\sigma_1}{\sigma_2}(y - m_2)\right),$$

where we have used the fact that the conditional mean of X given $Y = y$ is $m_1 + \rho_{X,Y}(\sigma_1/\sigma_2)(y - m_2)$. Therefore

$$E[(X - m_1)(Y - m_2) \mid Y] = \rho_{X,Y}\frac{\sigma_1}{\sigma_2}(Y - m_2)^2$$

and

$$\text{COV}(X, Y) = E[E[(X - m_1)(Y - m_2) \mid Y]] = \rho_{X,Y}\frac{\sigma_1}{\sigma_2}E[(Y - m_2)^2]$$
$$= \rho_{X,Y}\sigma_1\sigma_2.$$

The above equation is consistent with the definition of the correlation coefficient, $\rho_{X,Y} = \text{COV}(X, Y)/\sigma_1\sigma_2$. Thus the $\rho_{X,Y}$ in Eq. (5.61a) is indeed the correlation coefficient between X and Y.

Example 5.46

The amount of yearly rainfall in city 1 and in city 2 is modeled by a pair of jointly Gaussian random variables, X and Y, with pdf given by Eq. (5.61a). Find the most likely value of X given that we know $Y = y$.

The most likely value of X given $Y = y$ is the value of x for which $f_X(x \mid y)$ is maximum. The conditional pdf of X given $Y = y$ is given by Eq. (5.64), which is maximum at the conditional mean

$$E[X \mid y] = m_1 + \rho_{X,Y}\frac{\sigma_1}{\sigma_2}(y - m_2).$$

Note that this "maximum likelihood" estimate is a linear function of the observation y.

Example 5.47   Estimation of Signal in Noise

Let $Y = X + N$ where X (the "signal") and N (the "noise') are independent zero-mean Gaussian random variables with different variances. Find the correlation coefficient between the observed signal Y and the desired signal X. Find the value of x that maximizes $f_X(x \mid y)$.

The mean and variance of Y and the covariance of X and Y are:

$$E[Y] = E[X] + E[N] = 0$$

$$\sigma_Y^2 = E[Y^2] = E[(X+N)^2] = E[X^2 + 2XN + N^2] = E[X^2] + E[N^2] = \sigma_X^2 + \sigma_N^2.$$

$$COV(X, Y) = E[(X - E[X])(E(Y - E[Y])] = E[XY] = E[X(X+N)] = \sigma_X^2.$$

Therefore, the correlation coefficient is:

$$\rho_{X,Y} = \frac{COV(X,Y)}{\sigma_X \sigma_Y} = \frac{\sigma_X}{\sigma_Y} = \frac{\sigma_X}{(\sigma_X^2 + \sigma_N^2)^{1/2}} = \frac{1}{\left(1 + \dfrac{\sigma_N^2}{\sigma_X^2}\right)^{1/2}}.$$

Note that $\rho_{X,Y}^2 = \sigma_X^2/\sigma_Y^2 = 1 - \sigma_N^2/\sigma_Y^2$.

To find the joint pdf of X and Y consider the following linear transformation:

$$\begin{array}{lll} X = X & & X = X \\ Y = X + N & \text{which has inverse} & N = -X + Y. \end{array}$$

From Eq. (5.52) we have:

$$f_{X,Y}(x, y) = \left.\frac{f_{X,N}(x, y)}{\det A}\right|_{x=x, n=y-x} = \left.\frac{e^{-x^2/2\sigma_X^2}}{\sqrt{2\pi}\sigma_X} \frac{e^{-n^2/2\sigma_N^2}}{\sqrt{2\pi}\sigma_N}\right|_{x=x, n=y-x}$$

$$= \frac{e^{-x^2/2\sigma_X^2}}{\sqrt{2\pi}\sigma_X} \frac{e^{-(y-x)^2/2\sigma_N^2}}{\sqrt{2\pi}\sigma_N}.$$

The conditional pdf of the signal X given the observation Y is then:

$$f_X(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{e^{-x^2/2\sigma_X^2}}{\sqrt{2\pi}\sigma_X} \frac{e^{-(y-x)^2/2\sigma_N^2}}{\sqrt{2\pi}\sigma_N} \frac{\sqrt{2\pi}\sigma_Y}{e^{-y^2/2\sigma_Y^2}}$$

$$= \frac{\exp\left\{-\frac{1}{2}\left(\left(\frac{x}{\sigma_X}\right)^2 + \left(\frac{y-x}{\sigma_N}\right)^2 - \left(\frac{y}{\sigma_Y}\right)^2\right)\right\}}{\sqrt{2\pi}\sigma_N\sigma_X/\sigma_Y} = \frac{\exp\left\{-\frac{1}{2}\frac{\sigma_Y^2}{\sigma_X^2\sigma_N^2}\left(x - \frac{\sigma_X^2}{\sigma_Y^2}y\right)^2\right\}}{\sqrt{2\pi}\sigma_N\sigma_X/\sigma_Y}$$

$$= \frac{\exp\left\{-\frac{1}{2(1-\rho_{X,Y}^2)\sigma_X^2}\left(x - \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_X^2}\right)y\right)^2\right\}}{\sqrt{1 - \rho_{X,Y}^2}\,\sigma_X}.$$

This pdf has its maximum value, when the argument of the exponent is zero, that is,

$$x = \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2}\right)y = \left(\frac{1}{1 + \dfrac{\sigma_N^2}{\sigma_X^2}}\right)y.$$
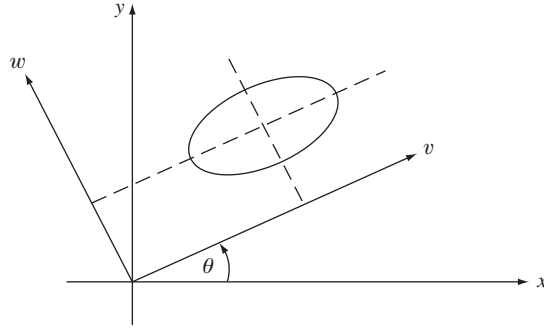
FIGURE 5.27
A rotation of the coordinate system transforms a pair of
dependent Gaussian random variables into a pair of independent
Gaussian random variables.

The signal-to-noise ratio (SNR) is defined as the ratio of the variance of X and the variance of N. At high SNRs this estimator gives $x \approx y$, and at very low signal-to-noise ratios, it gives $x \approx 0$.

Example 5.48   Rotation of Jointly Gaussian Random Variables

The ellipse corresponding to an arbitrary two-dimensional Gaussian vector forms an angle

$$\theta = \frac{1}{2} \arctan\left(\frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}\right)$$

relative to the x-axis. Suppose we define a new coordinate system whose axes are aligned with those of the ellipse as shown in Fig. 5.27. This is accomplished by using the following rotation matrix:

$$\begin{bmatrix} V \\ W \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}.$$

To show that the new random variables are independent it suffices to show that they have covariance zero:

$$\text{COV}(V, W) = E[(V - E[V])(W - E[W])]$$

$$= E[\{(X - m_1)\cos\theta + (Y - m_2)\sin\theta\}$$

$$\times \{-(X - m_1)\sin\theta + (Y - m_2)\cos\theta\}]$$

$$= -\sigma_1^2 \sin\theta \cos\theta + \text{COV}(X, Y)\cos^2\theta$$

$$-\text{COV}(X, Y)\sin^2\theta + \sigma_2^2 \sin\theta \cos\theta$$

$$= \frac{(\sigma_2^2 - \sigma_1^2)\sin 2\theta + 2\,\text{COV}(X, Y)\cos 2\theta}{2}$$

$$= \frac{\cos 2\theta[(\sigma_2^2 - \sigma_1^2)\tan 2\theta + 2\,\text{COV}(X, Y)]}{2}.$$

If we let the angle of rotation θ be such that

$$\tan 2\theta = \frac{2 \, \text{COV}(X, Y)}{\sigma_1^2 - \sigma_2^2},$$

then the covariance of V and W is zero as required.

---

*5.10   GENERATING INDEPENDENT GAUSSIAN RANDOM VARIABLES

We now present a method for generating unit-variance, uncorrelated (and hence independent) jointly Gaussian random variables. Suppose that X and Y are two independent zero-mean, unit-variance jointly Gaussian random variables with pdf:

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-(x^2 + y^2)/2}.$$

In Example 5.44 we saw that the transformation

$$R = \sqrt{X^2 + Y^2} \qquad \text{and} \qquad \Theta = \tan^{-1} Y/X$$

leads to the pair of independent random variables

$$f_{R,\Theta}(r, \theta) = \frac{1}{2\pi} r e^{-r^2/2} = f_R(r) f_\Theta(\theta),$$

where R is a Rayleigh random variable and Θ is a uniform random variable. The above transformation is invertible. Therefore we can also start with independent Rayleigh and uniform random variables and produce zero-mean, unit-variance independent Gaussian random variables through the transformation:

$$X = R \cos \Theta \quad \text{and} \quad Y = R \sin \Theta. \tag{5.65}$$

Consider $W = R^2$ where R is a Rayleigh random variable. From Example 5.41 we then have that: W has pdf

$$f_W(w) = \frac{f_R(\sqrt{w})}{2\sqrt{w}} = \frac{\sqrt{w} e^{-\sqrt{w}^2/2}}{2\sqrt{w}} = \frac{1}{2} e^{-w/2}.$$

$W = R^2$ has an exponential distribution with $\lambda = 1/2$.

Therefore we can generate R^2 by generating an exponential random variable with parameter $1/2$, and we can generate Θ by generating a random variable that is uniformly distributed in the interval $(0, 2\pi)$. If we substitute these random variables into Eq. (5.65), we then obtain a pair of independent zero-mean, unit-variance Gaussian random variables. The above discussion thus leads to the following algorithm:

1. Generate U_1 and U_2, two independent random variables uniformly distributed in the unit interval.
2. Let $R^2 = -2 \log U_1$ and $\Theta = 2\pi U_2$.
3. Let $X = R \cos \Theta = (-2 \log U_1)^{1/2} \cos 2\pi U_2$ and $Y = R \sin \Theta = (-2 \log U_1)^{1/2} \sin 2\pi U_2$.

Then X and Y are independent, zero-mean, unit-variance Gaussian random variables. By repeating the above procedure we can generate any number of such random variables.

Example 5.49

Use Octave or MATLAB to generate 1000 independent zero-mean, unit-variance Gaussian random variables. Compare a histogram of the observed values with the pdf of a zero-mean unit-variance random variable.

The Octave commands below show the steps for generating the Gaussian random variables. A set of histogram range values K from -4 to 4 is created and used to build a normalized histogram Z. The points in Z are then plotted and compared to the value predicted to fall in each interval by the Gaussian pdf. These plots are shown in Fig. 5.28, which shows excellent agreement.

```
> U1=rand(1000,1);          % Create a 1000-element vector U₁ (step 1).
> U2=rand(1000,1);          % Create a 1000-element vector U₂ (step 1).
> R2=-2*log(U1);            % Find R² (step 2).
> TH=2*pi*U2;               % Find θ (step 2).
> X=sqrt(R2).*sin(TH);      % Generate X (step 3).
```

The Octave commands:
- `> U1=rand(1000,1);` — % Create a 1000-element vector U_1 (step 1).
- `> U2=rand(1000,1);` — % Create a 1000-element vector U_2 (step 1).
- `> R2=-2*log(U1);` — % Find R^2 (step 2).
- `> TH=2*pi*U2;` — % Find θ (step 2).
- `> X=sqrt(R2).*sin(TH);` — % Generate X (step 3).
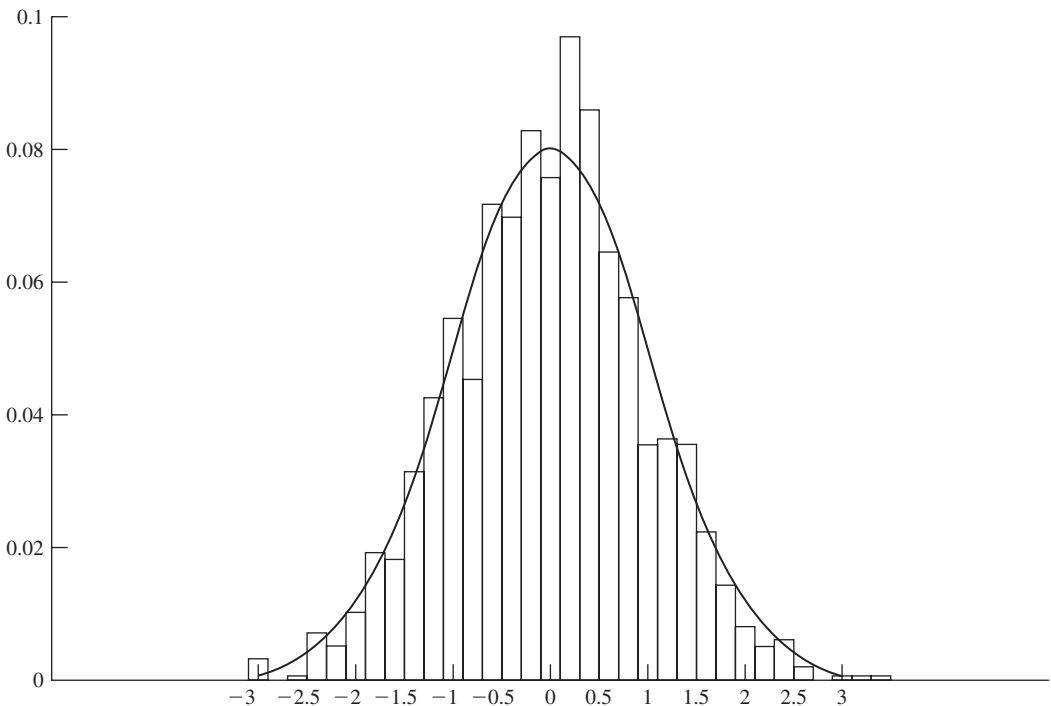
FIGURE 5.28
Histogram of 1000 observations of a Gaussian random variable.
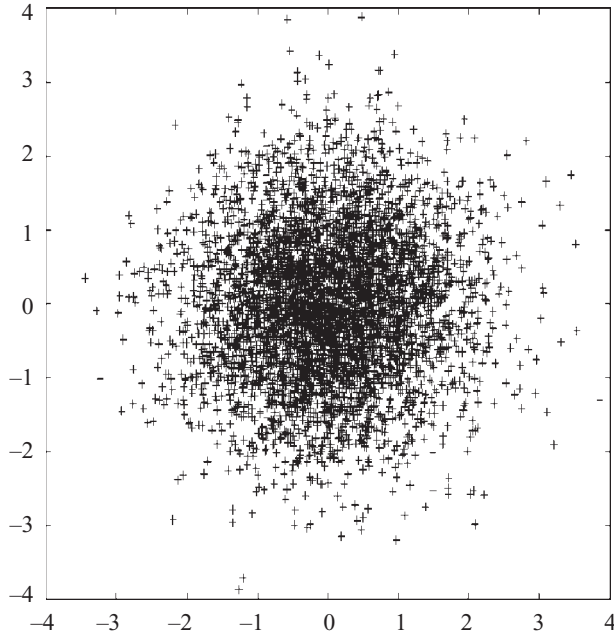
FIGURE 5.29
Scattergram of 5000 pairs of jointly Gaussian random variables.

```
> Y=sqrt(R2).*cos(TH);           % Generate Y (step 3).
> K=-4:.2:4;                     % Create histogram range values K.
> Z=hist(X,K)/1000              % Create normalized histogram Z based on K.
> bar(K,Z)                      % Plot Z.
> hold on
> stem(K,.2*normal_pdf(K,0,1))   % Compare to values predicted by pdf.
```

We also plotted the X values vs. the Y values for 5000 pairs of generated random variables in a scattergram as shown in Fig. 5.29. Good agreement with the circular symmetry of the jointly Gaussian pdf of zero-mean, unit-variance pairs is observed.

In the next chapter we will show how to generate a vector of jointly Gaussian random variables with an arbitrary covariance matrix.

SUMMARY

- The joint statistical behavior of a pair of random variables X and Y is specified by the joint cumulative distribution function, the joint probability mass function, or the joint probability density function. The probability of any event involving the joint behavior of these random variables can be computed from these functions.

- The statistical behavior of individual random variables from **X** is specified by the marginal cdf, marginal pdf, or marginal pmf that can be obtained from the joint cdf, joint pdf, or joint pmf of **X**.
- Two random variables are independent if the probability of a product-form event is equal to the product of the probabilities of the component events. Equivalent conditions for the independence of a set of random variables are that the joint cdf, joint pdf, or joint pmf factors into the product of the corresponding marginal functions.
- The covariance and the correlation coefficient of two random variables are measures of the linear dependence between the random variables.
- If **X** and **Y** are independent, then **X** and **Y** are uncorrelated, but not vice versa. If **X** and **Y** are jointly Gaussian and uncorrelated, then they are independent.
- The statistical behavior of **X**, given the exact values of X or Y, is specified by the conditional cdf, conditional pmf, or conditional pdf. Many problems lend themselves to a solution that involves conditioning on the value of one of the random variables. In these problems, the expected value of random variables can be obtained by conditional expectation.
- The joint pdf of a pair of jointly Gaussian random variables is determined by the means, variances, and covariance. All marginal pdf's and conditional pdf's are also Gaussian pdf's.
- Independent Gaussian random variables can be generated by a transformation of uniform random variables.

CHECKLIST OF IMPORTANT TERMS

Central moments of X and Y	Joint pmf
Conditional cdf	Jointly continuous random variables
Conditional expectation	Jointly Gaussian random variables
Conditional pdf	Linear transformation
Conditional pmf	Marginal cdf
Correlation of X and Y	Marginal pdf
Covariance X and Y	Marginal pmf
Independent random variables	Orthogonal random variables
Joint cdf	Product-form event
Joint moments of X and Y	Uncorrelated random variables
Joint pdf	

ANNOTATED REFERENCES

Papoulis [1] is the standard reference for electrical engineers for the material on random variables. References [2] and [3] present many interesting examples involving multiple random variables. The book by Jayant and Noll [4] gives numerous applications of probability concepts to the digital coding of waveforms.

1. A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 2002.

2. L. Breiman, *Probability and Stochastic Processes*, Houghton Mifflin, Boston, 1969.

3. H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences*, vol. 1, Wiley, New York, 1979.

4. N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs, N.J., 1984.

5. N. Johnson et al., *Continuous Multivariate Distributions*, Wiley, New York, 2000.

6. H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice Hall, Englewood Cliffs, N.J., 1986.

7. H. Anton, *Elementary Linear Algebra*, 9th ed., Wiley, New York, 2005.

8. C. H. Edwards, Jr., and D. E. Penney, *Calculus and Analytic Geometry*, 4th ed., Prentice Hall, Englewood Cliffs, N.J., 1994.

PROBLEMS

Section 5.1: Two Random Variables

5.1. Let X be the maximum and let Y be the minimum of the number of heads obtained when Carlos and Michael each flip a fair coin twice.

   (a) Describe the underlying space S of this random experiment and show the mapping from S to S_{XY}, the range of the pair (X, Y).

   (b) Find the probabilities for all values of (X, Y).

   (c) Find $P[X = Y]$.

   (d) Repeat parts b and c if Carlos uses a biased coin with $P[\text{heads}] = 3/4$.

5.2. Let X be the difference and let Y be the sum of the number of heads obtained when Carlos and Michael each flip a fair coin twice.

   (a) Describe the underlying space S of this random experiment and show the mapping from S to S_{XY}, the range of the pair (X, Y).

   (b) Find the probabilities for all values of (X, Y).

   (c) Find $P[X + Y = 1]$, $P[X + Y = 2]$.

5.3. The input X to a communication channel is "-1"or "1", with respective probabilities 1/4 and 3/4. The output of the channel Y is equal to: the corresponding input X with probability $1 - p - p_e$; $-X$ with probability p; 0 with probability p_e.

   (a) Describe the underlying space S of this random experiment and show the mapping from S to S_{XY}, the range of the pair (X, Y).

   (b) Find the probabilities for all values of (X, Y).

   (c) Find $P[X \neq Y]$, $P[Y = 0]$.

5.4. (a) Specify the range of the pair (N_1, N_2) in Example 5.2.

   (b) Specify and sketch the event "more revenue comes from type 1 requests than type 2 requests."

5.5. (a) Specify the range of the pair (Q, R) in Example 5.3.

   (b) Specify and sketch the event "last packet is more than half full."

5.6. Let the pair of random variables H and W be the height and weight in Example 5.1. The body mass index is a measure of body fat and is defined by BMI $= W/H^2$ where W is in kilograms and H is in meters. Determine and sketch on the plane the following events: $A = \{\text{"obese," BMI} \geq 30\}$; $B = \{\text{"overweight," } 25 \leq \text{BMI} < 30\}$; $C = \{\text{"normal," } 18.5 \leq \text{BMI} < 25\}$; and $D = \{\text{"underweight," BMI} < 18.5\}$.

5.7. Let (X, Y) be the two-dimensional noise signal in Example 5.4. Specify and sketch the events:

   (a) "Maximum noise magnitude is greater than 5."

   (b) "The noise power $X^2 + Y^2$ is greater than 4."

   (c) "The noise power $X^2 + Y^2$ is greater than 4 and less than 9."

5.8. For the pair of random variables (X, Y) sketch the region of the plane corresponding to the following events. Identify which events are of product form.

   (a) $\{X + Y > 3\}$.

   (b) $\{e^X > Ye^3\}$.

   (c) $\{\min(X, Y) > 0\} \cup \{\max\{X, Y\} < 0\}$.

   (d) $\{|X - Y| \geq 1\}$.

   (e) $\{|X/Y| > 2\}$.

   (f) $\{X/Y < 2\}$.

   (g) $\{X^3 > Y\}$.

   (h) $\{XY < 0\}$.

   (i) $\{\max(|X|, Y) < 3\}$.

Section 5.2: Pairs of Discrete Random Variables

5.9. (a) Find and sketch $p_{X,Y}(x, y)$ in Problem 5.1 when using a fair coin.

   (b) Find $p_X(x)$ and $p_Y(y)$.

   (c) Repeat parts a and b if Carlos uses a biased coin with $P[\text{heads}] = 3/4$.

5.10. (a) Find and sketch $p_{X,Y}(x, y)$ in Problem 5.2 when using a fair coin.

   (b) Find $p_X(x)$ and $p_Y(y)$.

   (c) Repeat parts a and b if Carlos uses a biased coin with $P[\text{heads}] = 3/4$.

5.11. (a) Find the marginal pmf's for the pairs of random variables with the indicated joint pmf.

	(i)				(ii)				(iii)		
X/Y	−1	0	1	X/Y	−1	0	1	X/Y	−1	0	1
−1	1/6	1/6	0	−1	1/9	1/9	1/9	−1	1/3	0	0
0	0	0	1/3	0	1/9	1/9	1/9	0	0	1/3	0
1	1/6	1/6	0	1	1/9	1/9	1/9	1	0	0	1/3

   (b) Find the probability of the events $A = \{X > 0\}, B = \{X \geq Y\}$, and $C = \{X = -Y\}$ for the above joint pmf's.

5.12. A modem transmits a two-dimensional signal (X, Y) given by:

$$X = r\cos(2\pi\Theta/8) \quad \text{and} \quad Y = r\sin(2\pi\Theta/8)$$

where Θ is a discrete uniform random variable in the set $\{0, 1, 2, \ldots, 7\}$.

   (a) Show the mapping from S to S_{XY}, the range of the pair (X, Y).

   (b) Find the joint pmf of X and Y.

   (c) Find the marginal pmf of X and of Y.

   (d) Find the probability of the following events: $A = \{X = 0\}, B = \{Y \leq r/\sqrt{2}\}$, $C = \{X \geq r/\sqrt{2}, Y \geq r/\sqrt{2}\}, D = \{X < -r/\sqrt{2}\}$.

5.13. Let N_1 be the number of Web page requests arriving at a server in a 100-ms period and let N_2 be the number of Web page requests arriving at a server in the next 100-ms period. Assume that in a 1-ms interval either zero or one page request takes place with respective probabilities $1 - p = 0.95$ and $p = 0.05$, and that the requests in different 1-ms intervals are independent of each other.

   (a) Describe the underlying space S of this random experiment and show the mapping from S to S_{XY}, the range of the pair (X, Y).

   (b) Find the joint pmf of X and Y.

   (c) Find the marginal pmf for X and for Y.

   (d) Find the probability of the events $A = \{X \geq Y\}$, $B = \{X = Y = 0\}$, $C = \{X > 5, Y > 3\}$.

   (e) Find the probability of the event $D = \{X + Y = 10\}$.

5.14. Let N_1 be the number of Web page requests arriving at a server in the period $(0, 100)$ ms and let N_2 be the *total* combined number of Web page requests arriving at a server in the period $(0, 200)$ ms. Assume arrivals occur as in Problem 5.13.

   (a) Describe the underlying space S of this random experiment and show the mapping from S to S_{XY}, the range of the pair (X, Y).

   (b) Find the joint pmf of N_1 and N_2.

   (c) Find the marginal pmf for N_1 and N_2.

   (d) Find the probability of the events $A = \{N_1 < N_2\}$, $B = \{N_2 = 0\}$, $C = \{N_1 > 5, N_2 > 3\}$, $D = \{|N_2 - 2N_1| < 2\}$.

5.15. At even time instants, a robot moves either $+\Delta$ cm or $-\Delta$ cm in the x-direction according to the outcome of a coin flip; at odd time instants, a robot moves similarly according to another coin flip in the y-direction. Assuming that the robot begins at the origin, let X and Y be the coordinates of the location of the robot after $2n$ time instants.

   (a) Describe the underlying space S of this random experiment and show the mapping from S to S_{XY}, the range of the pair (X, Y).

   (b) Find the marginal pmf of the coordinates X and Y.

   (c) Find the probability that the robot is within distance $\sqrt{2}$ of the origin after $2n$ time instants.

Section 5.3: The Joint cdf of x and y

5.16. (a) Sketch the joint cdf for the pair (X, Y) in Problem 5.1 and verify that the properties of the joint cdf are satisfied. You may find it helpful to first divide the plane into regions where the cdf is constant.

   (b) Find the marginal cdf of X and of Y.

5.17. A point (X, Y) is selected at random inside a triangle defined by $\{(x, y): 0 \leq y \leq x \leq 1\}$. Assume the point is equally likely to fall anywhere in the triangle.

   (a) Find the joint cdf of X and Y.

   (b) Find the marginal cdf of X and of Y.

   (c) Find the probabilities of the following events in terms of the joint cdf: $A = \{X \leq 1/2, Y \leq 3/4\}$; $B = \{1/4 < X \leq 3/4, 1/4 < Y \leq 3/4\}$.

5.18. A dart is equally likely to land at any point (X_1, X_2) inside a circular target of unit radius. Let R and Θ be the radius and angle of the point (X_1, X_2).

   (a) Find the joint cdf of R and Θ.

   (b) Find the marginal cdf of R and Θ.

(c) Use the joint cdf to find the probability that the point is in the first quadrant of the real plane and that the radius is greater than 0.5.

5.19. Find an expression for the probability of the events in Problem 5.8 parts c, h, and i in terms of the joint cdf of X and Y.

5.20. The pair (X, Y) has joint cdf given by:

$$F_{X,Y}(x, y) = \begin{cases} (1 - 1/x^2)(1 - 1/y^2) & \text{for } x > 1, y > 1 \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Sketch the joint cdf.
(b) Find the marginal cdf of X and of Y.
(c) Find the probability of the following events: $\{X < 3, Y \le 5\}$, $\{X > 4, Y > 3\}$.

5.21. Is the following a valid cdf? Why?

$$F_{X,Y}(x, y) = \begin{cases} (1 - 1/x^2y^2) & \text{for } x > 1, y > 1 \\ 0 & \text{elsewhere.} \end{cases}$$

5.22. Let $F_X(x)$ and $F_Y(y)$ be valid one-dimensional cdf's. Show that $F_{X,Y}(x, y) = F_X(x)F_Y(y)$ satisfies the properties of a two-dimensional cdf.

5.23. The number of users logged onto a system N and the time T until the next user logs off have joint probability given by:

$$P[N = n, X \le t] = (1 - \rho)\rho^{n-1}(1 - e^{-n\lambda t}) \qquad \text{for } n = 1, 2, \dots \quad t > 0.$$

(a) Sketch the above joint probability.
(b) Find the marginal pmf of N.
(c) Find the marginal cdf of X.
(d) Find $P[N \le 3, X > 3/\lambda]$.

5.24. A factory has n machines of a certain type. Let p be the probability that a machine is working on any given day, and let N be the total number of machines working on a certain day. The time T required to manufacture an item is an exponentially distributed random variable with rate $k\alpha$ if k machines are working. Find and $P[T \le t]$. Find $P[T \le t]$ as $t \to \infty$ and explain the result.

Section 5.4: The Joint pdf of Two Continuous Random Variables

5.25. The amplitudes of two signals X and Y have joint pdf:

$$f_{X,Y}(x, y) = e^{-x/2}ye^{-y^2} \qquad \text{for } x > 0, y > 0.$$

(a) Find the joint cdf.
(b) Find $P[X^{1/2} > Y]$.
(c) Find the marginal pdfs.

5.26. Let X and Y have joint pdf:

$$f_{X,Y}(x, y) = k(x + y) \qquad \text{for } 0 \le x \le 1, 0 \le y \le 1.$$

(a) Find k.
(b) Find the joint cdf of (X, Y).
(c) Find the marginal pdf of X and of Y.
(d) Find $P[X < Y], P[Y < X^2], P[X + Y > 0.5]$.

5.27. Let X and Y have joint pdf:

$$f_{X,Y}(x, y) = kx(1 - x)y \qquad \text{for } 0 < x < 1, 0 < y < 1.$$

(a) Find k.
(b) Find the joint cdf of (X, Y).
(c) Find the marginal pdf of X and of Y.
(d) Find $P[Y < X^{1/2}]$, $P[X < Y]$.

5.28. The random vector (X, Y) is uniformly distributed (i.e., $f(x, y) = k$) in the regions shown in Fig. P5.1 and zero elsewhere.
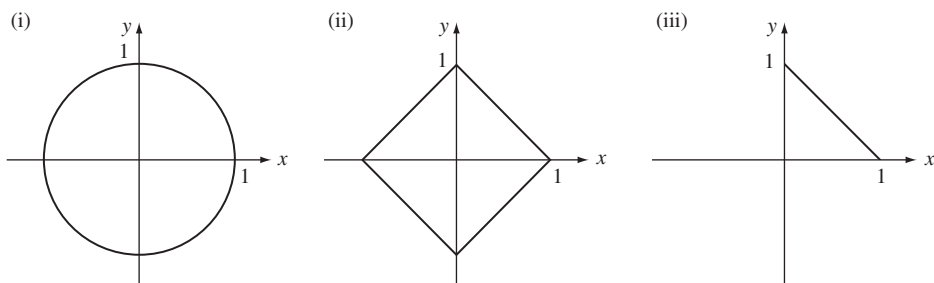
FIGURE P5.1

(a) Find the value of k in each case.
(b) Find the marginal pdf for X and for Y in each case.
(c) Find $P[X > 0, Y > 0]$.

5.29. (a) Find the joint cdf for the vector random variable introduced in Example 5.16.
(b) Use the result of part a to find the marginal cdf of X and of Y.

5.30. Let X and Y have the joint pdf:

$$f_{X,Y}(x, y) = ye^{-y(1+x)} \qquad \text{for } x > 0, y > 0.$$

Find the marginal pdf of X and of Y.

5.31. Let X and Y be the pair of random variables in Problem 5.17.
(a) Find the joint pdf of X and Y.
(b) Find the marginal pdf of X and of Y.
(c) Find $P[Y < X^2]$.

5.32. Let R and Θ be the pair of random variables in Problem 5.18.
(a) Find the joint pdf of R and Θ.
(b) Find the marginal pdf of R and of Θ.

5.33. Let (X, Y) be the jointly Gaussian random variables discussed in Example 5.18. Find $P[X^2 + Y^2 > r^2]$ when $\rho = 0$. *Hint:* Use polar coordinates to compute the integral.

5.34. The general form of the joint pdf for two jointly Gaussian random variables is given by Eq. (5.61a). Show that X and Y have marginal pdfs that correspond to Gaussian random variables with means m_1 and m_2 and variances σ_1^2 and σ_2^2 respectively.

5.35. The input X to a communication channel is $+1$ or -1 with probability p and $1 - p$, respectively. The received signal Y is the sum of X and noise N which has a Gaussian distribution with zero mean and variance $\sigma^2 = 0.25$.

    (a) Find the joint probability $P[X = j, Y \leq y]$.

    (b) Find the marginal pmf of X and the marginal pdf of Y.

    (c) Suppose we are given that $Y > 0$. Which is more likely, $X = 1$ or $X = -1$?

5.36. A modem sends a two-dimensional signal $\mathbf{X}$ from the set $\{(1, 1), (1, -1), (-1, 1), (-1, -1)\}$. The channel adds a noise signal (N_1, N_2), so the received signal is $\mathbf{Y} = \mathbf{X} + \mathbf{N} = (X_1 + N_1, X_2 + N_2)$. Assume that (N_1, N_2) have the jointly Gaussian pdf in Example 5.18 with $\rho = 0$. Let the distance between $\mathbf{X}$ and $\mathbf{Y}$ be $d(\mathbf{X}, \mathbf{Y}) = \{(X_1 - Y_1)^2 + (X_2 - Y_2)^2\}^{1/2}$.

    (a) Suppose that $\mathbf{X} = (1, 1)$. Find and sketch region for the event $\{\mathbf{Y}$ is closer to $(1, 1)$ than to the other possible values of $\mathbf{X}\}$. Evaluate the probability of this event.

    (b) Suppose that $\mathbf{X} = (1, 1)$. Find and sketch region for the event $\{\mathbf{Y}$ is closer to $(1, -1)$ than to the other possible values of $\mathbf{X}\}$. Evaluate the probability of this event.

    (c) Suppose that $\mathbf{X} = (1, 1)$. Find and sketch region for the event $\{d(\mathbf{X}, \mathbf{Y}) > 1\}$. Evaluate the probability of this event. Explain why this probability is an upper bound on the probability that $\mathbf{Y}$ is closer to a signal other than $\mathbf{X} = (1, 1)$.

Section 5.5: Independence of Two Random Variables

5.37. Let X be the number of full pairs and let Y be the remainder of the number of dots observed in a toss of a fair die. Are X and Y independent random variables?

5.38. Let X and Y be the coordinates of the robot in Problem 5.15 after $2n$ time instants. Determine whether X and Y are independent random variables.

5.39. Let X and Y be the coordinates of the two-dimensional modem signal (X, Y) in Problem 5.12.

    (a) Determine if X and Y are independent random variables.

    (b) Repeat part a if even values of Θ are twice as likely as odd values.

5.40. Determine which of the joint pmfs in Problem 5.11 correspond to independent pairs of random variables.

5.41. Michael takes the 7:30 bus every morning. The arrival time of the bus at the stop is uniformly distributed in the interval $[7:27, 7:37]$. Michael's arrival time at the stop is also uniformly distributed in the interval $[7:25, 7:40]$. Assume that Michael's and the bus's arrival times are independent random variables.

    (a) What is the probability that Michael arrives more than 5 minutes before the bus?

    (b) What is the probability that Michael misses the bus?

5.42. Are R and Θ independent in Problem 5.18?

5.43. Are X and Y independent in Problem 5.20?

5.44. Are the signal amplitudes X and Y independent in Problem 5.25?

5.45. Are X and Y independent in Problem 5.26?

5.46. Are X and Y independent in Problem 5.27?

5.47. Let X and Y be independent random variables. Find an expression for the probability of the following events in terms of $F_X(x)$ and $F_Y(y)$.

(a) $\{a < X \leq b\} \cap \{Y > d\}$.

(b) $\{a < X \leq b\} \cap \{c \leq Y < d\}$.

(c) $\{|X| < a\} \cap \{c \leq Y \leq d\}$.

5.48. Let X and Y be independent random variables that are uniformly distributed in $[-1, 1]$. Find the probability of the following events:

(a) $P[X^2 < 1/2, |Y| < 1/2]$.

(b) $P[4X < 1, Y < 0]$.

(c) $P[XY < 1/2]$.

(d) $P[\max(X, Y) < 1/3]$.

5.49. Let X and Y be random variables that take on values from the set $\{-1, 0, 1\}$.

(a) Find a joint pmf for which X and Y are independent.

(b) Are X^2 and Y^2 independent random variables for the pmf in part a?

(c) Find a joint pmf for which X and Y are not independent, but for which X^2 and Y^2 are independent.

5.50. Let X and Y be the jointly Gaussian random variables introduced in Problem 5.34.

(a) Show that X and Y are independent random variables if and only if $\rho = 0$.

(b) Suppose $\rho = 0$, find $P[XY < 0]$.

5.51. Two fair dice are tossed repeatedly until a pair occurs. Let K be the number of tosses required and let X be the number showing up in the pair. Find the joint pmf of K and X and determine whether K and X are independent.

5.52. The number of devices L produced in a day is geometric distributed with probability of success p. Let N be the number of working devices and let M be the number of defective devices produced in a day.

(a) Are N and M independent random variables?

(b) Find the joint pmf of N and M.

(c) Find the marginal pmfs of N and M. (See hint in Problem 5.87b.)

(d) Are L and M independent random variables?

5.53. Let N_1 be the number of Web page requests arriving at a server in a 100-ms period and let N_2 be the number of Web page requests arriving at a server in the next 100-ms period. Use the result of Problem 5.13 parts a and b to develop a model where N_1 and N_2 are independent Poisson random variables.

5.54. (a) Show that Eq. (5.22) implies Eq. (5.21).

(b) Show that Eq. (5.21) implies Eq. (5.22).

5.55. Verify that Eqs. (5.22) and (5.23) can be obtained from each other.

Section 5.6: Joint Moments and Expected Values of a Function of Two Random Variables

5.56. (a) Find $E[(X + Y)^2]$.

(b) Find the variance of $X + Y$.

(c) Under what condition is the variance of the sum equal to the sum of the individual variances?

5.57. Find $E[|X - Y|]$ if X and Y are independent exponential random variables with parameters $\lambda_1 = 1$ and $\lambda_2 = 2$, respectively.

5.58. Find $E[X^2 e^Y]$ where X and Y are independent random variables, X is a zero-mean, unit-variance Gaussian random variable, and Y is a uniform random variable in the interval $[0, 3]$.

5.59. For the discrete random variables X and Y in Problem 5.1, find the correlation and covariance, and indicate whether the random variables are independent, orthogonal, or uncorrelated.

5.60. For the discrete random variables X and Y in Problem 5.2, find the correlation and covariance, and indicate whether the random variables are independent, orthogonal, or uncorrelated.

5.61. For the three pairs of discrete random variables in Problem 5.11, find the correlation and covariance of X and Y, and indicate whether the random variables are independent, orthogonal, or uncorrelated.

5.62. Let N_1 and N_2 be the number of Web page requests in Problem 5.13. Find the correlation and covariance of N_1 and N_2, and indicate whether the random variables are independent, orthogonal, or uncorrelated.

5.63. Repeat Problem 5.62 for N_1 and N_2, the number of Web page requests in Problem 5.14.

5.64. Let N and T be the number of users logged on and the time till the next logoff in Problem 5.23. Find the correlation and covariance of N and T, and indicate whether the random variables are independent, orthogonal, or uncorrelated.

5.65. Find the correlation and covariance of X and Y in Problem 5.26. Determine whether X and Y are independent, orthogonal, or uncorrelated.

5.66. Repeat Problem 5.65 for X and Y in Problem 5.27.

5.67. For the three pairs of continuous random variables X and Y in Problem 5.28, find the correlation and covariance, and indicate whether the random variables are independent, orthogonal, or uncorrelated.

5.68. Find the correlation coefficient between X and $Y = aX + b$. Does the answer depend on the sign of a?

5.69. Propose a method for estimating the covariance of two random variables.

5.70. **(a)** Complete the calculations for the correlation coefficient in Example 5.28.
   (b) Repeat the calculations if X and Y have the pdf:

$$f_{X,Y}(x, y) = e^{-(x+|y|)} \qquad \text{for } x > 0, -x < y < x.$$

5.71. The output of a channel $Y = X + N$, where the input X and the noise N are independent, zero-mean random variables.
   (a) Find the correlation coefficient between the input X and the output Y.
   (b) Suppose we estimate the input X by a linear function $g(Y) = aY$. Find the value of a that minimizes the mean squared error $E[(X - aY)^2]$.
   (c) Express the resulting mean-square error in terms of σ_X/σ_N.

5.72. In Example 5.27 let $X = \cos \Theta/4$ and $Y = \sin \Theta/4$. Are X and Y uncorrelated?

5.73. **(a)** Show that $\text{COV}(X, E[Y \mid X]) = \text{COV}(X, Y)$.
   (b) Show that $E[Y \mid X = x] = E[Y]$, for all x, implies that X and Y are uncorrelated.

5.74. Use the fact that $E[(tX + Y)^2] \geq 0$ for all t to prove the Cauchy-Schwarz inequality:

$$(E[XY])^2 \leq E[X^2]E[Y^2].$$

Hint: Consider the discriminant of the quadratic equation in t that results from the above inequality.

Section 5.7: Conditional Probability and Conditional Expectation

5.75. (a) Find $p_Y(y \mid x)$ and $p_X(x \mid y)$ in Problem 5.1 assuming fair coins are used.

  (b) Find $p_Y(y \mid x)$ and $p_X(x \mid y)$ in Problem 5.1 assuming Carlos uses a coin with $p = 3/4$.

  (c) What is the effect on $p_X(x \mid y)$ of Carlos using a biased coin?

  (d) Find $E[Y \mid X = x]$ and $E[X \mid Y = y]$ in part a; then find $E[X]$ and $E[Y]$.

  (e) Find $E[Y \mid X = x]$ and $E[X \mid Y = y]$ in part b; then find $E[X]$ and $E[Y]$.

5.76. (a) Find $p_X(x \mid y)$ for the communication channel in Problem 5.3.

  (b) For each value of y, find the value of x that maximizes $p_X(x \mid y)$. State any assumptions about p and p_e.

  (c) Find the probability of error if a receiver uses the decision rule from part b.

5.77. (a) In Problem 5.11(i), which conditional pmf given X provides the most information about Y: $p_Y(y \mid -1)$, $p_Y(y \mid 0)$, or $p_Y(y \mid +1)$? Explain why.

  (b) Compare the conditional pmfs in Problems 5.11(ii) and (iii) and explain which of these two cases is "more random."

  (c) Find $E[Y \mid X = x]$ and $E[X \mid Y = y]$ in Problems 5.11(i), (ii), (iii); then find $E[X]$ and $E[Y]$.

  (d) Find $E[Y^2 \mid X = x]$ and $E[X^2 \mid Y = y]$ in Problems 5.11(i), (ii), (iii); then find VAR$[X]$ and VAR$[Y]$.

5.78. (a) Find the conditional pmf of N_1 given N_2 in Problem 5.14.

  (b) Find $P[N_1 = k \mid N_2 = 2k]$ for $k = 5, 10, 20$. *Hint:* Use Stirling's fromula.

  (c) Find $E[N_1 \mid N_2 = k]$, then find $E[N_1]$.

5.79. In Example 5.30, let Y be the number of defects inside the region R and let Z be the number of defects outside the region.

  (a) Find the pmf of Z given Y.

  (b) Find the joint pmf of Y and Z.

  (c) Are Y and Z independent random variables? Is the result intuitive?

5.80. (a) Find $f_Y(y \mid x)$ in Problem 5.26.

  (b) Find $P[Y > X \mid x]$.

  (c) Find $P[Y > X]$ using part b.

  (d) Find $E[Y \mid X = x]$.

5.81. (a) Find $f_Y(y \mid x)$ in Problem 5.28(i).

  (b) Find $E[Y \mid X = x]$ and $E[Y]$.

  (c) Repeat parts a and b of Problem 5.28(ii).

  (d) Repeat parts a and b of Problem 5.28(iii).

5.82. (a) Find $f_Y(y \mid x)$ in Example 5.27.

  (b) Find $E[Y \mid X = x]$.

  (c) Find $E[Y]$.

  (d) Find $E[XY \mid X = x]$.

  (e) Find $E[XY]$.

5.83. Find $f_Y(y \mid x)$ and $f_X(x \mid y)$ for the jointly Gaussian pdf in Problem 5.34.

5.84. (a) Find $f_X(t \mid N = n)$ in Problem 5.23.

  (b) Find $E[X^t \mid N = n]$.

  (c) Find the value of n that maximizes $P[N = n \mid t < X < t + dt]$.

5.85. (a) Find $p_Y(y \mid x)$ and $p_X(x \mid y)$ in Problem 5.12.

    (b) Find $E[Y \mid X = x]$.

    (c) Find $E[XY \mid X = x]$ and $E[XY]$.

5.86. A customer enters a store and is equally likely to be served by one of three clerks. The time taken by clerk 1 is a constant random variable with mean two minutes; the time for clerk 2 is exponentially distributed with mean two minutes; and the time for clerk 3 is Pareto distributed with mean two minutes and $\alpha = 2.5$.

    (a) Find the pdf of T, the time taken to service a customer.

    (b) Find $E[T]$ and $\mathrm{VAR}[T]$.

5.87. A message requires N time units to be transmitted, where N is a geometric random variable with pmf $p_i = (1 - a)a^{i-1}, i = 1, 2, \ldots$. A single new message arrives during a time unit with probability p, and no messages arrive with probability $1 - p$. Let K be the number of new messages that arrive during the transmission of a single message.

    (a) Find $E[K]$ and $\mathrm{VAR}[K]$ using conditional expectation.

    (b) Find the pmf of K. *Hint:* $(1 - \beta)^{-(k+1)} = \sum_{n=k}^{\infty} \binom{n}{k} \beta^{n-k}$.

    (c) Find the conditional pmf of N given $K = k$.

    (d) Find the value of n that maximizes $P[N = n \mid X = k]$.

5.88. The number of defects in a VLSI chip is a Poisson random variable with rate r. However, r is itself a gamma random variable with parameters α and λ.

    (a) Use conditional expectation to find $E[N]$ and $\mathrm{VAR}[N]$.

    (b) Find the pmf for N, the number of defects.

5.89. (a) In Problem 5.35, find the conditional pmf of the input X of the communication channel given that the output is in the interval $y < Y \le y + dy$.

    (b) Find the value of X that is more probable given $y < Y \le y + dy$.

    (c) Find an expression for the probability of error if we use the result of part b to decide what the input to the channel was.

Section 5.8: Functions of Two Random Variables

5.90. Two toys are started at the same time each with a different battery. The first battery has a lifetime that is exponentially distributed with mean 100 minutes; the second battery has a Rayleigh-distributed lifetime with mean 100 minutes.

    (a) Find the cdf to the time T until the battery in a toy first runs out.

    (b) Suppose that both toys are still operating after 100 minutes. Find the cdf of the time T_2 that subsequently elapses until the battery in a toy first runs out.

    (c) In part b, find the cdf of the total time that elapses until a battery first fails.

5.91. (a) Find the cdf of the time that elapses until both batteries run out in Problem 5.90a.

    (b) Find the cdf of the remaining time until both batteries run out in Problem 5.90b.

5.92. Let K and N be independent random variables with nonnegative integer values.

    (a) Find an expression for the pmf of $M = K + N$.

    (b) Find the pmf of M if K and N are binomial random variables with parameters (k, p) and (n, p).

    (c) Find the pmf of M if K and N are Poisson random variables with parameters α_1 and α_2, respectively.

5.93. The number X of goals the Bulldogs score against the Flames has a geometric distribution with mean 2; the number of goals Y that the Flames score against the Bulldogs is also geometrically distributed but with mean 4.

    (a) Find the pmf of the $Z = X - Y$. Assume X and Y are independent.

    (b) What is the probability that the Bulldogs beat the Flames? Tie the Flames?

    (c) Find $E[Z]$.

5.94. Passengers arrive at an airport taxi stand every minute according to a Bernoulli random variable. A taxi will not leave until it has two passengers.

    (a) Find the pmf until the time T when the taxi has two passengers.

    (b) Find the pmf for the time that the first customer waits.

5.95. Let X and Y be independent random variables that are uniformly distributed in the interval $[0, 1]$. Find the pdf of $Z = XY$.

5.96. Let X_1, X_2, and X_3 be independent and uniformly distributed in $[-1, 1]$.

    (a) Find the cdf and pdf of $Y = X_1 + X_2$.

    (b) Find the cdf of $Z = Y + X_3$.

5.97. Let X and Y be independent random variables with gamma distributions and parameters (α_1, λ) and (α_2, λ), respectively. Show that $Z = X + Y$ is gamma-distributed with parameters $(\alpha_1 + \alpha_2, \lambda)$. *Hint:* See Eq. (4.59).

5.98. Signals X and Y are independent. X is exponentially distributed with mean 1 and Y is exponentially distributed with mean 1.

    (a) Find the cdf of $Z = |X - Y|$.

    (b) Use the result of part a to find $E[Z]$.

5.99. The random variables X and Y have the joint pdf

$$f_{X,Y}(x, y) = e^{-(x+y)} \qquad \text{for } 0 < y < x < 1.$$

Find the pdf of $Z = X + Y$.

5.100. Let X and Y be independent Rayleigh random variables with parameters $\alpha = \beta = 1$. Find the pdf of $Z = X/Y$.

5.101. Let X and Y be independent Gaussian random variables that are zero mean and unit variance. Show that $Z = X/Y$ is a Cauchy random variable.

5.102. Find the joint cdf of $W = \min(X, Y)$ and $Z = \max(X, Y)$ if X and Y are independent and X is uniformly distributed in $[0, 1]$ and Y is uniformly distributed in $[0, 1]$.

5.103. Find the joint cdf of $W = \min(X, Y)$ and $Z = \max(X, Y)$ if X and Y are independent exponential random variables with the same mean.

5.104. Find the joint cdf of $W = \min(X, Y)$ and $Z = \max(X, Y)$ if X and Y are the independent Pareto random variables with the same distribution.

5.105. Let $W = X + Y$ and $Z = X - Y$.

    (a) Find an expression for the joint pdf of W and Z.

    (b) Find $f_{W,Z}(z, w)$ if X and Y are independent exponential random variables with parameter $\lambda = 1$.

    (c) Find $f_{W,Z}(z, w)$ if X and Y are independent Pareto random variables with the same distribution.

5.106. The pair (X, Y) is uniformly distributed in a ring centered about the origin and inner and outer radii $r_1 < r_2$. Let R and Θ be the radius and angle corresponding to (X, Y). Find the joint pdf of R and Θ.

5.107. Let X and Y be independent, zero-mean, unit-variance Gaussian random variables. Let $V = aX + bY$ and $W = cX + eY$.

    (a) Find the joint pdf of V and W, assuming the transformation matrix A is invertible.

    (b) Suppose A is not invertible. What is the joint pdf of V and W?

5.108. Let X and Y be independent Gaussian random variables that are zero mean and unit variance. Let $W = X^2 + Y^2$ and let $\Theta = \tan^{-1}(Y/X)$. Find the joint pdf of W and Θ.

5.109. Let X and Y be the random variables introduced in Example 5.4. Let $R = (X^2 + Y^2)^{1/2}$ and let $\Theta = \tan^{-1}(Y/X)$.

    (a) Find the joint pdf of R and Θ.

    (b) What is the joint pdf of X and Y?

Section 5.9: Pairs of Jointly Gaussian Variables

5.110. Let X and Y be jointly Gaussian random variables with pdf

$$f_{X,Y}(x, y) = \frac{\exp\{-2x^2 - y^2/2\}}{2\pi c} \quad \text{for all } x, y.$$

Find $\text{VAR}[X]$, $\text{VAR}[Y]$, and $\text{COV}(X, Y)$.

5.111. Let X and Y be jointly Gaussian random variables with pdf

$$f_{X,Y}(x, y) = \frac{\exp\left\{\frac{-1}{2}[x^2 + 4y^2 - 3xy + 3y - 2x + 1]\right\}}{2\pi} \quad \text{for all } x, y.$$

Find $E[X], E[Y], \text{VAR}[X], \text{VAR}[Y]$, and $\text{COV}(X, Y)$.

5.112. Let X and Y be jointly Gaussian random variables with $E[Y] = 0$, $\sigma_1 = 1$, $\sigma_2 = 2$, and $E[X \mid Y] = Y/4 + 1$. Find the joint pdf of X and Y.

5.113. Let X and Y be zero-mean, independent Gaussian random variables with $\sigma^2 = 1$.

    (a) Find the value of r for which the probability that (X, Y) falls inside a circle of radius r is $1/2$.

    (b) Find the conditional pdf of (X, Y) given that (X, Y) is not inside a ring with inner radius r_1 and outer radius r_2.

5.114. Use a plotting program (as provided by Octave or MATLAB) to show the pdf for jointly Gaussian zero-mean random variables with the following parameters:

    (a) $\sigma_1 = 1, \sigma_2 = 1, \rho = 0$.

    (b) $\sigma_1 = 1, \sigma_2 = 1, \rho = 0.8$.

    (c) $\sigma_1 = 1, \sigma_2 = 1, \rho = -0.8$.

    (d) $\sigma_1 = 1, \sigma_2 = 2, \rho = 0$.

    (e) $\sigma_1 = 1, \sigma_2 = 2, \rho = 0.8$.

    (f) $\sigma_1 = 1, \sigma_2 = 10, \rho = 0.8$.

5.115. Let X and Y be zero-mean, jointly Gaussian random variables with $\sigma_1 = 1$, $\sigma_2 = 2$, and correlation coefficient ρ.

    (a) Plot the principal axes of the constant-pdf ellipse of (X, Y).

    (b) Plot the conditional expectation of Y given $X = x$.

    (c) Are the plots in parts a and b the same or different? Why?

5.116. Let X and Y be zero-mean, unit-variance jointly Gaussian random variables for which $\rho = 1$. Sketch the joint cdf of X and Y. Does a joint pdf exist?

5.117. Let $h(x, y)$ be a joint Gaussian pdf for zero-mean, unit-variance Gaussian random variables with correlation coefficient ρ_1. Let $g(x, y)$ be a joint Gaussian pdf for zero-mean, unit-variance Gaussian random variables with correlation coefficient $\rho_2 \neq \rho_1$. Suppose the random variables X and Y have joint pdf

$$f_{X,Y}(x, y) = \{h(x, y) + g(x, y)\}/2.$$

(a) Find the marginal pdf for X and for Y.

(b) Explain why X and Y are not jointly Gaussian random variables.

5.118. Use conditional expectation to show that for X and Y zero-mean, jointly Gaussian random variables, $E[X^2Y^2] = E[X^2]E[Y^2] + 2E[XY]^2$.

5.119. Let $\mathbf{X} = (X, Y)$ be the zero-mean jointly Gaussian random variables in Problem 5.110. Find a transformation A such that $\mathbf{Z} = A\mathbf{X}$ has components that are zero-mean, unit-variance Gaussian random variables.

5.120. In Example 5.47, suppose we estimate the value of the signal X from the noisy observation Y by:

$$\hat{X} = \frac{1}{1 + \sigma_N^2/\sigma_X^2} Y.$$

(a) Evaluate the mean square estimation error: $E[(X - \hat{X})^2]$.

(b) How does the estimation error in part a vary with signal-to-noise ratio σ_X/σ_N?

Section 5.10: Generating Independent Gaussian Random Variables

5.121. Find the inverse of the cdf of the Rayleigh random variable to derive the transformation method for generating Rayleigh random variables. Show that this method leads to the same algorithm that was presented in Section 5.10.

5.122. Reproduce the results presented in Example 5.49.

5.123. Consider the two-dimensional modem in Problem 5.36.

(a) Generate 10,000 discrete random variables uniformly distributed in the set $\{1, 2, 3, 4\}$. Assign each outcome in this set to one of the signals $\{(1, 1), (1, -1), (-1, 1), (-1, -1)\}$. The sequence of discrete random variables then produces a sequence of 10,000 signal points $\mathbf{X}$.

(b) Generate 10,000 noise pairs $\mathbf{N}$ of independent zero-mean, unit-variance jointly Gaussian random variables.

(c) Form the sequence of 10,000 received signals $\mathbf{Y} = (Y_1, Y_2) = \mathbf{X} + \mathbf{N}$.

(d) Plot the scattergram of received signal vectors. Is the plot what you expected?

(e) Estimate the transmitted signal by the quadrant that $\mathbf{Y}$ falls in: $\hat{X} = (\text{sgn}(Y_1), \text{sgn}(Y_2))$.

(f) Compare the estimates with the actually transmitted signals to estimate the probability of error.

5.124. Generate a sequence of 1000 pairs of independent zero-mean Gaussian random variables, where X has variance 2 and N has variance 1. Let $Y = X + N$ be the noisy signal from Example 5.47.

(a) Estimate X using the estimator in Problem 5.120, and calculate the sequence of estimation errors.

(b) What is the pdf of the estimation error?

(c) Compare the mean, variance, and relative frequencies of the estimation error with the result from part b.

5.125. Let $X_1, X_2, \ldots, X_{1000}$ be a sequence of zero-mean, unit-variance independent Gaussian random variables. Suppose that the sequence is "smoothed" as follows:

$$Y_n = (X_n + X_{N-1})/2 \text{ where } X_0 = 0.$$

   (a) Find the pdf of (Y_n, Y_{n+1}).

   (b) Generate the sequence of X_n and the corresponding sequence Y_n. Plot the scattergram of (Y_n, Y_{n+1}). Does it agree with the result from part a?

   (c) Repeat parts a and b for $Z_n = (X_n - X_{N-1})/2$.

5.126. Let X and Y be independent, zero-mean, unit-variance Gaussian random variables. Find the linear transformation to generate jointly Gaussian random variables with means m_1, m_2, variances σ_1^2, σ_2^2, and correlation coefficient ρ. *Hint:* Use the conditional pdf in Eq. (5.64).

5.127. **(a)** Use the method developed in Problem 5.126 to generate 1000 pairs of jointly Gaussian random variables with $m_1 = 1, m_2 = -1$, variances $\sigma_1^2 = 1, \sigma_2^2 = 2$, and correlation coefficient $\rho = -1/2$.

   (b) Plot a two-dimensional scattergram of the 1000 pairs and compare to equal-pdf contour lines for the theoretical pdf.

5.128. Let H and W be the height and weight of adult males. Studies have shown that H (in cm) and $V = \ln W$ (W in kg) are jointly Gaussian with parameters $m_H = 174$ cm, $m_V = 4.4$, $\sigma_H^2 = 42.36, \sigma_V^2 = 0.021$, and $\text{COV}(H, V) = 0.458$.

   (a) Use the method in part a to generate 1000 pairs (H, V). Plot a scattergram to check the joint pdf.

   (b) Convert the (H, V) pairs into (H, W) pairs.

   (c) Calculate the body mass index for each outcome, and estimate the proportion of the population that is underweight, normal, overweight, or obese. (See Problem 5.6.)

Problems Requiring Cumulative Knowledge

5.129. The random variables X and Y have joint pdf:

$$f_{X,Y}(x, y) = c \sin(x + y) \qquad 0 \le x \le \pi/2, 0 \le y \le \pi/2.$$

   (a) Find the value of the constant c.

   (b) Find the joint cdf of X and Y.

   (c) Find the marginal pdf's of X and of Y.

   (d) Find the mean, variance, and covariance of X and Y.

5.130. An inspector selects an item for inspection according to the outcome of a coin flip: The item is inspected if the outcome is heads. Suppose that the time between item arrivals is an exponential random variable with mean one. Assume the time to inspect an item is a constant value t.

   (a) Find the pmf for the number of item arrivals between consecutive inspections.

   (b) Find the pdf for the time X between item inspections. *Hint:* Use conditional expectation.

   (c) Find the value of p, so that with a probability of 90% an inspection is completed before the next item is selected for inspection.

5.131. The lifetime X of a device is an exponential random variable with mean $= 1/R$. Suppose that due to irregularities in the production process, the parameter R is random and has a gamma distribution.

   (a) Find the joint pdf of X and R.

   (b) Find the pdf of X.

   (c) Find the mean and variance of X.

5.132. Let X and Y be samples of a random signal at two time instants. Suppose that X and Y are independent zero-mean Gaussian random variables with the same variance. When signal "0" is present the variance is σ_0^2, and when signal "1" is present the variance is $\sigma_1^2 > \sigma_0^2$. Suppose signals 0 and 1 occur with probabilities p and $1 - p$, respectively. Let $R^2 = X^2 + Y^2$ be the total energy of the two observations.

   (a) Find the pdf of R^2 when signal 0 is present; when signal 1 is present. Find the pdf of R^2.

   (b) Suppose we use the following "signal detection" rule: If $R^2 > T$, then we decide signal 1 is present; otherwise, we decide signal 0 is present. Find an expression for the probability of error in terms of T.

   (c) Find the value of T that minimizes the probability of error.

5.133. Let $U_0, U_1, \ldots$ be a sequence of independent zero-mean, unit-variance Gaussian random variables. A "low-pass filter" takes the sequence U_i and produces the output sequence $X_n = (U_n + U_{n-1})/2$, and a "high-pass filter" produces the output sequence $Y_n = (U_n - U_{n-1})/2$.

   (a) Find the joint pdf of X_n and X_{n-1}; of X_n and $X_{n+m}, m > 1$.

   (b) Repeat part a for Y_n.

   (c) Find the joint pdf of X_n and Y_m.

Vector Random Variables

In the previous chapter we presented methods for dealing with two random variables. In this chapter we extend these methods to the case of n random variables in the following ways:

- By representing n random variables as a vector, we obtain a compact notation for the joint pmf, cdf, and pdf as well as marginal and conditional distributions.
- We present a general method for finding the pdf of transformations of vector random variables.
- Summary information of the distribution of a vector random variable is provided by an expected value vector and a covariance matrix.
- We use linear transformations and characteristic functions to find alternative representations of random vectors and their probabilities.
- We develop optimum estimators for estimating the value of a random variable based on observations of other random variables.
- We show how jointly Gaussian random vectors have a compact and easy-to-work-with pdf and characteristic function.

6.1 VECTOR RANDOM VARIABLES

The notion of a random variable is easily generalized to the case where several quantities are of interest. A **vector random variable X** is a function that assigns a vector of real numbers to each outcome ζ in S, the sample space of the random experiment. We use uppercase boldface notation for vector random variables. By convention **X** is a column vector (n rows by 1 column), so the vector random variable with components $X_1, X_2, \ldots, X_n$ corresponds to

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = [X_1, X_2, \ldots, X_n]^{\mathrm{T}},$$

where "$^{\mathrm{T}}$" denotes the transpose of a matrix or vector. We will sometimes write $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ to save space and omit the transpose unless dealing with matrices. Possible values of the vector random variable are denoted by $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ where x_i corresponds to the value of X_i.

Example 6.1   Arrivals at a Packet Switch

Packets arrive at each of three input ports of a packet switch according to independent Bernoulli trials with $p = 1/2$. Each arriving packet is equally likely to be destined to any of three output ports. Let $\mathbf{X} = (X_1, X_2, X_3)$ where X_i is the total number of packets arriving for output port i. $\mathbf{X}$ is a vector random variable whose values are determined by the pattern of arrivals at the input ports.

Example 6.2   Joint Poisson Counts

A random experiment consists of finding the number of defects in a semiconductor chip and identifying their locations. The outcome of this experiment consists of the vector $\zeta = (n, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$, where the first component specifies the total number of defects and the remaining components specify the coordinates of their location. Suppose that the chip consists of M regions. Let $N_1(\zeta), N_2(\zeta), \ldots, N_M(\zeta)$ be the number of defects in each of these regions, that is, $N_k(\zeta)$ is the number of $\mathbf{y}$'s that fall in region k. The vector $\mathbf{N}(\zeta) = (N_1, N_2, \ldots, N_M)$ is then a vector random variable.

Example 6.3   Samples of an Audio Signal

Let the outcome ζ of a random experiment be an audio signal $X(t)$. Let the random variable $X_k = X(kT)$ be the sample of the signal taken at time kT. An MP3 codec processes the audio in blocks of n samples $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. $\mathbf{X}$ is a vector random variable.

6.1.1   Events and Probabilities

Each event A involving $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ has a corresponding region in an n-dimensional real space R^n. As before, we use "rectangular" product-form sets in R^n as building blocks. For the n-dimensional random variable $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, we are interested in events that have the **product form**

$$A = \{X_1 \text{ in } A_1\} \cap \{X_2 \text{ in } A_2\} \cap \cdots \cap \{X_n \text{ in } A_n\}, \qquad (6.1)$$

where each A_k is a one-dimensional event (i.e., subset of the real line) that involves X_k only. The event A occurs when all of the events $\{X_k \text{ in } A_k\}$ occur jointly.

We are interested in obtaining the probabilities of these product-form events:

$$P[A] = P[\mathbf{X} \in A] = P[\{X_1 \text{ in } A_1\} \cap \{X_2 \text{ in } A_2\} \cap \cdots \cap \{X_n \text{ in } A_n\}]$$
$$\triangleq P[X_1 \text{ in } A_1, X_2 \text{ in } A_2, \ldots, X_n \text{ in } A_n]. \qquad (6.2)$$

In principle, the probability in Eq. (6.2) is obtained by finding the probability of the equivalent event in the underlying sample space, that is,

$$P[A] = P[\{\zeta \text{ in } S : \mathbf{X}(\zeta) \text{ in } A\}]$$

$$= P[\{\zeta \text{ in } S : X_1(\zeta) \in A_1, X_2(\zeta) \in A_2, \ldots, X_n(\zeta) \in A_n\}]. \qquad (6.3)$$

Equation (6.2) forms the basis for the definition of the n-dimensional joint probability mass function, cumulative distribution function, and probability density function. The probabilities of other events can be expressed in terms of these three functions.

6.1.2   Joint Distribution Functions

The **joint cumulative distribution function** of $X_1, X_2, \ldots, X_n$ is defined as the probability of an n-dimensional semi-infinite rectangle associated with the point $(x_1, \ldots, x_n)$:

$$F_{\mathbf{X}}(\mathbf{x}) \triangleq F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = P[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n]. \quad (6.4)$$

The joint cdf is defined for discrete, continuous, and random variables of mixed type. The probability of product-form events can be expressed in terms of the joint cdf.

   The joint cdf generates a family of **marginal cdf's** for subcollections of the random variables $X_1, \ldots, X_n$. These marginal cdf's are obtained by setting the appropriate entries to $+\infty$ in the joint cdf in Eq. (6.4). For example:

Joint cdf for $X_1, \ldots, X_{n-1}$ is given by $F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_{n-1}, \infty)$ and
Joint cdf for X_1 and X_2 is given by $F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \infty, \ldots, \infty)$.

Example 6.4

A radio transmitter sends a signal to a receiver using three paths. Let X_1, X_2, and X_3 be the signals that arrive at the receiver along each path. Find $P[\max(X_1, X_2, X_3) \leq 5]$.

   The maximum of three numbers is less than 5 if and only if each of the three numbers is less than 5; therefore

$$P[A] = P[\{X_1 \leq 5\} \cap \{X_2 \leq 5\} \cap \{X_3 \leq 5\}]$$

$$= F_{X_1, X_2, X_3}(5, 5, 5).$$

The **joint probability mass function** of n discrete random variables is defined by

$$p_X(\mathbf{x}) \triangleq p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]. \quad (6.5)$$

The probability of any n-dimensional event A is found by summing the pmf over the points in the event

$$P[\mathbf{X} \text{ in } A] = \sum_{\mathbf{x} \text{ in } A} \cdots \sum p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n). \qquad (6.6)$$

The joint pmf generates a family of **marginal pmf's** that specifies the joint probabilities for subcollections of the n random variables. For example, the one-dimensional pmf of X_j is found by adding the joint pmf over all variables other than x_j:

$$p_{X_j}(x_j) = P[X_j = x_j] = \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} p_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n). \quad (6.7)$$

The two-dimensional joint pmf of any pair X_j and X_k is found by adding the joint pmf over all $n - 2$ other variables, and so on. Thus, the marginal pmf for $X_1, \ldots, X_{n-1}$ is given by

$$p_{X_1, \ldots, X_{n-1}}(x_1, x_2, \ldots, x_{n-1}) = \sum_{x_n} p_{X_1, \ldots, X_n}(x_1, x_2, \ldots, x_n). \quad (6.8)$$

A family of **conditional pmf's** is obtained from the joint pmf by conditioning on different subcollections of the random variables. For example, if $p_{X_1, \ldots, X_{n-1}}(x_1, \ldots, x_{n-1}) > 0$:

$$p_{X_n}(x_n \mid x_1, \ldots, x_{n-1}) = \frac{p_{X_1, \ldots, X_n}(x_1, \ldots, x_n)}{p_{X_1, \ldots, X_{n-1}}(x_1, \ldots, x_{n-1})}. \quad (6.9a)$$

Repeated applications of Eq. (6.9a) yield the following very useful expression:

$$p_{X_1, \ldots, X_n}(x_1, \ldots, x_n) =$$

$$p_{X_n}(x_n \mid x_1, \ldots, x_{n-1}) p_{X_{n-1}}(x_{n-1} \mid x_1, \ldots, x_{n-2}) \cdots p_{X_2}(x_2 \mid x_1) p_{X_1}(x_1). \quad (6.9b)$$

Example 6.5 Arrivals at a Packet Switch

Find the joint pmf of $\mathbf{X} = (X_1, X_2, X_3)$ in Example 6.1. Find $P[X_1 > X_3]$.

Let N be the total number of packets arriving in the three input ports. Each input port has an arrival with probability $p = 1/2$, so N is binomial with pmf:

$$p_N(n) = \binom{3}{n} \frac{1}{2^3} \quad \text{for} \quad 0 \le n \le 3.$$

Given $N = n$, the number of packets arriving for each output port has a multinomial distribution:

$$p_{X_1, X_2, X_3}(i, j, k \mid i + j + k = n) = \begin{cases} \dfrac{n!}{i! \, j! \, k!} \dfrac{1}{3^n} & \text{for} \quad i + j + k = n, i \ge 0, j \ge 0, k \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$

The joint pmf of $\mathbf{X}$ is then:

$$p_{\mathbf{X}}(i, j, k) = p_{\mathbf{X}}(i, j, k \mid n) \binom{3}{n} \frac{1}{2^3} \quad \text{for} \quad i \ge 0, j \ge 0, k \ge 0, i + j + k = n \le 3.$$

The explicit values of the joint pmf are:

$$p_{\mathbf{X}}(0, 0, 0) = \frac{0!}{0! \, 0! \, 0!} \frac{1}{3^0} \binom{3}{0} \frac{1}{2^3} = \frac{1}{8}$$

$$p_{\mathbf{X}}(1,0,0) = p_{\mathbf{X}}(0,1,0) = p_{\mathbf{X}}(0,0,1) = \frac{1!}{0!\,0!\,1!}\frac{1}{3^1}\binom{3}{1}\frac{1}{2^3} = \frac{3}{24}$$

$$p_{\mathbf{X}}(1,1,0) = p_{\mathbf{X}}(1,0,1) = p_{\mathbf{X}}(0,1,1) = \frac{2!}{0!\,1!\,1!}\frac{1}{3^2}\binom{3}{2}\frac{1}{2^3} = \frac{6}{72}$$

$$p_{\mathbf{X}}(2,0,0) = p_{\mathbf{X}}(0,2,0) = p_{\mathbf{X}}(0,0,2) = 3/72$$

$$p_{\mathbf{X}}(1,1,1) = 6/216$$

$$p_{\mathbf{X}}(0,1,2) = p_{\mathbf{X}}(0,2,1) = p_{\mathbf{X}}(1,0,2) = p_{\mathbf{X}}(1,2,0) = p_{\mathbf{X}}(2,0,1) = p_{\mathbf{X}}(2,1,0) = 3/216$$

$$p_{\mathbf{X}}(3,0,0) = p_{\mathbf{X}}(0,3,0) = p_{\mathbf{X}}(0,0,3) = 1/216.$$

Finally:

$$\begin{aligned}
P[X_1 > X_3] &= p_{\mathbf{X}}(1,0,0) + p_{\mathbf{X}}(1,1,0) + p_{\mathbf{X}}(2,0,0) + p_{\mathbf{X}}(1,2,0) \\
&\quad + p_{\mathbf{X}}(2,0,1) + p_{\mathbf{X}}(2,1,0) + p_{\mathbf{X}}(3,0,0) \\
&= 8/27.
\end{aligned}$$

We say that the random variables $X_1, X_2, \ldots, X_n$ are **jointly continuous random variables** if the probability of any n-dimensional event A is given by an n-dimensional integral of a probability density function:

$$P[\mathbf{X} \text{ in } A] = \int_{\substack{\cdots \\ \mathbf{x} \text{ in } A}} \int f_{X_1, \ldots, X_n}(x_1', \ldots, x_n')\, dx_1' \ldots dx_n', \tag{6.10}$$

where $f_{X_1, \ldots, X_n}(x_1, \ldots, x_n)$ is the **joint probability density function**.

The joint cdf of $\mathbf{X}$ is obtained from the joint pdf by integration:

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1, \ldots, X_n}(x_1', \ldots, x_n')\, dx_1' \ldots dx_n'. \tag{6.11}$$

The joint pdf (if the derivative exists) is given by

$$f_{\mathbf{X}}(\mathbf{x}) \triangleq f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_{X_1, \ldots, X_n}(x_1, \ldots, x_n). \tag{6.12}$$

A family of **marginal pdf's** is associated with the joint pdf in Eq. (6.12). The marginal pdf for a subset of the random variables is obtained by integrating the other variables out. For example, the marginal pdf of X_1 is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1, X_2, \ldots, X_n}(x_1, x_2', \ldots, x_n')\, dx_2' \ldots dx_n'. \tag{6.13}$$

As another example, the marginal pdf for $X_1, \ldots, X_{n-1}$ is given by

$$f_{X_1, \ldots, X_{n-1}}(x_1, \ldots, x_{n-1}) = \int_{-\infty}^{\infty} f_{X_1, \ldots, X_n}(x_1, \ldots, x_{n-1}, x_n')\, dx_n'. \tag{6.14}$$

A family of **conditional pdf's** is also associated with the joint pdf. For example, the pdf of X_n given the values of $X_1, \ldots, X_{n-1}$ is given by

$$f_{X_n}(x_n | x_1, \ldots, x_{n-1}) = \frac{f_{X_1, \ldots, X_n}(x_1, \ldots, x_n)}{f_{X_1, \ldots, X_{n-1}}(x_1, \ldots, x_{n-1})} \tag{6.15a}$$

if $f_{X_1,\ldots,X_{n-1}}(x_1,\ldots,x_{n-1}) > 0$.

Repeated applications of Eq. (6.15a) yield an expression analogous to Eq. (6.9b):

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) =$$
$$f_{X_n}(x_n|x_1,\ldots,x_{n-1})f_{X_{n-1}}(x_{n-1}|x_1,\ldots,x_{n-2})\ldots f_{X_2}(x_2|x_1)f_{X_1}(x_1). \quad (6.15b)$$

Example 6.6

The random variables X_1, X_2, and X_3 have the joint Gaussian pdf

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{e^{-(x_1^2+x_2^2-\sqrt{2}\,x_1x_2+\frac{1}{2}x_3^2)}}{2\pi\sqrt{\pi}}.$$

Find the marginal pdf of X_1 and X_3. Find the conditional pdf of X_2 given X_1 and X_3.

The marginal pdf for the pair X_1 and X_3 is found by integrating the joint pdf over x_2:

$$f_{X_1,X_3}(x_1, x_3) = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-(x_1^2+x_2^2-\sqrt{2}x_1x_2)}}{2\pi/\sqrt{2}}\,dx_2.$$

The above integral was carried out in Example 5.18 with $\rho = -1/\sqrt{2}$. By substituting the result of the integration above, we obtain

$$f_{X_1,X_3}(x_1, x_3) = \frac{e^{-x_3^2/2}}{\sqrt{2\pi}} \frac{e^{-x_1^2/2}}{\sqrt{2\pi}}.$$

Therefore X_1 and X_3 are independent zero-mean, unit-variance Gaussian random variables.

The conditional pdf of X_2 given X_1 and X_3 is:

$$f_{X_2}(x_2|x_1, x_3) = \frac{e^{-(x_1^2+x_2^2-\sqrt{2}x_1x_2+\frac{1}{2}x_3^2)}}{2\pi\sqrt{\pi}} \frac{\sqrt{2\pi}\sqrt{2\pi}}{e^{-x_3^2/2}e^{-x_1^2/2}}$$

$$= \frac{e^{-(\frac{1}{2}x_1^2+x_2^2-\sqrt{2}x_1x_2)}}{\sqrt{\pi}} = \frac{e^{-(x_2-x_1/\sqrt{2}x_1)^2}}{\sqrt{\pi}}.$$

We conclude that X_2 given and X_3 is a Gaussian random variable with mean $x_1/\sqrt{2}$ and variance 1/2.

Example 6.7   Multiplicative Sequence

Let X_1 be uniform in $[0, 1]$, X_2 be uniform in $[0, X_1]$, and X_3 be uniform in $[0, X_2]$. (Note that X_3 is also the product of three uniform random variables.) Find the joint pdf of **X** and the marginal pdf of X_3.

For $0 < z < y < x < 1$, the joint pdf is nonzero and given by:

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = f_{X_3}(z|x, y)f_{X_2}(y|x)f_{X_1}(x) = \frac{1}{y}\frac{1}{x}1 = \frac{1}{xy}.$$

The joint pdf of X_2 and X_3 is nonzero for $0 < z < y < 1$ and is obtained by integrating x between y and 1:

$$f_{X_2,X_3}(x_2, x_3) = \int_y^1 \frac{1}{xy}\,dx = \frac{1}{y}\ln x\Big|_y^1 = \frac{1}{y}\ln\frac{1}{y}.$$

We obtain the pdf of X_3 by integrating y between z and 1:

$$f_{X_3}(x_3) = -\int_z^1 \frac{1}{y}\ln y\,dy = -\frac{1}{2}(\ln y)^2\Big|_z^1 = \frac{1}{2}(\ln z)^2.$$

Note that the pdf of X_3 is concentrated at the values close to $x = 0$.

6.1.3   Independence

The collection of random variables $X_1, \ldots, X_n$ is **independent** if

$$P[X_1 \text{ in } A_1, X_2 \text{ in } A_2, \ldots, X_n \text{ in } A_n] = P[X_1 \text{ in } A_1]P[X_2 \text{ in } A_2]\ldots P[X_n \text{ in } A_n]$$

for *any* one-dimensional events $A_1, \ldots, A_n$. It can be shown that $X_1, \ldots, X_n$ *are independent if and only if*

$$F_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1)\ldots F_{X_n}(x_n) \qquad (6.16)$$

for all $x_1, \ldots, x_n$. If the random variables are discrete, Eq. (6.16) is equivalent to

$$p_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1)\ldots p_{X_n}(x_n) \qquad \text{for all } x_1, \ldots, x_n.$$

If the random variables are jointly continuous, Eq. (6.16) is equivalent to

$$f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1)\ldots f_{X_n}(x_n)$$

for all $x_1, \ldots, x_n$.

Example 6.8

The n samples $X_1, X_2, \ldots, X_n$ of a noise signal have joint pdf given by

$$f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \frac{e^{-(x_1^2+\ldots+x_n^2)/2}}{(2\pi)^{n/2}} \qquad \text{for all } x_1, \ldots, x_n.$$

It is clear that the above is the product of n one-dimensional Gaussian pdf's. Thus $X_1, \ldots, X_n$ are independent Gaussian random variables.

6.2   FUNCTIONS OF SEVERAL RANDOM VARIABLES

Functions of vector random variables arise naturally in random experiments. For example $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ may correspond to observations from n repetitions of an experiment that generates a given random variable. We are almost always interested in the sample mean and the sample variance of the observations. In another example

$\mathbf{X} = (X_1, X_2, \ldots, X_n)$ may correspond to samples of a speech waveform and we may be interested in extracting features that are defined as functions of $\mathbf{X}$ for use in a speech recognition system.

6.2.1   One Function of Several Random Variables

Let the random variable Z be defined as a function of several random variables:

$$Z = g(X_1, X_2, \ldots, X_n). \tag{6.17}$$

The cdf of Z is found by finding the equivalent event of $\{Z \le z\}$, that is, the set $R_z = \{\mathbf{x}: g(\mathbf{x}) \le z\}$, then

$$F_Z(z) = P[\mathbf{X} \text{ in } R_z] = \int_{\mathbf{x} \text{ in } R_z} \cdots \int f_{X_1, \ldots, X_n}(x_1', \ldots, x_n') \, dx_1' \ldots dx_n'. \tag{6.18}$$

The pdf of Z is then found by taking the derivative of $F_Z(z)$.

Example 6.9   Maximum and Minimum of n Random Variables

Let $W = \max(X_1, X_2, \ldots, X_n)$ and $Z = \min(X_1, X_2, \ldots, X_n)$, where the X_i are independent random variables with the same distribution. Find $F_W(w)$ and $F_Z(z)$.

The maximum of $X_1, X_2, \ldots, X_n$ is less than x if and only if each X_i is less than x, so:

$$F_W(w) = P[\max(X_1, X_2, \ldots, X_n) \le w]$$

$$= P[X_1 \le w]P[X_2 \le w] \ldots P[X_n \le w] = (F_X(w))^n.$$

The minimum of $X_1, X_2, \ldots, X_n$ is greater than x if and only if each X_i is greater than x, so:

$$1 - F_Z(z) = P[\min(X_1, X_2, \ldots, X_n) > z]$$

$$= P[X_1 > z]P[X_2 > z] \ldots P[X_n > z] = (1 - F_X(z))^n$$

and

$$F_Z(z) = 1 - (1 - F_X(z))^n.$$

Example 6.10   Merging of Independent Poisson Arrivals

Web page requests arrive at a server from n independent sources. Source j generates packets with exponentially distributed interarrival times with rate λ_j. Find the distribution of the interarrival times between consecutive requests at the server.

Let the interarrival times for the different sources be given by $X_1, X_2, \ldots, X_n$. Each X_j satisfies the memoryless property, so the time that has elapsed since the last arrival from each source is irrelevant. The time until the next arrival at the multiplexer is then:

$$Z = \min(X_1, X_2, \ldots, X_n).$$

Therefore the pdf of Z is:

$$1 - F_Z(z) = P[\min(X_1, X_2, \ldots, X_n) > z]$$

$$= P[X_1 > z]P[X_2 > z] \ldots P[X_n > z]$$

$$= \left(1 - F_{X_1}(z)\right)\left(1 - F_{X_2}(z)\right)\ldots\left(1 - F_{X_n}(z)\right)$$
$$= e^{-\lambda_1 z}e^{-\lambda_2 z}\ldots e^{-\lambda_n z} = e^{-(\lambda_1+\lambda_2+\cdots+\lambda_n)z}.$$

The interarrival time is an exponential random variable with rate $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Example 6.11   Reliability of Redundant Systems

A computing cluster has n independent redundant subsystems. Each subsystem has an exponentially distributed lifetime with parameter λ. The cluster will operate as long as at least one subsystem is functioning. Find the cdf of the time until the system fails.

Let the lifetime of each subsystem be given by $X_1, X_2, \ldots, X_n$. The time until the last subsystem fails is:

$$W = \max(X_1, X_2, \ldots, X_n).$$

Therefore the cdf of W is:

$$F_W(w) = \left(F_X(w)\right)^n = (1 - e^{-\lambda w})^n = 1 - \binom{n}{1}e^{-\lambda w} + \binom{n}{2}e^{-2\lambda w} + \ldots.$$

6.2.2   Transformations of Random Vectors

Let $X_1, \ldots, X_n$ be random variables in some experiment, and let the random variables $Z_1, \ldots, Z_n$ be defined by a transformation that consists of n functions of $\mathbf{X} = (X_1, \ldots, X_n)$:

$$Z_1 = g_1(\mathbf{X}) \qquad Z_2 = g_2(\mathbf{X}) \qquad \ldots \qquad Z_n = g_n(\mathbf{X}).$$

The joint cdf of $\mathbf{Z} = (Z_1, \ldots, Z_n)$ at the point $\mathbf{z} = (z_1, \ldots, z_n)$ is equal to the probability of the region of $\mathbf{x}$ where $g_k(\mathbf{x}) \leq z_k$ for $k = 1, \ldots, n$:

$$F_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = P[g_1(\mathbf{X}) \leq z_1, \ldots, g_n(\mathbf{X}) \leq z_n]. \tag{6.19a}$$

If $X_1, \ldots, X_n$ have a joint pdf, then

$$F_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = \int\ldots\int_{x':g_k(x')\leq z_k} f_{X_1,\ldots,X_n}(x'_1, \ldots, x'_n)\, dx'_1 \ldots dx'. \tag{6.19b}$$

Example 6.12

Given a random vector $\mathbf{X}$, find the joint pdf of the following transformation:

$$Z_1 = g_1(X_1) = a_1 X_1 + b_1,$$
$$Z_2 = g_2(X_2) = a_2 X_2 + b_2,$$
$$\vdots$$
$$Z_n = g_n(X_n) = a_n X_n + b_n.$$

Note that $Z_k = a_k X_k + b_k, \le z_k$, if and only if $X_k \le (z_k - b_k)/a_k$, if $a_k > 0$, so

$$F_{Z_1, Z_2, \ldots, Z_n}(z_1, z_2, \ldots, z_n) = P\left[X_1 \le \frac{z_1 - b_1}{a_1}, X_2 \le \frac{z_2 - b_2}{a_2}, \ldots, X_n \le \frac{z_n - b_n}{a_n} \right]$$

$$= F_{X_1, X_2, \ldots, X_n}\left(\frac{z_1 - b_1}{a_1}, \frac{z_2 - b_2}{a_2}, \ldots, \frac{z_n - b_n}{a_n} \right)$$

$$f_{Z_1, Z_2, \ldots, Z_n}(z_1, z_2, \ldots, z_n) = \frac{\partial^n}{\partial z_1 \ldots \partial z_n} F_{Z_1, Z_2, \ldots, Z_n}(z_1, z_2, \ldots, z_n)$$

$$= \frac{1}{a_1 \ldots a_n} f_{X_1, X_2, \ldots, X_n}\left(\frac{z_1 - b_1}{a_1}, \frac{z_2 - b_2}{a_2}, \ldots, \frac{z_n - b_n}{a_n} \right).$$

*6.2.3 pdf of General Transformations

We now introduce a general method for finding the pdf of a transformation of n jointly continuous random variables. We first develop the two-dimensional case. Let the random variables V and W be defined by two functions of X and Y:

$$V = g_1(X, Y) \qquad \text{and} \qquad W = g_2(X, Y). \tag{6.20}$$

Assume that the functions $v(x, y)$ and $w(x, y)$ are invertible in the sense that the equations $v = g_1(x, y)$ and $w = g_2(x, y)$ can be solved for x and y, that is,

$$x = h_1(v, w) \text{ and } y = h_2(v, w).$$

The joint pdf of X and Y is found by finding the equivalent event of infinitesimal rectangles. The image of the infinitesimal rectangle is shown in Fig. 6.1(a). The image can be approximated by the parallelogram shown in Fig. 6.1(b) by making the approximation

$$g_k(x + dx, y) \simeq g_k(x, y) + \frac{\partial}{\partial x} g_k(x, y)\, dx \qquad k = 1, 2$$

and similarly for the y variable. The probabilities of the infinitesimal rectangle and the parallelogram are approximately equal, therefore

$$f_{X,Y}(x, y)\, dx\, dy = f_{V,W}(v, w)\, dP$$

and

$$f_{V,W}(v, w) = \frac{f_{X,Y}(h_1(v, w), (h_2(v, w))}{\left| \dfrac{dP}{dxdy} \right|}, \tag{6.21}$$

where dP is the area of the parallelogram. By analogy with the case of a linear transformation (see Eq. 5.59), we can match the derivatives in the above approximations with the coefficients in the linear transformations and conclude that the
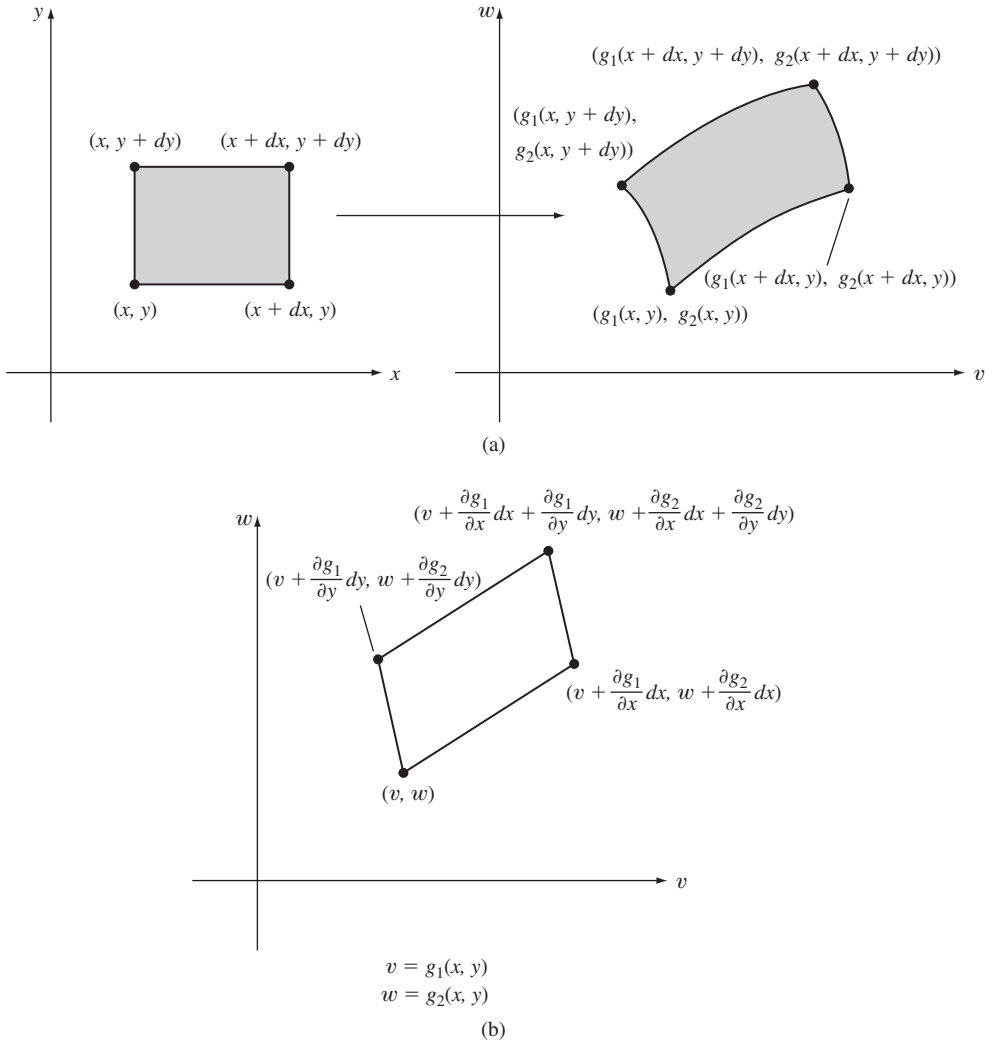
FIGURE 6.1
(a) Image of an infinitesimal rectangle under general transformation. (b) Approximation of image by a parallelogram.

"stretch factor" at the point (v, w) is given by the determinant of a matrix of partial derivatives:

$$J(x, y) = \det \begin{bmatrix} \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \\[2mm] \dfrac{\partial w}{\partial x} & \dfrac{\partial w}{\partial y} \end{bmatrix}.$$

The determinant $J(x, y)$ is called the **Jacobian** of the transformation. The Jacobian of the inverse transformation is given by

$$J(v, w) = \det \begin{bmatrix} \dfrac{\partial x}{\partial v} & \dfrac{\partial x}{\partial w} \\[2mm] \dfrac{\partial y}{\partial v} & \dfrac{\partial y}{\partial w} \end{bmatrix}.$$

It can be shown that

$$|J(v, w)| = \frac{1}{|J(x, y)|}.$$

We therefore conclude that the joint pdf of V and W can be found using either of the following expressions:

$$f_{V,W}(v, w) = \frac{f_{X,Y}(h_1(v, w), (h_2(v, w))}{|J(x, y)|} \tag{6.22a}$$

$$= f_{X,Y}(h_1(v, w), (h_2(v, w))|J(v, w)|. \tag{6.22b}$$

It should be noted that Eq. (6.21) is applicable even if Eq. (6.20) has more than one solution; the pdf is then equal to the sum of terms of the form given by Eqs. (6.22a) and (6.22b), with each solution providing one such term.

Example 6.13

Server 1 receives m Web page requests and server 2 receives k Web page requests. Web page transmission times are exponential random variables with mean $1/\mu$. Let X be the total time to transmit files from server 1 and let Y be the total time for server 2. Find the joint pdf for T, the total transmission time, and W, the proportion of the total transmission time contributed by server 1:

$$T = X + Y \quad \text{and} \quad W = \frac{X}{X + Y}.$$

From Chapter 4, the sum of j independent exponential random variables is an Erlang random variable with parameters j and μ. Therefore X and Y are independent Erlang random variables with parameters m and μ, and k and μ, respectively:

$$f_X(x) = \frac{\mu e^{-\mu x}(\mu x)^{m-1}}{(m - 1)!} \quad \text{and} \quad f_Y(y) = \frac{\mu e^{-\mu y}(\mu y)^{k-1}}{(k - 1)!}.$$

We solve for X and Y in terms of T and W:

$$X = TW \quad \text{and} \quad Y = T(1 - W).$$

The Jacobian of the transformation is:

$$J(x, y) = \det \begin{bmatrix} 1 & 1 \\[2mm] \dfrac{y}{(x + y)^2} & \dfrac{-x}{(x + y)^2} \end{bmatrix}$$

$$= \frac{-x}{(x + y)^2} - \frac{y}{(x + y)^2} = \frac{-1}{x + y} = \frac{-1}{t}.$$

The joint pdf of T and W is then:

$$f_{T,W}(t, w) = \frac{1}{|J(x, y)|}\left[\frac{\mu e^{-\mu x}(\mu x)^{m-1}}{(m-1)!}\frac{\mu e^{-\mu y}(\mu y)^{k-1}}{(k-1)!}\right]_{\substack{x=tw \\ y=t(1-w)}}$$

$$= t\frac{\mu e^{-\mu tw}(\mu tw)^{m-1}}{(m-1)!}\frac{\mu e^{-\mu t(1-w)}(\mu t(1-w))^{k-1}}{(k-1)!}$$

$$= \frac{\mu e^{-\mu t}(\mu t)^{m+k-1}}{(m+k-1)!}\frac{(m+k-1)!}{(m-1)!(k-1)!}(w)^{m-1}(1-w)^{k-1}.$$

We see that T and W are independent random variables. As expected, T is Erlang with parameters $m + k$ and μ, since it is the sum of $m + k$ independent Erlang random variables. W is the beta random variable introduced in Chapter 3.

The method developed above can be used even if we are interested in only one function of a random variable. By defining an "auxiliary" variable, we can use the transformation method to find the joint pdf of both random variables, and then we can find the marginal pdf involving the random variable of interest. The following example demonstrates the method.

Example 6.14   Student's *t*-distribution

Let X be a zero-mean, unit-variance Gaussian random variable and let Y be a chi-square random variable with n degrees of freedom. Assume that X and Y are independent. Find the pdf of $V = X/\sqrt{Y/n}$.

Define the auxiliary function of $W = Y$. The variables X and Y are then related to V and W by

$$X = V\sqrt{W/n} \quad \text{and} \quad Y = W.$$

The Jacobian of the inverse transformation is

$$|J(v, w)| = \begin{vmatrix} \sqrt{w/n} & (v/2)\sqrt{wn} \\ 0 & 1 \end{vmatrix} = \sqrt{w/n}.$$

Since $f_{X,Y}(x, y) = f_X(x)f_Y(y)$, the joint pdf of V and W is thus

$$f_{V,W}(v, w) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}\frac{(y/2)^{n/2-1}e^{-y/2}}{2\Gamma(n/2)}|J(v, w)|\Bigg|_{\substack{x = v\sqrt{w/n} \\ y = w}}$$

$$= \frac{(w/2)^{(n-1)/2}e^{-[(w/2)(1+v2/n)]}}{2\sqrt{n\pi}\Gamma(n/2)}.$$

The pdf of V is found by integrating the joint pdf over w:

$$f_V(v) = \frac{1}{2\sqrt{n\pi}\Gamma(n/2)}\int_0^\infty (w/2)^{(n-1)/2}e^{-[(w/2)(1+v2/n)]}\, dw.$$

If we let $w' = (w/2)(v^2/n + 1)$, the integral becomes

$$f_V(v) = \frac{(1 + v^2/n)^{-(n+1)/2}}{\sqrt{n\pi}\Gamma(n/2)}\int_0^\infty (w')^{(n-1)/2}e^{-w'}\, dw'.$$

By noting that the above integral is the gamma function evaluated at $(n + 1)/2$, we finally obtain the *Student's t-distribution*:

$$f_V(v) = \frac{(1 + v^2/n)^{-(n+1)/2}\Gamma((n + 1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)}.$$

This pdf is used extensively in statistical calculations. (See Chapter 8.)

Next consider the problem of finding the joint pdf for n functions of n random variables $\mathbf{X} = (X_1, \ldots, X_n)$:

$$Z_1 = g_1(\mathbf{X}), \quad Z_2 = g_2(\mathbf{X}), \ldots, \quad Z_n = g_n(\mathbf{X}).$$

We assume as before that the set of equations

$$z_1 = g_1(\mathbf{x}), \quad z_2 = g_2(\mathbf{x}), \ldots, \quad z_n = g_n(\mathbf{x}). \tag{6.23}$$

has a unique solution given by

$$x_1 = h_1(\mathbf{x}), \quad x_2 = h_2(\mathbf{x}), \ldots, \quad x_n = h_n(\mathbf{x}).$$

The joint pdf of $\mathbf{Z}$ is then given by

$$f_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = \frac{f_{X_1,\ldots,X_n}(h_1(\mathbf{z}), h_2(\mathbf{z}), \ldots, h_n(\mathbf{z}))}{|J(x_1, x_2, \ldots, x_n)|} \tag{6.24a}$$

$$= f_{X_1,\ldots,X_n}(h_1(\mathbf{z}), h_2(\mathbf{z}), \ldots, h_n(\mathbf{z}))|J(z_1, z_2, \ldots, z_n)|, \tag{6.24b}$$

where $|J(x_1, \ldots, x_n)|$ and $|J(z_1, \ldots, z_n)|$ are the determinants of the transformation and the inverse transformation, respectively,

$$J(x_1, \ldots, x_n) = \det \begin{bmatrix} \dfrac{\partial g_1}{\partial x_1} & \cdots & \dfrac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial g_n}{\partial x_1} & \cdots & \dfrac{\partial g_n}{\partial x_n} \end{bmatrix}$$

and

$$J(z_1, \ldots, z_n) = \det \begin{bmatrix} \dfrac{\partial h_1}{\partial z_1} & \cdots & \dfrac{\partial h_1}{\partial z_n} \\ \vdots & & \vdots \\ \dfrac{\partial h_n}{\partial z_1} & \cdots & \dfrac{\partial h_n}{\partial z_n} \end{bmatrix}.$$

In the special case of a linear transformation we have:

$$\mathbf{Z} = \mathbf{A}\mathbf{X} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ . & . & \dots & . \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{bmatrix}.$$

The components of $\mathbf{Z}$ are:

$$Z_j = a_{j1}X_1 + a_{j2}X_2 + \dots + a_{jn}X_n.$$

Since $dz_j/dx_i = a_{ji}$, the Jacobian is then simply:

$$J(x_1, x_2, \dots, x_n) = \det \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ . & . & \dots & . \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \det \mathbf{A}.$$

Assuming that $\mathbf{A}$ is invertible,[1] we then have that:

$$f_{\mathbf{Z}}(\mathbf{z}) = \left. \frac{f_{\mathbf{X}}(\mathbf{x})}{|\det \mathbf{A}|} \right|_{\mathbf{x} = \mathbf{A}^{-1}\mathbf{z}} = \frac{f_{\mathbf{X}}(\mathbf{A}^{-1}\mathbf{z})}{|\det \mathbf{A}|}.$$

Example 6.15   Sum of Random Variables

Given a random vector $\mathbf{X} = (X_1, X_2, X_3)$, find the joint pdf of the sum:

$$Z = X_1 + X_2 + X_3.$$

We will use the transformation by introducing auxiliary variables as follows:

$$Z_1 = X_1, Z_2 = X_1 + X_2, Z_3 = X_1 + X_2 + X_3.$$

The inverse transformation is given by:

$$X_1 = Z_1, X_2 = Z_2 - Z_1, X_3 = Z_3 - Z_2.$$

The Jacobian matrix is:

$$J(x_1, x_2, x_3) = \det \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} = 1.$$

Therefore the joint pdf of $\mathbf{Z}$ is

$$f_{\mathbf{Z}}(z_1, z_2, z_3) = f_{\mathbf{X}}(z_1, z_2 - z_1, z_3 - z_2).$$

The pdf of Z_3 is obtained by integrating with respect to z_1 and z_2:

$$f_{Z_3}(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{X}}(z_1, z_2 - z_1, z - z_2) \, dz_1 dz_2.$$

This expression can be simplified further if X_1, X_2, and X_3 are independent random variables.

---

[1]Appendix C provides a summary of definitions and useful results from linear algebra.

6.3   EXPECTED VALUES OF VECTOR RANDOM VARIABLES

In this section we are interested in the characterization of a vector random variable through the expected values of its components and of functions of its components. We focus on the characterization of a vector random variable through its mean vector and its covariance matrix. We then introduce the joint characteristic function for a vector random variable.

The expected value of a function $g(\mathbf{X}) = g(X_1, \ldots, X_n)$ of a vector random variable $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is given by:

$$
E[Z] = \begin{cases} \displaystyle\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n) f_{\mathbf{X}}(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n & \mathbf{X} \text{ jointly} \\ & \quad\text{continuous} \\ \displaystyle\sum_{x_1} \cdots \sum_{x_n} g(x_1, x_2, \ldots, x_n) p_{\mathbf{X}}(x_1, x_2, \ldots, x_n) & \mathbf{X} \text{ discrete.} \end{cases}
$$

$$(6.25)$$

An important example is $g(\mathbf{X})$ equal to the sum of functions of $\mathbf{X}$. The procedure leading to Eq. (5.26) and a simple induction argument show that:

$$
E[g_1(\mathbf{X}) + g_2(\mathbf{X}) + \cdots + g_n(\mathbf{X})] = E[g_1(\mathbf{X})] + \cdots + E[g_n(\mathbf{X})]. \quad (6.26)
$$

Another important example is $g(\mathbf{X})$ equal to the product of n individual functions of the components. If $X_1, \ldots, X_n$ are *independent* random variables, then

$$
E[g_1(X_1)g_2(X_2)\ldots g_n(X_n)] = E[g_1(X_1)]E[g_2(X_2)]\ldots E[g_n(X_n)]. \quad (6.27)
$$

6.3.1   Mean Vector and Covariance Matrix

The mean, variance, and covariance provide useful information about the distribution of a random variable and are easy to estimate, so we are frequently interested in characterizing multiple random variables in terms of their first and second moments. We now introduce the mean vector and the covariance matrix. We then investigate the mean vector and the covariance matrix of a linear transformation of a random vector.

For $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, the **mean vector** is defined as the column vector of expected values of the components X_k:

$$
\mathbf{m_X} = E[\mathbf{X}] = E\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \triangleq \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}. \quad (6.28a)
$$

Note that we define the vector of expected values as a column vector. In previous sections we have sometimes written $\mathbf{X}$ as a row vector, but in this section and wherever we deal with matrix transformations, we will represent $\mathbf{X}$ and its expected value as a column vector.

The **correlation matrix** has the second moments of **X** as its entries:

$$\mathbf{R_X} = \begin{bmatrix} E[X_1^2] & E[X_1X_2] & \ldots & E[X_1X_n] \\ E[X_2X_1] & E[X_2^2] & \ldots & E[X_2X_n] \\ . & . & \ldots & . \\ E[X_nX_1] & E[X_nX_2] & \ldots & E[X_n^2] \end{bmatrix}. \tag{6.28b}$$

The **covariance matrix** has the second-order central moments as its entries:

$$\mathbf{K_X} = \begin{bmatrix} E[(X_1-m_1)^2] & E[(X_1-m_1)(X_2-m_2)] & \ldots & E[(X_1-m_1)(X_n-m_n)] \\ E[(X_2-m_2)(X_1-m_1)] & E[(X_2-m_2)^2] & \ldots & E[(X_2-m_2)(X_n-m_n)] \\ . & . & \ldots & . \\ E[(X_n-m_n)(X_1-m_1)] & E[(X_n-m_n)(X_2-m_2)] & \ldots & E[(X_n-m_n)^2] \end{bmatrix}. \tag{6.28c}$$

Both $\mathbf{R_X}$ and $\mathbf{K_X}$ are $n \times n$ symmetric matrices. The diagonal elements of $\mathbf{K_X}$ are given by the variances $\text{VAR}[X_k] = E[(X_k - m_k)^2]$ of the elements of **X**. If these elements are uncorrelated, then $\text{COV}(X_j, X_k) = 0$ for $j \neq k$, and $\mathbf{K_X}$ is a diagonal matrix. If the random variables $X_1, \ldots, X_n$ are independent, then they are uncorrelated and $\mathbf{K_X}$ is diagonal. Finally, if the vector of expected values is **0**, that is, $m_k = E[X_k] = 0$ for all k, then $\mathbf{R_X} = \mathbf{K_X}$.

Example 6.16

Let $\mathbf{X} = (X_1, X_2, X_3)$ be the jointly Gaussian random vector from Example 6.6. Find $E[\mathbf{X}]$ and $\mathbf{K_X}$.
We rewrite the joint pdf as follows:

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = \frac{e^{-(x_1^2 + x_2^2 - 2\frac{1}{\sqrt{2}}x_1x_2)}}{2\pi\sqrt{1 - \left(-\dfrac{1}{\sqrt{2}}\right)^2}} \frac{e^{-x_3^2/2}}{\sqrt{2\pi}}.$$

We see that X_3 is a Gaussian random variable with zero mean and unit variance, and that it is independent of X_1 and X_2. We also see that X_1 and X_2 are jointly Gaussian with zero mean and unit variance, and with correlation coefficient

$$\rho_{X_1X_2} = -\frac{1}{\sqrt{2}} = \frac{\text{COV}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \text{COV}(X_1, X_2).$$

Therefore the vector of expected values is: $\mathbf{m_X} = \mathbf{0}$, and

$$\mathbf{K_X} = \begin{bmatrix} 1 & -\dfrac{1}{\sqrt{2}} & 0 \\ -\dfrac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We now develop compact expressions for $\mathbf{R_X}$ and $\mathbf{K_X}$. If we multiply $\mathbf{X}$, an $n \times 1$ matrix, and $\mathbf{X}^{\mathrm{T}}$, a $1 \times n$ matrix, we obtain the following $n \times n$ matrix:

$$\mathbf{XX}^{\mathrm{T}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} [X_1, X_2, \ldots, X_n] = \begin{bmatrix} X_1^2 & X_1X_2 & \ldots & X_1X_n \\ X_2X_1 & X_2^2 & \ldots & X_2X_n \\ . & . & \ldots & . \\ X_nX_1 & X_nX_2 & \ldots & X_n^2 \end{bmatrix}.$$

If we define the expected value of a matrix to be the matrix of expected values of the matrix elements, then we can write the correlation matrix as:

$$\mathbf{R_X} = E[\mathbf{XX}^{\mathrm{T}}]. \tag{6.29a}$$

The covariance matrix is then:

$$\begin{aligned} \mathbf{K_X} &= E[(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^{\mathrm{T}}] \\ &= E[\mathbf{XX}^{\mathrm{T}}] - \mathbf{m_X}\, E[\mathbf{X}^{\mathrm{T}}] - E[\mathbf{X}]\mathbf{m_X}^{\mathrm{T}} + \mathbf{m_X}\mathbf{m_X}^{\mathrm{T}} \\ &= \mathbf{R_X} - \mathbf{m_X}\mathbf{m_X}^{\mathrm{T}}. \end{aligned} \tag{6.29b}$$

6.3.2   Linear Transformations of Random Vectors

Many engineering systems are linear in the sense that will be elaborated on in Chapter 10. Frequently these systems can be reduced to a linear transformation of a vector of random variables where the "input" is $\mathbf{X}$ and the "output" is $\mathbf{Y}$:

$$\mathbf{Y} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_n \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & \ldots & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \mathbf{AX}.$$

The expected value of the kth component of $\mathbf{Y}$ is the inner product (dot product) of the kth row of $\mathbf{A}$ and $\mathbf{X}$:

$$E[Y_k] = E\left[\sum_{j=1}^{n} a_{kj}X_j\right] = \sum_{j=1}^{n} a_{kj}E[X_j].$$

Each component of $E[\mathbf{Y}]$ is obtained in this manner, so:

$$\mathbf{m_Y} = E[\mathbf{Y}] = \begin{bmatrix} \sum_{j=1}^{n} a_{1j}E[X_j] \\ \sum_{j=1}^{n} a_{2j}E[X_j] \\ \vdots \\ \sum_{j=1}^{n} a_{nj}E[X_j] \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_n \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & \ldots & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}$$

$$= \mathbf{A}E[\mathbf{X}] = \mathbf{Am_X}. \tag{6.30a}$$

The covariance matrix of **Y** is then:

$$\mathbf{K_Y} = E[(\mathbf{Y} - \mathbf{m_Y})(\mathbf{Y} - \mathbf{m_Y})^T] = E[(\mathbf{AX} - \mathbf{Am_X})(\mathbf{AX} - \mathbf{Am_X})^T]$$
$$= E[\mathbf{A}(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^T\mathbf{A}^T] = \mathbf{A}E[(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^T]\mathbf{A}^T$$
$$= \mathbf{AK_XA}^T, \tag{6.30b}$$

where we used the fact that the transpose of a matrix multiplication is the product of the transposed matrices in reverse order: $\{\mathbf{A}(\mathbf{X} - \mathbf{m_X})\}^T = (\mathbf{X} - \mathbf{m_X})^T\mathbf{A}^T$.

The **cross-covariance** matrix of two random vectors **X** and **Y** is defined as:

$$\mathbf{K_{XY}} = E[(\mathbf{X} - \mathbf{m_X})(\mathbf{Y} - \mathbf{m_Y})^T] = E[\mathbf{XY}^T] - \mathbf{m_X}\mathbf{m_Y}^T = \mathbf{R_{XY}} - \mathbf{m_X}\mathbf{m_Y}^T.$$

We are interested in the cross-covariance between **X** and $\mathbf{Y} = \mathbf{AX}$:

$$\mathbf{K_{XY}} = E[\mathbf{X} - \mathbf{m_X})(\mathbf{Y} - \mathbf{m_Y})^T] = E[(\mathbf{X} - \mathbf{m_X})(\mathbf{X} - \mathbf{m_X})^T\mathbf{A}^T]$$
$$= \mathbf{K_XA}^T. \tag{6.30c}$$

Example 6.17   Transformation of Uncorrelated Random Vector

Suppose that the components of **X** are uncorrelated and have unit variance, then $\mathbf{K_X} = \mathbf{I}$, the identity matrix. The covariance matrix for $\mathbf{Y} = \mathbf{AX}$ is

$$\mathbf{K_Y} = \mathbf{AK_XA}^T = \mathbf{AIA}^T = \mathbf{AA}^T. \tag{6.31}$$

In general $\mathbf{K_Y} = \mathbf{AA}^T$ is not a diagonal matrix and so the components of **Y** are correlated. In Section 6.6 we discuss how to find a matrix **A** so that Eq. (6.31) holds for a given $\mathbf{K_Y}$. We can then generate a random vector **Y** with any desired covariance matrix $\mathbf{K_Y}$.

Suppose that the components of **X** are correlated so $\mathbf{K_X}$ is not a diagonal matrix. In many situations we are interested in finding a transformation matrix **A** so that $\mathbf{Y} = \mathbf{AX}$ has uncorrelated components. This requires finding **A** so that $\mathbf{K_Y} = \mathbf{AK_XA}^T$ is a diagonal matrix. In the last part of this section we show how to find such a matrix **A**.

Example 6.18   Transformation to Uncorrelated Random Vector

Suppose the random vector X_1, X_2, and X_3 in Example 6.16 is transformed using the matrix:

$$\mathbf{A} = \begin{bmatrix} \dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} & 0 \\ \dfrac{1}{\sqrt{2}} & -\dfrac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Find the $E[\mathbf{Y}]$ and $\mathbf{K_Y}$.

Since $\mathbf{m_X} = \mathbf{0}$, then $E[\mathbf{Y}] = \mathbf{Am_X} = \mathbf{0}$. The covariance matrix of $\mathbf{Y}$ is:

$$\mathbf{K_Y} = \mathbf{AK_XA^T} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & -\dfrac{1}{\sqrt{2}} & 0 \\ -\dfrac{1}{\sqrt{2}} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 - \dfrac{1}{\sqrt{2}} & 1 + \dfrac{1}{\sqrt{2}} & 0 \\ 1 - \dfrac{1}{\sqrt{2}} & -\left(1 + \dfrac{1}{\sqrt{2}}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 - \dfrac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 1 + \dfrac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The linear transformation has produced a vector of random variables $\mathbf{Y} = (Y_1, Y_2, Y_3)$ with components that are uncorrelated.

---

*6.3.3 Joint Characteristic Function

The **joint characteristic function** of n random variables is defined as

$$\Phi_{X_1, X_2, \ldots, X_n}(\omega_1, \omega_2, \ldots, \omega_n) = E[e^{j(\omega_1 X_1 + \omega_2 X_2 + \cdots + \omega_n X_n)}]. \tag{6.32a}$$

In this section we develop the properties of the joint characteristic function of two random variables. These properties generalize in straightforward fashion to the case of n random variables. Therefore consider

$$\Phi_{X,Y}(\omega_1, \omega_2) = E[e^{j(\omega_1 X + \omega_2 Y)}]. \tag{6.32b}$$

If X and Y are jointly continuous random variables, then

$$\Phi_{X,Y}(\omega_1, \omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)e^{j(\omega_1 x + \omega_2 y)}\, dx\, dy. \tag{6.32c}$$

Equation (6.32c) shows that the joint characteristic function is the two-dimensional Fourier transform of the joint pdf of X and Y. The inversion formula for the Fourier transform implies that the joint pdf is given by

$$f_{X,Y}(x, y) = \frac{1}{4\pi^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi_{X,Y}(\omega_1, \omega_2)e^{-j(\omega_1 x + \omega_2 y)}\, d\omega_1\, d\omega_2. \tag{6.33}$$

Note in Eq. (6.32b) that the marginal characteristic functions can be obtained from joint characteristic function:

$$\Phi_X(\omega) = \Phi_{X,Y}(\omega, 0) \quad \Phi_Y(\omega) = \Phi_{X,Y}(0, \omega). \tag{6.34}$$

If X and Y are independent random variables, then the joint characteristic function is the product of the marginal characteristic functions since

$$\Phi_{X,Y}(\omega_1, \omega_2) = E[e^{j(\omega_1 X + \omega_2 Y)}] = E[e^{j\omega_1 X}e^{j\omega_2 Y}]$$

$$= E[e^{j\omega_1 X}]E[e^{j\omega_2 Y}] = \Phi_X(\omega_1)\Phi_Y(\omega_2), \tag{6.35}$$

where the third equality follows from Eq. (6.27).

The characteristic function of the sum $Z = aX + bY$ can be obtained from the joint characteristic function of X and Y as follows:

$$\Phi_Z(\omega) = E[e^{j\omega(aX+bY)}] = E[e^{j(\omega aX+\omega bY)}] = \Phi_{X,Y}(a\omega, b\omega). \qquad (6.36a)$$

If X and Y are independent random variables, the characteristic function of $Z = aX + bY$ is then

$$\Phi_Z(\omega) = \Phi_{X,Y}(a\omega, b\omega) = \Phi_X(a\omega)\Phi_Y(b\omega). \qquad (6.36b)$$

In Section 8.1 we will use the above result in dealing with sums of random variables.

The joint moments of X and Y (if they exist) can be obtained by taking the derivatives of the joint characteristic function. To show this we rewrite Eq. (6.32b) as the expected value of a product of exponentials and we expand the exponentials in a power series:

$$\Phi_{X,Y}(\omega_1, \omega_2) = E[e^{j\omega_1 X}e^{j\omega_2 Y}]$$

$$= E\left[\sum_{i=0}^{\infty} \frac{(j\omega_1 X)^i}{i!} \sum_{k=0}^{\infty} \frac{(j\omega_2 Y)^k}{k!} \right]$$

$$= \sum_{i=0}^{\infty} \sum_{k=0}^{\infty} E[X^i Y^k] \frac{(j\omega_1)^i}{i!} \frac{(j\omega_2)^k}{k!}.$$

It then follows that the moments can be obtained by taking an appropriate set of derivatives:

$$E[X^i Y^k] = \frac{1}{j^{i+k}} \frac{\partial^i \partial^k}{\partial \omega_1^i \partial \omega_2^k} \Phi_{X,Y}(\omega_1, \omega_2) \big|_{\omega_1=0,\omega_2=0}. \qquad (6.37)$$

Example 6.19

Suppose U and V are independent zero-mean, unit-variance Gaussian random variables, and let

$$X = U + V \qquad Y = 2U + V.$$

Find the joint characteristic function of X and Y, and find $E[XY]$.

The joint characteristic function of X and Y is

$$\Phi_{X,Y}(\omega_1, \omega_2) = E[e^{j(\omega_1 X + \omega_2 Y)}] = E[e^{j\omega_1(U+V)}e^{j\omega_2(2U+V)}]$$

$$= E[e^{j((\omega_1+2\omega_2)U + (\omega_1+\omega_2)V)}].$$

Since U and V are independent random variables, the joint characteristic function of U and V is equal to the product of the marginal characteristic functions:

$$\Phi_{X,Y}(\omega_1, \omega_2) = E[e^{j((\omega_1+2\omega_2)U)}]E[e^{j((\omega_1+\omega_2)V)}]$$

$$= \Phi_U(\omega_1 + 2\omega_2)\Phi_V(\omega_1 + \omega_2)$$

$$= e^{-\frac{1}{2}(\omega_1+2\omega_2)^2}e^{-\frac{1}{2}(\omega_1+\omega_2)^2}$$

$$= e^{\{-\ \frac{1}{2}(2\omega_1^2 + 6\omega_1\omega_2 + 5\omega_2^2)\}}.$$

where marginal characteristic functions were obtained from Table 4.1.

The correlation $E[XY]$ is found from Eq. (6.37) with $i = 1$ and $k = 1$:

$$E[XY] = \frac{1}{j^2} \frac{\partial^2}{\partial \omega_1 \partial \omega_2} \Phi_{X,Y}(\omega_1, \omega_2)\big|_{\omega_1 = 0, \omega_2 = 0}$$

$$= -\exp\{-\tfrac{1}{2}(2\omega_1^2 + 6\omega_1\omega_2 + 5\omega_2^2)\}[6\omega_1 + 10\omega_2]\left(\frac{1}{4}\right)[4\omega_1 + 6\omega_2]$$

$$+ \frac{1}{2}\exp\{-\tfrac{1}{2}(2\omega_1^2 + 6\omega_1\omega_2 + 5\omega_2^2)\}[6]\big|_{\omega_1 = 0, \omega_2 = 0}$$

$$= 3.$$

You should verify this answer by evaluating $E[XY] = E[(U + V)(2U + V)]$ directly.

---

*6.3.4   Diagonalization of Covariance Matrix

Let $\mathbf{X}$ be a random vector with covariance $\mathbf{K_X}$. We are interested in finding an $n \times n$ matrix $\mathbf{A}$ such that $\mathbf{Y} = \mathbf{AX}$ has a covariance matrix that is diagonal. The components of $\mathbf{Y}$ are then uncorrelated.

We saw that $\mathbf{K_X}$ is a real-valued symmetric matrix. In Appendix C we state results from linear algebra that $\mathbf{K_X}$ is then a diagonalizable matrix, that is, there is a matrix $\mathbf{P}$ such that:

$$\mathbf{P}^T\mathbf{K_X}\mathbf{P} = \Lambda \quad \text{and} \quad \mathbf{P}^T\mathbf{P} = \mathbf{I} \tag{6.38a}$$

where Λ is a diagonal matrix and $\mathbf{I}$ is the identity matrix. Therefore if we let $\mathbf{A} = \mathbf{P}^T$, then from Eq. (6.30b) we obtain a diagonal $\mathbf{K_Y}$.

We now show how $\mathbf{P}$ is obtained. First, we find the eigenvalues and eigenvectors of $\mathbf{K_X}$ from:

$$\mathbf{K_X}\mathbf{e}_i = \lambda_i \mathbf{e}_i \tag{6.38b}$$

where $\mathbf{e}_i$ are $n \times 1$ column vectors.[2] We can normalize each eigenvector $\mathbf{e}_i$ so that $\mathbf{e}_i{}^T\mathbf{e}_i$, the sum of the square of its components, is 1. The normalized eigenvectors are then orthonormal, that is,

$$\mathbf{e}_i{}^T\mathbf{e}_j = \delta_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \tag{6.38c}$$

Let $\mathbf{P}$ be the matrix whose columns are the eigenvectors of $\mathbf{K_X}$ and let Λ be the diagonal matrix of eigenvalues:

$$\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] \qquad \Lambda = \text{diag}[\lambda_1].$$

From Eq. (6.38b) we have:

$$\mathbf{K_X}\mathbf{P} = \mathbf{K_X}[\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] = [\mathbf{K_X}\mathbf{e}_1, \mathbf{K_X}\mathbf{e}_2, \ldots, \mathbf{K_X}\mathbf{e}_n]$$

$$= [\lambda_1\mathbf{e}_1, \lambda_2\mathbf{e}_2, \ldots, \lambda_n\mathbf{e}_n] = \mathbf{P}\Lambda \tag{6.39a}$$

where the second equality follows from the fact that each column of $\mathbf{K_X}\mathbf{P}$ is obtained by multiplying a column of $\mathbf{P}$ by $\mathbf{K_X}$. By premultiplying both sides of the above equations by $\mathbf{P}^T$, we obtain:

$$\mathbf{P}^T\mathbf{K_X}\mathbf{P} = \mathbf{P}^T\mathbf{P}\Lambda = \Lambda. \tag{6.39b}$$

[2]See Appendix C.

We conclude that if we let $\mathbf{A} = \mathbf{P}^T$, and

$$\mathbf{Y} = \mathbf{A}\mathbf{X} = \mathbf{P}^T\mathbf{X}, \tag{6.40a}$$

then the random variables in $\mathbf{Y}$ are uncorrelated since

$$\mathbf{K_Y} = \mathbf{P}^T\mathbf{K_X}\mathbf{P} = \Lambda. \tag{6.40b}$$

In summary, any covariance matrix $\mathbf{K}_X$. can be diagonalized by a linear transformation. The matrix $\mathbf{A}$ in the transformation is obtained from the eigenvectors of $\mathbf{K_X}$.

Equation (6.40b) provides insight into the invertibility of $\mathbf{K_X}$ and $\mathbf{K_Y}$. From linear algebra we know that the determinant of a product of $n \times n$ matrices is the product of the determinants, so:

$$\det \mathbf{K_Y} = \det \mathbf{P}^T \det \mathbf{K_X} \det \mathbf{P} = \det \Lambda = \lambda_1\lambda_2\ldots\lambda_n,$$

where we used the fact that $\det \mathbf{P}^T \det \mathbf{P} = \det \mathbf{I} = 1$. Recall that a matrix is invertible if and only if its determinant is nonzero. Therefore $\mathbf{K_Y}$ is not invertible if and only if one or more of the eigenvalues of $\mathbf{K_X}$ is zero.

Now suppose that one of the eigenvalues is zero, say $\lambda_k = 0$. Since $\text{VAR}[Y_k] = \lambda_k = 0$, then $Y_k = 0$. But Y_k is defined as a linear combination, so

$$0 = Y_k = a_{k1}X_1 + a_{k2}X_2 + \cdots + a_{kn}X_n.$$

We conclude that the components of $\mathbf{X}$ are linearly dependent. Therefore, one or more of the components in $\mathbf{X}$ are redundant and can be expressed as a linear combination of the other components.

It is interesting to look at the vector $\mathbf{X}$ expressed in terms of $\mathbf{Y}$. Multiply both sides of Eq. (6.40a) by $\mathbf{P}$ and use the fact that $\mathbf{PP}^T = \mathbf{I}$:

$$\mathbf{X} = \mathbf{PP}^T\mathbf{X} = \mathbf{PY} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n]\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{k=1}^{n} Y_k\mathbf{e}_k. \tag{6.41}$$

This equation is called the **Karhunen-Loeve expansion**. The equation shows that a random vector $\mathbf{X}$ can be expressed as a weighted sum of the eigenvectors of $\mathbf{K_X}$, where the coefficients are *uncorrelated* random variables Y_k. Furthermore, the eigenvectors form an orthonormal set. Note that if any of the eigenvalues are zero, $\text{VAR}[Y_k] = \lambda_k = 0$, then $Y_k = 0$, and the corresponding term can be dropped from the expansion in Eq. (6.41). In Chapter 10, we will see that this expansion is very useful in the processing of random signals.

6.4   JOINTLY GAUSSIAN RANDOM VECTORS

The random variables $X_1, X_2, \ldots, X_n$ are said to be jointly Gaussian if their joint pdf is given by

$$f_{\mathbf{X}}(\mathbf{x}) \triangleq f_{X_1, X_2, \ldots, X_n}(x_1, \ldots, x_n) = \frac{\exp\{-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T K^{-1}(\mathbf{x} - \mathbf{m})\}}{(2\pi)^{n/2}|K|^{1/2}}, \tag{6.42a}$$

where **x** and **m** are column vectors defined by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad \mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}$$

and K is the covariance matrix that is defined by

$$K = \begin{bmatrix} \text{VAR}(X_1) & \text{COV}(X_1, X_2) & \cdots & \text{COV}(X_1, X_n) \\ \text{COV}(X_2, X_1) & \text{VAR}(X_2) & \cdots & \text{COV}(X_2, X_n) \\ \vdots & \vdots & & \vdots \\ \text{COV}(X_n, X_1) & \cdots & & \text{VAR}(X_n) \end{bmatrix}. \qquad (6.42b)$$

The $(.)^{\mathrm{T}}$ in Eq. (6.42a) denotes the transpose of a matrix or vector. Note that the co-variance matrix is a symmetric matrix since $\text{COV}(X_i, X_j) = \text{COV}(X_j, X_i)$.

Equation (6.42a) shows that *the pdf of jointly Gaussian random variables is completely specified by the individual means and variances and the pairwise covariances*. It can be shown using the joint characteristic function that all the marginal pdf's associated with Eq. (6.42a) are also Gaussian and that these too are completely specified by the same set of means, variances, and covariances.

Example 6.20

Verify that the two-dimensional Gaussian pdf given in Eq. (5.61a) has the form of Eq. (6.42a).

The covariance matrix for the two-dimensional case is given by

$$K = \begin{bmatrix} \sigma_1^2 & \rho_{X,Y}\sigma_1\sigma_2 \\ \rho_{X,Y}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

where we have used the fact the $\text{COV}(X_1, X_2) = \rho_{X,Y}\sigma_1\sigma_2$. The determinant of K is $\sigma_1^2 \sigma_2^2(1 - \rho_{X,Y}^2)$ so the denominator of the pdf has the correct form. The inverse of the covariance matrix is also a real symmetric matrix:

$$K^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho_{X,Y}^2)} \begin{bmatrix} \sigma_2^2 & -\rho_{X,Y}\sigma_1\sigma_2 \\ -\rho_{X,Y}\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}.$$

The term in the exponent is therefore

$$\frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho_{X,Y}^2)} (x - m_1, y - m_2) \begin{bmatrix} \sigma_2^2 & -\rho_{X,Y}\sigma_1\sigma_2 \\ -\rho_{X,Y}\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} \begin{bmatrix} x - m_1 \\ y - m_2 \end{bmatrix}$$

$$= \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho_{X,Y}^2)} (x - m_1, y - m_2) \begin{bmatrix} \sigma_2^2(x - m_1) - \rho_{X,Y}\sigma_1\sigma_2(y - m_2) \\ -\rho_{X,Y}\sigma_1\sigma_2(x - m_1) + \sigma_1^2(y - m_2) \end{bmatrix}$$

$$= \frac{((x - m_1)/\sigma_1)^2 - 2\rho_{X,Y}((x - m_1)/\sigma_1)((y - m_2)/\sigma_2) + ((y - m_2)/\sigma_2)^2}{(1 - \rho_{X,Y}^2)}.$$

Thus the two-dimensional pdf has the form of Eq. (6.42a).

Example 6.21

The vector of random variables (X, Y, Z) is jointly Gaussian with zero means and covariance matrix:

$$K = \begin{bmatrix} \text{VAR}(X) & \text{COV}(X, Y) & \text{COV}(X, Z) \\ \text{COV}(Y, X) & \text{VAR}(Y) & \text{COV}(Y, Z) \\ \text{COV}(Z, X) & \text{COV}(Z, Y) & \text{VAR}(Z) \end{bmatrix} = \begin{bmatrix} 1.0 & 0.2 & 0.3 \\ 0.2 & 1.0 & 0.4 \\ 0.3 & 0.4 & 1.0 \end{bmatrix}.$$

Find the marginal pdf of X and Z.

We can solve this problem two ways. The first involves integrating the pdf directly to obtain the marginal pdf. The second involves using the fact that the marginal pdf for X and Z is also Gaussian and has the same set of means, variances, and covariances. We will use the second approach.

The pair (X, Z) has zero-mean vector and covariance matrix:

$$K' = \begin{bmatrix} \text{VAR}(X) & \text{COV}(X, Z) \\ \text{COV}(Z, X) & \text{VAR}(Z) \end{bmatrix} = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}.$$

The joint pdf of X and Z is found by substituting a zero-mean vector and this covariance matrix into Eq. (6.42a).

Example 6.22   Independence of Uncorrelated Jointly Gaussian Random Variables

Suppose $X_1, X_2, \ldots, X_n$ are jointly Gaussian random variables with $\text{COV}(X_i, X_j) = 0$ for $i \neq j$. Show that $X_1, X_2, \ldots, X_n$ are independent random variables.

From Eq. (6.42b) we see that the covariance matrix is a diagonal matrix:

$$K = \text{diag}[\text{VAR}(X_i)] = \text{diag}[\sigma_i^2]$$

Therefore

$$K^{-1} = \text{diag}\left[\frac{1}{\sigma_i^2}\right]$$

and

$$(\mathbf{x} - \mathbf{m})^{\text{T}} K^{-1} (\mathbf{x} - \mathbf{m}) = \sum_{i=1}^{n} \left(\frac{x_i - m_i}{\sigma_i}\right)^2.$$

Thus from Eq. (6.42a)

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}[(x_i - m_i)/\sigma_i]^2\right\}}{(2\pi)^{n/2}}|K|^{1/2} = \prod_{i=1}^{n}\frac{\exp\left\{-\frac{1}{2}[(x_i - m_i)/\sigma_i]^2\right\}}{\sqrt{2\pi\sigma_i^2}} = \prod_{i=1}^{n}f_{X_i}(x_i).$$

Thus $X_1, X_2, \ldots, X_n$ are independent Gaussian random variables.

Example 6.23   Conditional pdf of Gaussian Random Variable

Find the conditional pdf of X_n given $X_1, X_2, \ldots, X_{n-1}$.

Let $\mathbf{K}_n$ be the covariance matrix for $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ and $\mathbf{K}_{n-1}$ be the covariance matrix for $\mathbf{X}_{n-1} = (X_1, X_2, \ldots, X_{n-1})$. Let $\mathbf{Q}_n = \mathbf{K}_n^{-1}$ and $\mathbf{Q}_{n-1} = \mathbf{K}_{n-1}^{-1}$, then the latter matrices are

submatrices of the former matrices as shown below:

$$
\mathbf{K}_n = \begin{bmatrix} & & & K_{1n} \\ & \mathbf{K}_{n-1} & & K_{2n} \\ & & & \vdots \\ K_{1n} & K_{2n} & \dots & K_{nn} \end{bmatrix}
\qquad
\mathbf{Q}_n = \begin{bmatrix} & & & Q_{1n} \\ & \mathbf{Q}_{n-1} & & Q_{2n} \\ & & & \vdots \\ Q_{1n} & Q_{2n} & \dots & Q_{nn} \end{bmatrix}
$$

Below we will use the subscript n or $n-1$ to distinguish between the two random vectors and their parameters. The marginal pdf of X_n given $X_1, X_2, \dots, X_{n-1}$ is given by:

$$
f_{X_n}(x_n \,|\, x_1, \dots, x_{n-1}) = \frac{f_{\mathbf{X}_n}(\mathbf{x}_n)}{f_{\mathbf{X}_{n-1}}(\mathbf{x}_{n-1})}
$$

$$
= \frac{\exp\{-\tfrac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^{\mathrm{T}} \mathbf{Q}_n (\mathbf{x}_n - \mathbf{m}_n)\}}{(2\pi)^{n/2} |\mathbf{K}_n|^{1/2}} \frac{(2\pi)^{(n-1)1/2} |\mathbf{K}_{n-1}|^{1/2}}{\exp\{-\tfrac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^{\mathrm{T}} \mathbf{Q}_{n-1} (\mathbf{x}_{n-1} - \mathbf{m}_{n-1})\}}
$$

$$
= \frac{\exp\{-\tfrac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^{\mathrm{T}} \mathbf{Q}_n (\mathbf{x}_n - \mathbf{m}_n) + \tfrac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^{\mathrm{T}} \mathbf{Q}_{n-1} (\mathbf{x}_{n-1} - \mathbf{m}_{n-1})\}}{\sqrt{2\pi} |\mathbf{K}_n|^{1/2} / |\mathbf{K}_{n-1}|^{1/2}}.
$$

In Problem 6.60 we show that the terms in the above expression are given by:

$$
\tfrac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^{\mathrm{T}} \mathbf{Q}_n (\mathbf{x}_n - \mathbf{m}_n) - \tfrac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^{\mathrm{T}} \mathbf{Q}_{n-1} (\mathbf{x}_{n-1} - \mathbf{m}_{n-1})
$$

$$
= Q_{nn}\{(x_n - m_n) + B\}^2 - Q_{nn}B^2 \tag{6.43}
$$

where $B = \dfrac{1}{Q_{nn}} \displaystyle\sum_{j=1}^{n-1} Q_{jn}(x_j - m_j)$ and $|\mathbf{K}_n|/|\mathbf{K}_{n-1}| = 1/Q_{nn}$.

This implies that X_n has mean $m_n - B$, and variance $1/Q_{nn}$. The term $Q_{nn}B^2$ is part of the normalization constant. We therefore conclude that:

$$
f_{X_n}(x_n \,|\, x_1, \dots, x_{n-1}) = \frac{\exp\left\{-\dfrac{Q_{nn}}{2}\left(x - m_n + \dfrac{1}{Q_{nn}}\displaystyle\sum_{j=1}^{n-1} Q_{jn}(x_j - m_j)\right)^2\right\}}{\sqrt{2\pi / Q_{nn}}}
$$

We see that the conditional mean of X_n is a linear function of the "observations" $x_1, x_2, \dots, x_{n-1}$.

*6.4.1   Linear Transformation of Gaussian Random Variables

A very important property of jointly Gaussian random variables is that *the linear transformation of any n jointly Gaussian random variables results in n random variables that are also jointly Gaussian.* This is easy to show using the matrix notation in Eq. (6.42a). Let $\mathbf{X} = (X_1, \dots, X_n)$ be jointly Gaussian with covariance matrix K_X and mean vector $\mathbf{m_X}$ and define $\mathbf{Y} = (Y_1, \dots, Y_n)$ by

$$
\mathbf{Y} = A\mathbf{X},
$$

where A is an invertible $n \times n$ matrix. From Eq. (5.60) we know that the pdf of $\mathbf{Y}$ is given by

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

$$= \frac{\exp\{-\frac{1}{2}(A^{-1}\mathbf{y} - \mathbf{mX})^{\mathrm{T}} K_X^{-1}(A^{-1}\mathbf{y} - \mathbf{mX})\}}{(2\pi)^{n/2}|A||K_X|^{1/2}}. \tag{6.44}$$

From elementary properties of matrices we have that

$$(A^{-1}\mathbf{y} - \mathbf{m_X}) = A^{-1}(\mathbf{y} - A\mathbf{m_X})$$

and

$$(A^{-1}\mathbf{y} - \mathbf{m_X})^{\mathrm{T}} = (\mathbf{y} - A\mathbf{m_X})^{\mathrm{T}} A^{-1\mathrm{T}}.$$

The argument in the exponential is therefore equal to

$$(\mathbf{y} - A\mathbf{m_X})^{\mathrm{T}} A^{-1\mathrm{T}} K_X^{-1} A^{-1}(\mathbf{y} - A\mathbf{m_X}) = (\mathbf{y} - A\mathbf{m_X})^{\mathrm{T}}(AK_X A^{\mathrm{T}})^{-1}(\mathbf{y} - A\mathbf{m_X})$$

since $A^{-1\mathrm{T}} K_X^{-1} = (AK_X A^{\mathrm{T}})^{-1}$. Letting $K_Y = AK_X A^{\mathrm{T}}$ and $\mathbf{m_Y} = A\mathbf{m_X}$ and noting that $\det(K_Y) = \det(AK_X A^{\mathrm{T}}) = \det(A)\det(K_X)\det(A^{\mathrm{T}}) = \det(A)^2 \det(K_X)$, we finally have that the pdf of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{e^{-(1/2)(\mathbf{y}-\mathbf{m_Y})^{\mathrm{T}} K_Y^{-1}(\mathbf{y}-\mathbf{m_Y})}}{(2\pi)^{n/2}|K_Y|^{1/2}}. \tag{6.45}$$

Thus the pdf of $\mathbf{Y}$ has the form of Eq. (6.42a) and therefore $Y_1, \ldots, Y_n$ are jointly Gaussian random variables with mean vector and covariance matrix:

$$\mathbf{m_Y} = A\mathbf{m_X} \quad \text{and} \quad K_Y = AK_X A^{\mathrm{T}}.$$

This result is consistent with the mean vector and covariance matrix we obtained before in Eqs. (6.30a) and (6.30b).

In many problems we wish to transform $\mathbf{X}$ to a vector $\mathbf{Y}$ of independent Gaussian random variables. Since K_X is a symmetric matrix, it is always possible to find a matrix A such that $AK_X A^{\mathrm{T}} = \Lambda$ is a diagonal matrix. (See Section 6.6.) For such a matrix A, the pdf of $\mathbf{Y}$ will be

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{e^{-(1/2)(\mathbf{y}-\mathbf{n})^{\mathrm{T}}\Lambda^{-1}(\mathbf{y}-\mathbf{n})}}{(2\pi)^{n/2}|\Lambda|^{1/2}}$$

$$= \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - n_i)^2/\lambda_i\right\}}{[(2\pi\lambda_1)(2\pi\lambda_2)\ldots(2\pi\lambda_n)]^{1/2}}, \tag{6.46}$$

where $\lambda_1, \ldots, \lambda_n$ are the diagonal components of Λ. We assume that these values are all nonzero. The above pdf implies that $Y_1, \ldots, Y_n$ are *independent* random variables

with means n_i and variance λ_i. In conclusion, *it is possible to linearly transform a vector of jointly Gaussian random variables into a vector of independent Gaussian random variables*.

It is always possible to select the matrix A that diagonalizes K so that $\det(A) = 1$. The transformation $A\mathbf{X}$ then corresponds to a rotation of the coordinate system so that the principal axes of the ellipsoid corresponding to the pdf are aligned to the axes of the system. Example 5.48 provides an $n = 2$ example of rotation.

In computer simulation models we frequently need to generate jointly Gaussian random vectors with specified covariance matrix and mean vector. Suppose that $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ has components that are zero-mean, unit-variance Gaussian random variables, so its mean vector is $\mathbf{0}$ and its covariance matrix is the identity matrix $\mathbf{I}$. Let $\mathbf{K}$ denote the desired covariance matrix. Using the methods discussed in Section 6.3, it is possible to find a matrix $\mathbf{A}$ so that $\mathbf{A}^T\mathbf{A} = \mathbf{K}$. Therefore $\mathbf{Y} = \mathbf{A}^T\mathbf{U}$ has zero mean vector and covariance $\mathbf{K}$. From Eq. (6.46) we have that $\mathbf{Y}$ is also a jointly Gaussian random vector with zero mean vector and covariance $\mathbf{K}$. If we require a nonzero mean vector $\mathbf{m}$, we use $\mathbf{Y} + \mathbf{m}$.

Example 6.24   Sum of Jointly Gaussian Random Variables

Let $X_1, X_2, \ldots, X_n$ be jointly Gaussian random variables with joint pdf given by Eq. (6.42a). Let

$$Z = a_1X_1 + a_2X_2 + \cdots + a_nX_n.$$

We will show that Z is always a Gaussian random variable.

We find the pdf of Z by introducing auxiliary random variables. Let

$$Z_2 = X_2, \quad Z_3 = X_3, \ldots, \quad Z_n = X_n.$$

If we define $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$, then

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

where

$$A = \begin{bmatrix} a_1 & a_2 & \ldots & \cdot & a_n \\ 0 & 1 & \ldots & \cdot & 0 \\ \cdot & \cdot & \ldots & \cdot & \cdot \\ 0 & \cdot & \ldots & 0 & 1 \end{bmatrix}.$$

From Eq. (6.45) we have that $\mathbf{Z}$ is jointly Gaussian with mean $\mathbf{n} = A\mathbf{m}$, and covariance matrix $C = AKA^T$. Furthermore, it then follows that the marginal pdf of Z is a Gaussian pdf with mean given by the first component of $\mathbf{n}$ and variance given by the 1-1 component of the covariance matrix C. By carrying out the above matrix multiplications, we find that

$$E[Z] = \sum_{i=1}^{n} a_i E[X_i] \tag{6.47a}$$

$$\text{VAR}[Z] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \, \text{COV}(X_i, X_j). \tag{6.47b}$$

6.4.2   Joint Characteristic Function of a Gaussian Random Variable

The joint characteristic function is very useful in developing the properties of jointly Gaussian random variables. We now show that **the joint characteristic function of n jointly Gaussian random variables $X_1, X_2, \ldots, X_n$** is given by

$$\Phi_{X_1, X_2, \ldots, X_n}(\omega_1, \omega_2, \ldots, \omega_n) = e^{j\sum_{i=1}^{n}\omega_i m_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}\omega_i\omega_k \text{COV}(X_i, X_k)}, \qquad (6.48a)$$

which can be written more compactly as follows:

$$\Phi_{\mathbf{X}}(\boldsymbol{\omega}) \triangleq \Phi_{X_1, X_2, \ldots, X_n}(\omega_1, \omega_2, \ldots, \omega_n) = e^{j\boldsymbol{\omega}^{\mathrm{T}}\mathbf{m} - \frac{1}{2}\boldsymbol{\omega}^{\mathrm{T}}K\boldsymbol{\omega}}, \qquad (6.48b)$$

where $\mathbf{m}$ is the vector of means and K is the covariance matrix defined in Eq. (6.42b).

Equation (6.48) can be verified by direct integration (see Problem 6.65). We use the approach in [Papoulis] to develop Eq. (6.48) by using the result from Example 6.24 that a linear combination of jointly Gaussian random variables is always Gaussian. Consider the sum

$$Z = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n.$$

The characteristic function of Z is given by

$$\Phi_Z(\omega) = E[e^{j\omega Z}] = E[e^{j(\omega a_1 X_1 + \omega a_2 X_2 + \cdots + \omega a_n X_n)}]$$

$$= \Phi_{X_1, \ldots, X_n}(a_1\omega, a_2\omega, \ldots, a_n\omega).$$

On the other hand, since Z is a Gaussian random variable with mean and variance given Eq. (6.47), we have

$$\Phi_Z(\omega) = e^{j\omega E[Z] - \frac{1}{2}\text{VAR}[Z]\omega^2}$$

$$= e^{j\omega\sum_{i=1}^{n}a_i m_i - \frac{1}{2}\omega^2\sum_{i=1}^{n}\sum_{k=1}^{n}a_i a_k \text{COV}(X_i, X_k)}. \qquad (6.49)$$

By equating both expressions for $\Phi_Z(\omega)$ with $\omega = 1$, we finally obtain

$$\Phi_{X_1, X_2, \ldots, X_n}(a_1, a_2, \ldots, a_n) = e^{j\sum_{i=1}^{n}a_i m_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n}a_i a_k \text{COV}(X_i, X_k)}$$

$$= e^{j\mathbf{a}^{\mathrm{T}}\mathbf{m} - \frac{1}{2}\mathbf{a}^{\mathrm{T}}K\mathbf{a}}. \qquad (6.50)$$

By replacing the a_i's with ω_i's we obtain Eq. (6.48).

The marginal characteristic function of any subset of the random variables $X_1, X_2, \ldots, X_n$ can be obtained by setting appropriate ω_i's to zero. Thus, for example, the marginal characteristic function of $X_1, X_2, \ldots, X_m$ for $m < n$ is obtained by setting $\omega_{m+1} = \omega_{m+2} = \cdots = \omega_n = 0$. Note that the resulting characteristic function again corresponds to that of jointly Gaussian random variables with mean and covariance terms corresponding the reduced set $X_1, X_2, \ldots, X_m$.

The derivation leading to Eq. (6.50) suggests an alternative definition for jointly Gaussian random vectors:

Definition: $\mathbf{X}$ is a jointly Gaussian random vector if and only every linear combination $Z = \boldsymbol{a}^{\mathrm{T}}\mathbf{X}$ is a Gaussian random variable.

In Example 6.24 we showed that if $\mathbf{X}$ is a jointly Gaussian random vector then the linear combination $Z = \boldsymbol{a}^{\mathrm{T}}\mathbf{X}$ is a Gaussian random variable. Suppose that we do not know the joint pdf of $\mathbf{X}$ but we are given that $Z = \boldsymbol{a}^{\mathrm{T}}\mathbf{X}$ is a Gaussian random variable for any choice of coefficients $\boldsymbol{a}^{\mathrm{T}} = (a_1, a_2, \ldots, a_n)$. This implies that Eqs. (6.48) and (6.49) hold, which together imply Eq. (6.50) which states that $\mathbf{X}$ has the characteristic function of a jointly Gaussian random vector.

The above definition is slightly broader than the definition using the pdf in Eq. (6.44). The definition based on the pdf requires that the covariance in the exponent be invertible. The above definition leads to the characteristic function of Eq. (6.50) which does not require that the covariance be invertible. Thus the above definition allows for cases where the covariance matrix is not invertible.

6.5   ESTIMATION OF RANDOM VARIABLES

In this book we will encounter two basic types of estimation problems. In the first type, we are interested in *estimating the parameters of one or more random variables,* e.g., probabilities, means, variances, or covariances. In Chapter 1, we stated that relative frequencies can be used to estimate the probabilities of events, and that sample averages can be used to estimate the mean and other moments of a random variable. In Chapters 7 and 8 we will consider this type of estimation further. In this section, we are concerned with the second type of estimation problem, where we are interested in *estimating the value of an inaccessible random variable X in terms of the observation of an accessible random variable Y.* For example, X could be the input to a communication channel and Y could be the observed output. In a prediction application, X could be a future value of some quantity and Y its present value.

6.5.1   MAP and ML Estimators

We have considered estimation problems informally earlier in the book. For example, in estimating the output of a discrete communications channel we are interested in finding the most probable input given the observation $Y = y$, that is, the value of input x that maximizes $P[X = x \,|\, Y = y]$:

$$\max_x P[X = x \,|\, Y = y].$$

In general we refer to the above estimator for X in terms of Y as the **maximum a posteriori (MAP) estimator**. The a posteriori probability is given by:

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

and so the MAP estimator requires that we know the a priori probabilities $P[X = x]$. In some situations we know $P[Y = y \,|\, X = x]$ but we do not know the a priori probabilities, so we select the estimator value x as the value that maximizes the likelihood of the observed value $Y = y$:

$$\max_x P[Y = y \,|\, X = x].$$

We refer to this estimator of X in terms of Y as the **maximum likelihood (ML) estimator**.

We can define MAP and ML estimators when X and Y are continuous random variables by replacing events of the form $\{Y = y\}$ by $\{y < Y < y + dy\}$. If X and Y are continuous, the **MAP estimator for X given the observation Y** is given by:

$$\max_{x} f_X(X = x \,|\, Y = y),$$

and the **ML estimator for X given the observation Y** is given by:

$$\max_{x} f_X(Y = y \,|\, X = x).$$

Example 6.25   Comparison of ML and MAP Estimators

Let X and Y be the random pair in Example 5.16. Find the MAP and ML estimators for X in terms of Y.

From Example 5.32, the conditional pdf of X given Y is given by:

$$f_X(x \,|\, y) = e^{-(x-y)} \quad \text{for} \quad y \le x$$

which decreases as x increases beyond y. Therefore the MAP estimator is $\hat{X}_{\text{MAP}} = y$. On the other hand, the conditional pdf of Y given X is:

$$f_Y(y \,|\, x) = \frac{e^{-y}}{1 - e^{-x}} \quad \text{for} \quad 0 < y \le x.$$

As x increases beyond y, the denominator becomes larger so the conditional pdf decreases. Therefore the ML estimator is $\hat{X}_{\text{ML}} = y$. In this example the ML and MAP estimators agree.

Example 6.26   Jointly Gaussian Random Variables

Find the MAP and ML estimator of X in terms of Y when X and Y are jointly Gaussian random variables.

The conditional pdf of X given Y is given by:

$$f_X(x \,|\, y) = \frac{\exp\left\{ -\dfrac{1}{2(1 - \rho^2)\sigma_X^2} \left(x - \rho\dfrac{\sigma_X}{\sigma_Y}(y - m_Y) - m_X \right)^2 \right\}}{\sqrt{2\pi\sigma_X^2\,(1 - \rho^2)}}$$

which is maximized by the value of x for which the exponent is zero. Therefore

$$\hat{X}_{\text{MAP}} = \rho\frac{\sigma_X}{\sigma_Y}(y - m_Y) + m_X.$$

The conditional pdf of Y given X is:

$$f_Y(y \,|\, x) = \frac{\exp\left\{ -\dfrac{1}{2(1 - \rho^2)\sigma_Y^2} \left(y - \rho\dfrac{\sigma_Y}{\sigma_X}(x - m_X) - m_Y \right)^2 \right\}}{\sqrt{2\pi\sigma_Y^2\,(1 - \rho^2)}}.$$

which is also maximized for the value of x for which the exponent is zero:

$$0 = y - \rho\frac{\sigma_Y}{\sigma_X}(x - m_X) - m_Y.$$

The ML estimator for X given $Y = y$ is then:

$$\hat{X}_{ML} = \frac{\sigma_X}{\rho \sigma_Y}(y - m_Y) + m_X.$$

Therefore we conclude that $\hat{X}_{ML} \neq \hat{X}_{MAP}$. In other words, knowledge of the a priori probabilities of X will affect the estimator.

6.5.2   Minimum MSE Linear Estimator

The estimate for X is given by a function of the observation $\hat{X} = g(Y)$. In general, the *estimation error*, $X - \hat{X} = X - g(Y)$, is nonzero, and there is a *cost* associated with the error, $c(X - g(Y))$. We are usually interested in finding the function $g(Y)$ that minimizes the expected value of the cost, $E[c(X - g(Y))]$. For example, if X and Y are the discrete input and output of a communication channel, and c is zero when $X = g(Y)$ and one otherwise, then the expected value of the cost corresponds to the probability of error, that is, that $X \neq g(Y)$. When X and Y are continuous random variables, we frequently use the **mean square error (MSE)** as the cost:

$$e = E[(X - g(Y))^2].$$

In the remainder of this section we focus on this particular cost function. We first consider the case where $g(Y)$ is constrained to be a linear function of Y, and then consider the case where $g(Y)$ can be any function, whether linear or nonlinear.

First, consider the problem of estimating a random variable X by a *constant a* so that the mean square error is minimized:

$$\min_a E[(X - a)^2] = E[X^2] - 2aE[X] + a^2. \tag{6.51}$$

The best a is found by taking the derivative with respect to a, setting the result to zero, and solving for a. The result is

$$a^* = E[X], \tag{6.52}$$

which makes sense since the expected value of X is the center of mass of the pdf. The mean square error for this estimator is equal to $E[(X - a^*)^2] = \text{VAR}(X)$.

Now consider estimating X by a *linear* function $g(Y) = aY + b$:

$$\min_{a,b} E[(X - aY - b)^2]. \tag{6.53a}$$

Equation (6.53a) can be viewed as the approximation of $X - aY$ by the constant b. This is the minimization posed in Eq. (6.51) and the best b is

$$b^* = E[X - aY] = E[X] - aE[Y]. \tag{6.53b}$$

Substitution into Eq. (6.53a) implies that the best a is found by

$$\min_a E[\{(X - E[X]) - a(Y - E[Y])\}^2].$$

We once again differentiate with respect to a, set the result to zero, and solve for a:

$$0 = \frac{d}{da} E[(X - E[X]) - a(Y - E[Y])^2]$$

$$= -2E[\{(X - E[X]) - a(Y - E[Y])\}(Y - E[Y])]$$
$$= -2(\text{COV}(X, Y) - a\text{VAR}(Y)). \tag{6.54}$$

The best coefficient a is found to be

$$a^* = \frac{\text{COV}(X, Y)}{\text{VAR}(Y)} = \rho_{X,Y} \frac{\sigma_X}{\sigma_Y},$$

where $\sigma_Y = \sqrt{\text{VAR}(Y)}$ and $\sigma_X = \sqrt{\text{VAR}(X)}$. Therefore, **_the minimum mean square error (mmse) linear estimator_** for X in terms of Y is

$$\hat{X} = a^*Y + b^*$$

$$= \rho_{X,Y}\sigma_X \frac{Y - E[Y]}{\sigma_Y} + E[X]. \tag{6.55}$$

The term $(Y - E[Y])/\sigma_Y$ is simply a zero-mean, unit-variance version of Y. Thus $\sigma_X(Y - E[Y])/\sigma_Y$ is a rescaled version of Y that has the variance of the random variable that is being estimated, namely σ_X^2. The term $E[X]$ simply ensures that the estimator has the correct mean. The key term in the above estimator is the correlation coefficient: $\rho_{X,Y}$ *specifies the sign and extent of the estimate of Y relative to* $\sigma_X(Y - E[Y])/\sigma_Y$. If X and Y are uncorrelated (i.e., $\rho_{X,Y} = 0$) then the best estimate for X is its mean, $E[X]$. On the other hand, if $\rho_{X,Y} = \pm1$ then the best estimate is equal to $\pm\sigma_X(Y - E[Y])/\sigma_Y + E[X]$.

We draw our attention to the second equality in Eq. (6.54):

$$E[\{(X - E[X]) - a^*(Y - E[Y])\}(Y - E[Y])] = 0. \tag{6.56}$$

This equation is called the **orthogonality condition** because it states that the error of the best linear estimator, the quantity inside the braces, is orthogonal to the observation $Y - E[Y]$. The orthogonality condition is a fundamental result in mean square estimation.

The mean square error of the best *linear* estimator is

$$e_L^* = E[((X - E[X]) - a^*(Y - E[Y]))^2]$$

$$= E[((X - E[X]) - a^*(Y - E[Y]))(X - E[X])]$$

$$\quad - a^*E[((X - E[X]) - a^*(Y - E[Y]))(Y - E[Y])]$$

$$= E[((X - E[X]) - a^*(Y - E[Y]))(X - E[X])]$$

$$= \text{VAR}(X) - a^* \text{COV}(X, Y)$$

$$= \text{VAR}(X)(1 - \rho_{X,Y}^2) \tag{6.57}$$

where the second equality follows from the orthogonality condition. Note that when $|\rho_{X,Y}| = 1$, the mean square error is zero. This implies that $P[|X - a^*Y - b^*| = 0] = P[X = a^*Y + b^*] = 1$, so that X is essentially a linear function of Y.

6.5.3   Minimum MSE Estimator

In general the estimator for X that minimizes the mean square error is a *nonlinear* function of Y. The estimator $g(Y)$ that best approximates X in the sense of minimizing mean square error must satisfy

$$\underset{g(.)}{\text{minimize}}\ E[(X - g(Y))^2].$$

The problem can be solved by using **conditional expectation**:

$$E[(X - g(Y))^2] = E[E[(X - g(Y))^2 | Y]]$$

$$= \int_{-\infty}^{\infty} E[(X - g(Y))^2 | Y = y]f_Y(y)dy.$$

The integrand above is positive for all y; therefore, the integral is minimized by minimizing $E[(X - g(Y))^2 | Y = y]$ for each y. But $g(y)$ is a constant as far as the conditional expectation is concerned, so the problem is equivalent to Eq. (6.51) and the "constant" that minimizes $E[(X - g(y))^2 | Y = y]$ is

$$g^*(y) = E[X | Y = y]. \tag{6.58}$$

The function $g^*(y) = E[X | Y = y]$ is called the **regression curve** which simply traces the conditional expected value of X given the observation $Y = y$.

The mean square error of the best estimator is:

$$e^* = E[(X - g^*(Y))^2] = \int_R E[(X - E[X | y])^2 | Y = y]f_Y(y)\, dy$$

$$= \int_{R^n} \text{VAR}[X | Y = y]f_Y(y)\, dy.$$

Linear estimators in general are suboptimal and have larger mean square errors.

Example 6.27   Comparison of Linear and Minimum MSE Estimators

Let X and Y be the random pair in Example 5.16. Find the best linear and nonlinear estimators for X in terms of Y, and of Y in terms of X.

Example 5.28 provides the parameters needed for the linear estimator: $E[X] = 3/2$, $E[Y] = 1/2$, $\text{VAR}[X] = 5/4$, $\text{VAR}[Y] = 1/4$, and $\rho_{X,Y} = 1/\sqrt{5}$. Example 5.32 provides the conditional pdf's needed to find the nonlinear estimator. The best linear and nonlinear estimators for X in terms of Y are:

$$\hat{X} = \frac{1}{\sqrt{5}}\frac{\sqrt{5}}{2}\frac{Y - 1/2}{1/2} + \frac{3}{2} = Y + 1$$

$$E[X | y] = \int_y^{\infty} xe^{-(x-y)}\, dx = y + 1 \text{ and so } E[X | Y] = Y + 1.$$

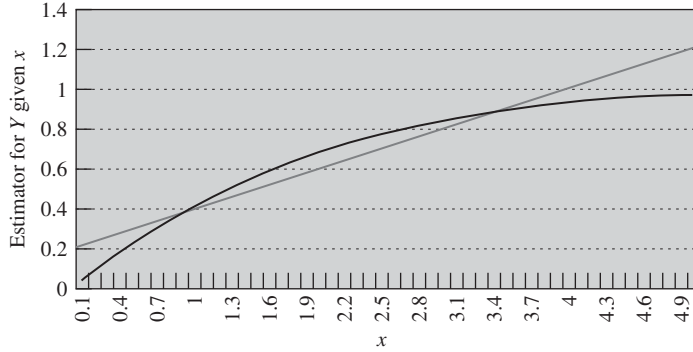Thus the optimum linear and nonlinear estimators are the same.

FIGURE 6.2
Comparison of linear and nonlinear estimators.

The best linear and nonlinear estimators for Y in terms of X are:

$$\hat{Y} = \frac{1}{\sqrt{5}}\frac{1}{2}\frac{X - 3/2}{\sqrt{5/2}} + \frac{1}{2} = (X + 1)/5.$$

$$E[Y\,|\,x] = \int_0^x y\frac{e^{-y}}{1 - e^{-x}}\,dy = \frac{1 - e^{-x} - xe^{-x}}{1 - e^{-x}} = 1 - \frac{xe^{-x}}{1 - e^{-x}}.$$

The optimum linear and nonlinear estimators are not the same in this case. Figure 6.2 compares the two estimators. It can be seen that the linear estimator is close to $E[Y\,|\,x]$ for lower values of x, where the joint pdf of X and Y are concentrated and that it diverges from $E[Y\,|\,x]$ for larger values of x.

Example 6.28

Let X be uniformly distributed in the interval $(-1, 1)$ and let $Y = X^2$. Find the best linear estimator for Y in terms of X. Compare its performance to the best estimator.

The mean of X is zero, and its correlation with Y is

$$E[XY] = E[XX^2] = \int_{-\frac{1}{2}}^{1} x^3/2\,dx = 0.$$

Therefore $\text{COV}(X, Y) = 0$ and the best linear estimator for Y is $E[Y]$ by Eq. (6.55). The mean square error of this estimator is the $\text{VAR}(Y)$ by Eq. (6.57).

The best estimator is given by Eq. (6.58):

$$E[Y\,|\,X = x] = E[X^2\,|\,X = x] = x^2.$$

The mean square error of this estimator is

$$E[(Y - g(X))^2] = E[(X^2 - X^2)^2] = 0.$$

Thus in this problem, the best linear estimator performs poorly while the nonlinear estimator gives the smallest possible mean square error, zero.

Example 6.29   Jointly Gaussian Random Variables

Find the minimum mean square error estimator of X in terms of Y when X and Y are jointly Gaussian random variables.

The minimum mean square error estimator is given by the conditional expectation of X given Y. From Eq. (5.63), we see that the conditional expectation of X given $Y = y$ is given by

$$E[X \,|\, Y = y] = E[X] + \rho_{X,Y} \frac{\sigma_X}{\sigma_Y}(Y - E[Y]).$$

This is identical to the best linear estimator. Thus *for jointly Gaussian random variables the minimum mean square error estimator is linear.*

6.5.4   Estimation Using a Vector of Observations

The MAP, ML, and mean square estimators can be extended to where a vector of observations is available. Here we focus on mean square estimation. We wish to estimate X by a function $g(\mathbf{Y})$ of a random vector of observations $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)^{\mathrm{T}}$ so that the **mean square error** is minimized:

$$\underset{g(\cdot)}{\text{minimize}} \; E[(X - g(\mathbf{Y}))^2].$$

To simplify the discussion we will assume that X and the Y_i have zero means. The same derivation that led to Eq. (6.58) leads to the *optimum minimum mean square estimator*:

$$g^*(\mathbf{y}) = E[X \,|\, \mathbf{Y} = \mathbf{y}]. \tag{6.59}$$

The minimum mean square error is then:

$$E[(X - g^*(Y))^2] = \int_{R^n} E[(X - E[X \,|\, \mathbf{Y}])^2 \,|\, \mathbf{Y} = \mathbf{y}] f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}$$

$$= \int_{R^n} \text{VAR}[X \,|\, \mathbf{Y} = \mathbf{y}] f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}.$$

Now suppose the estimate is a linear function of the observations:

$$g(\mathbf{Y}) = \sum_{k=1}^{n} a_k Y_k = \mathbf{a}^{\mathrm{T}} \mathbf{Y}.$$

The mean square error is now:

$$E[(X - g(\mathbf{Y}))^2] = E\left[\left(X - \sum_{k=1}^{n} a_k Y_k\right)^2\right].$$

We take derivatives with respect to a_k and again obtain the orthogonality conditions:

$$E\left[\left(X - \sum_{k=1}^{n} a_k Y_k\right) Y_j\right] = 0 \quad \text{for } j = 1, \ldots, n.$$

The orthogonality condition becomes:

$$E[XY_j] = E\left[\left(\sum_{k=1}^{n} a_k Y_k\right) Y_j\right] = \sum_{k=1}^{n} a_k E[Y_k Y_j] \quad \text{for } j = 1, \ldots, n.$$

We obtain a compact expression by introducing matrix notation:

$$E[X\mathbf{Y}] = \mathbf{R_Y} a \qquad \text{where } a = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}. \tag{6.60}$$

where $E[X\mathbf{Y}] = [E[XY_1], E[XY_2], \ldots, E[XY_n]]^{\mathrm{T}}$ and $\mathbf{R_Y}$ is the correlation matrix. Assuming $\mathbf{R_Y}$ is invertible, the optimum coefficients are:

$$a = \mathbf{R}_Y^{-1} E[X\mathbf{Y}]. \tag{6.61a}$$

We can use the methods from Section 6.3 to invert $\mathbf{R_Y}$. The mean square error of the optimum linear estimator is:

$$E[(X - a^{\mathrm{T}}\mathbf{Y})^2] = E[(X - a^{\mathrm{T}}\mathbf{Y})X] - E[(X - a^{\mathrm{T}}\mathbf{Y})a^{\mathrm{T}}\mathbf{Y}]$$
$$= E[(X - a^{\mathrm{T}}\mathbf{Y})X] = \mathrm{VAR}(X) - a^{\mathrm{T}}E[\mathbf{Y}X]. \tag{6.61b}$$

Now suppose that X has mean m_X and $\mathbf{Y}$ has mean vector $\mathbf{m_Y}$, so our estimator now has the form:

$$\hat{X} = g(\mathbf{Y}) = \sum_{k=1}^{n} a_k Y_k + b = a^{\mathrm{T}}\mathbf{Y} + b. \tag{6.62}$$

The same argument that led to Eq. (6.53b) implies that the optimum choice for b is:

$$b = E[X] - a^{\mathrm{T}}\mathbf{m_Y}.$$

Therefore the optimum linear estimator has the form:

$$\hat{X} = g(\mathbf{Y}) = a^{\mathrm{T}}(\mathbf{Y} - \mathbf{m_Y}) + m_X = a^{\mathrm{T}}\mathbf{Z} + m_X$$

where $\mathbf{Z} = \mathbf{Y} - \mathbf{m_Y}$ is a random vector with zero mean vector. The mean square error for this estimator is:

$$E[(X - g(\mathbf{Y}))^2] = E[(X - a^{\mathrm{T}}\mathbf{Z} - m_X)^2] = E[(W - a^{\mathrm{T}}\mathbf{Z})^2]$$

where $W = X - m_X$ has zero mean. We have reduced the general estimation problem to one with zero mean random variables, i.e., W and $\mathbf{Z}$, which has solution given by Eq. (6.61a). Therefore the optimum set of linear predictors is given by:

$$a = \mathbf{R}_z^{-1} E[W\mathbf{Z}] = \mathbf{K}_Y^{-1} E[(X - m_X)(\mathbf{Y} - \mathbf{m_Y})]. \tag{6.63a}$$

The mean square error is:

$$E[(X - a^{\mathrm{T}}\mathbf{Y} - b)^2] = E[(W - a^{\mathrm{T}}\mathbf{Z}\,W] = \mathrm{VAR}(W) - a^{\mathrm{T}}E[W\mathbf{Z}]$$
$$= \mathrm{VAR}(X) - a^{\mathrm{T}}E[(X - \mathbf{m_X})(\mathbf{Y} - \mathbf{m_Y})]. \tag{6.63b}$$

This result is of particular importance in the case where X and $\mathbf{Y}$ are jointly Gaussian random variables. In Example 6.23 we saw that the conditional expected value

of X given Y is a linear function of Y of the form in Eq. (6.62). Therefore in this case the optimum minimum mean square estimator corresponds to the optimum linear estimator.

Example 6.30   Diversity Receiver

A radio receiver has two antennas to receive noisy versions of a signal X. The desired signal X is a Gaussian random variable with zero mean and variance 2. The signals received in the first and second antennas are $Y_1 = X + N_1$ and $Y_2 = X + N_2$ where N_1 and N_2 are zero-mean, unit-variance Gaussian random variables. In addition, X, N_1, and N_2 are independent random variables. Find the optimum mean square error linear estimator for X based on a single antenna signal and the corresponding mean square error. Compare the results to the optimum mean square estimator for X based on both antenna signals $\mathbf{Y} = (Y_1, Y_2)$.

   Since all random variables have zero mean, we only need the correlation matrix and the cross-correlation vector in Eq. (6.61):

$$\mathbf{R}_Y = \begin{bmatrix} E[Y_1^2] & E[Y_1Y_2] \\ E[Y_1Y_2] & E[Y_2^2] \end{bmatrix}$$

$$= \begin{bmatrix} E[(X + N_1)^2] & E[(X + N_1)(X + N_2)] \\ E[(X + N_1)(X + N_2)] & E[(X + N_2)^2] \end{bmatrix}$$

$$= \begin{bmatrix} E[X^2] + E[N_1^2] & E[X^2] \\ E[X^2] & E[X^2] + E[N_2^2] \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

and

$$E[X\mathbf{Y}] = \begin{bmatrix} E[XY_1] \\ E[XY_2] \end{bmatrix} = \begin{bmatrix} E[X^2] \\ E[X^2] \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

The optimum estimator using a single antenna received signal involves solving the 1×1 version of the above system:

$$\hat{X} = \frac{E[X^2]}{E[X^2] + E[N_1^2]} Y_1 = \frac{2}{3} Y_1$$

and the associated mean square error is:

$$\mathrm{VAR}(X) - a^* \,\mathrm{COV}(Y_1, X) = 2 - \frac{2}{3} 2 = \frac{2}{3}.$$

The coefficients of the optimum estimator using two antenna signals are:

$$\boldsymbol{a} = \mathbf{R_Y}^{-1} E[X\mathbf{Y}] = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 3 & -2 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.4 \end{bmatrix}$$

and the optimum estimator is:

$$\hat{X} = 0.4Y_1 + 0.4Y_2.$$

The mean square error for the two antenna estimator is:

$$E[(X - \boldsymbol{a}^{\mathrm{T}}\mathbf{Y})^2] = \mathrm{VAR}(X) - \boldsymbol{a}^{\mathrm{T}} E[\mathbf{Y}X] = 2 - [0.4, 0.4] \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 0.4.$$

As expected, the two antenna system has a smaller mean square error. Note that the receiver adds the two received signals and scales the result by 0.4. The sum of the signals is:

$$\hat{X} = 0.4Y_1 + 0.4Y_2 = 0.4(2X + N_1 + N_2) = 0.8\left(X + \frac{N_1 + N_2}{2}\right)$$

so combining the signals keeps the desired signal portion, X, constant while averaging the two noise signals N_1 and N_2. The problems at the end of the chapter explore this topic further.

Example 6.31   Second-Order Prediction of Speech

Let $X_1, X_2, \ldots$ be a sequence of samples of a speech voltage waveform, and suppose that the samples are fed into the second-order predictor shown in Fig. 6.3. Find the set of predictor coefficients a and b that minimize the mean square value of the predictor error when X_n is estimated by $aX_{n-2} + bX_{n-1}$.

We find the best predictor for X_1, X_2, and X_3 and assume that the situation is identical for X_2, X_3, and X_4 and so on. It is common practice to model speech samples as having zero mean and variance σ^2, and a covariance that does not depend on the specific index of the samples, but rather on the separation between them:

$$\text{COV}(X_j, X_k) = \rho_{|j-k|}\sigma^2.$$

The equation for the optimum linear predictor coefficients becomes

$$\sigma^2\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \sigma^2\begin{bmatrix} \rho_2 \\ \rho_1 \end{bmatrix}.$$

Equation (6.61a) gives

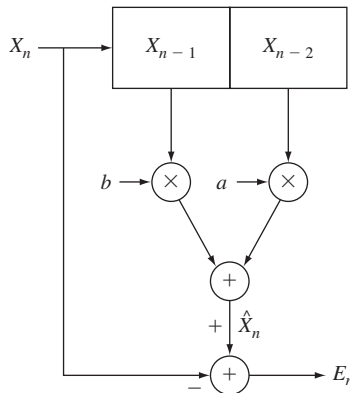$$a = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} \quad \text{and} \quad b = \frac{\rho_1(1 - \rho_1^2)}{1 - \rho_1^2}.$$

FIGURE 6.3
A two-tap linear predictor for processing speech.

In Problem 6.78, you are asked to show that the mean square error using the above values of a and b is

$$\sigma^2 \left\{ 1 - \rho_1^2 - \frac{(\rho_1^2 - \rho_2)^2}{1 - \rho_1^2} \right\}. \tag{6.64}$$

Typical values for speech signals are $\rho_1 = .825$ and $\rho_2 = .562$. The mean square value of the predictor output is then $.281\sigma^2$. The lower variance of the output $(.281\sigma^2)$ relative to the input variance (σ^2) shows that the linear predictor is effective in anticipating the next sample in terms of the two previous samples. The order of the predictor can be increased by using more terms in the linear predictor. Thus a third-order predictor has three terms and involves inverting a 3×3 correlation matrix, and an n-th order predictor will involve an $n \times n$ matrix. Linear predictive techniques are used extensively in speech, audio, image and video compression systems. We discuss linear prediction methods in greater detail in Chapter 10.

---

*6.6   GENERATING CORRELATED VECTOR RANDOM VARIABLES

Many applications involve vectors or sequences of correlated random variables. Computer simulation models of such applications therefore require methods for generating such random variables. In this section we present methods for generating vectors of random variables with specified covariance matrices. We also discuss the generation of jointly Gaussian vector random variables.

6.6.1   Generating Random Vectors with Specified Covariance Matrix

Suppose we wish to generate a random vector $\mathbf{Y}$ with an arbitrary valid covariance matrix $\mathbf{K_Y}$. Let $\mathbf{Y} = \mathbf{A^T X}$ as in Example 6.17, where $\mathbf{X}$ is a vector random variable with components that are uncorrelated, zero mean, and unit variance. $\mathbf{X}$ has covariance matrix equal to the identity matrix $\mathbf{K_X} = \mathbf{I}$, $\mathbf{m_Y} = \mathbf{Am_X} = \mathbf{0}$, and

$$\mathbf{K_Y} = \mathbf{A^T K_X A} = \mathbf{A^T A}.$$

Let $\mathbf{P}$ be the matrix whose columns are the eigenvectors of $\mathbf{K_Y}$ and let $\mathbf{\Lambda}$ be the diagonal matrix of eigenvalues, then from Eq. (6.39b) we have:

$$\mathbf{P^T K_Y P} = \mathbf{P^T P \Lambda} = \mathbf{\Lambda}.$$

If we premultiply the above equation by $\mathbf{P}$ and then postmultiply by $\mathbf{P^T}$, we obtain expression for an arbitrary covariance matrix $\mathbf{K_Y}$ in terms of its eigenvalues and eigenvectors:

$$\mathbf{P \Lambda P^T} = \mathbf{P P^T K_Y P P^T} = \mathbf{K_Y}. \tag{6.65}$$

Define the matrix $\mathbf{\Lambda}^{1/2}$ as the diagonal matrix of square roots of the eigenvalues:

$$\mathbf{\Lambda}^{1/2} \triangleq \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & \sqrt{\lambda_n} \end{bmatrix}.$$

In Problem 6.53 we show that any covariance matrix $\mathbf{K_Y}$ is positive semi-definite, which implies that it has nonnegative eigenvalues, and so taking the square root is always possible. If we now let

$$\mathbf{A} = (\mathbf{P}\Lambda^{1/2})^\mathrm{T} \tag{6.66}$$

then

$$\mathbf{A}^\mathrm{T}\mathbf{A} = \mathbf{P}\Lambda^{1/2}\Lambda^{1/2}\mathbf{P}^\mathrm{T} = \mathbf{P}\Lambda\mathbf{P}^\mathrm{T} = \mathbf{K_Y}.$$

Therefore $\mathbf{Y}$ has the desired covariance matrix $\mathbf{K_Y}$.

Example 6.32

Let $\mathbf{X} = (X_1, X_2)$ consist of two zero-mean, unit-variance, uncorrelated random variables. Find the matrix A such that $\mathbf{Y} = A\mathbf{X}$ has covariance matrix

$$K = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}.$$

First we need to find the eigenvalues of K which are determined from the following equation:

$$\det(\mathbf{K} - \lambda\mathbf{I}) = 0 = \det\begin{bmatrix} 4 - \lambda & 2 \\ 2 & 4 - \lambda \end{bmatrix} = (4 - \lambda)^2 - 4 = \lambda^2 - 8\lambda + 12$$

$$= (\lambda - 6)(\lambda - 2).$$

We find the eigenvalues to be $\lambda_1 = 2$ and $\lambda_2 = 6$. Next we need to find the eigenvectors corresponding to each eigenvalue:

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \lambda_1\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = 2\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

which implies that $2e_1 + 2e_2 = 0$. Thus any vector of the form $[1, -1]^\mathrm{T}$ is an eigenvector. We choose the normalized eigenvector corresponding to $\lambda_1 = 2$ as $\mathbf{e}_1 = [1/\sqrt{2}, -1/\sqrt{2}]^\mathrm{T}$. We similarly find the eigenvector corresponding to $\lambda_2 = 6$ as $\mathbf{e}_2 = [1/\sqrt{2}, 1/\sqrt{2}]^\mathrm{T}$.

The method developed in Section 6.3 requires that we form the matrix P whose columns consist of the eigenvectors of K:

$$P = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

Next it requires that we form the diagonal matrix with elements equal to the square root of the eigenvalues:

$$\Lambda^{1/2} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{6} \end{bmatrix}.$$

The desired matrix is then

$$\mathbf{A} = \mathbf{P}\Lambda^{1/2} = \begin{bmatrix} 1 & \sqrt{3} \\ -1 & \sqrt{3} \end{bmatrix}.$$

You should verify that $K = AA^\mathrm{T}$.

Example 6.33

Use Octave to find the eigenvalues and eigenvectors calculated in the previous example.

After entering the matrix K, we use the `eig(K)` function to find the matrix of eigenvectors P and eigenvalues Λ. We then find A and its transpose A^T. Finally we confirm that $A^T A$ gives the desired covariance matrix.

```
> K=[4, 2; 2, 4];
> [P,D] =eig (K)
P =
 -0.70711  0.70711
  0.70711  0.70711
D =
  2  0
  0  6
> A=(P*sqrt(D))'
A =
 -1.0000  1.0000
  1.7321  1.7321
> A'
ans =
 -1.0000  1.7321
  1.0000  1.7321
> A'*A
ans =
  4.0000  2.0000
  2.0000  4.0000
```

The above steps can be used to find the transformation A^T for any desired covariance matrix K. The only check required is to ascertain that K is a valid covariance matrix: (1) K is symmetric (trivial); (2) K has positive eigenvalues (easy to check numerically).

6.6.2 Generating Vectors of Jointly Gaussian Random Variables

In Section 6.4 we found that if $\mathbf{X}$ is a vector of jointly Gaussian random variables with covariance K_X, then $\mathbf{Y} = A\mathbf{X}$ is also jointly Gaussian with covariance matrix $K_Y = AK_X A^T$. If we assume that $\mathbf{X}$ consists of unit-variance, uncorrelated random variables, then $K_X = I$, the identity matrix, and therefore $K_Y = AA^T$.

We can use the method from the first part of this section to find A for any desired covariance matrix K_Y. We generate jointly Gaussian random vectors $\mathbf{Y}$ with arbitrary covariance matrix K_Y and mean vector $\mathbf{m_Y}$ as follows:

1. Find a matrix A such that $K_Y = AA^T$.
2. Use the method from Section 5.10 to generate $\mathbf{X}$ consisting of n independent, zero-mean, Gaussian random variables.
3. Let $\mathbf{Y} = A\mathbf{X} + \mathbf{m_Y}$.

Example 6.34

The Octave commands below show necessary steps for generating the Gaussian random variables with the covariance matrix from Example 6.30.

```
> U1=rand(1000, 1);            % Create a 1000-element vector U₁.
> U2=rand(1000, 1);            % Create a 1000-element vector U₂.
> R2=−2*log(U1);               % Find R².
> TH=2*pi*U2;                  % Find Θ.

> X1=sqrt(R2).*sin(TH);        % Generate X1.
> X2=sqrt(R2).*cos(TH);        % Generate X2.
> Y1=X1+sqrt(3)*X2             % Generate Y1.
> Y2=−X1+sqrt(3)*X2            % Generate Y2.
> plot(Y1,Y2,'+')             % Plot scattergram.
```

We plotted the Y_1 values vs. the Y_2 values for 1000 pairs of generated random variables in a scattergram as shown in Fig. 6.4. Good agreement with the elliptical symmetry of the desired jointly Gaussian pdf is observed.
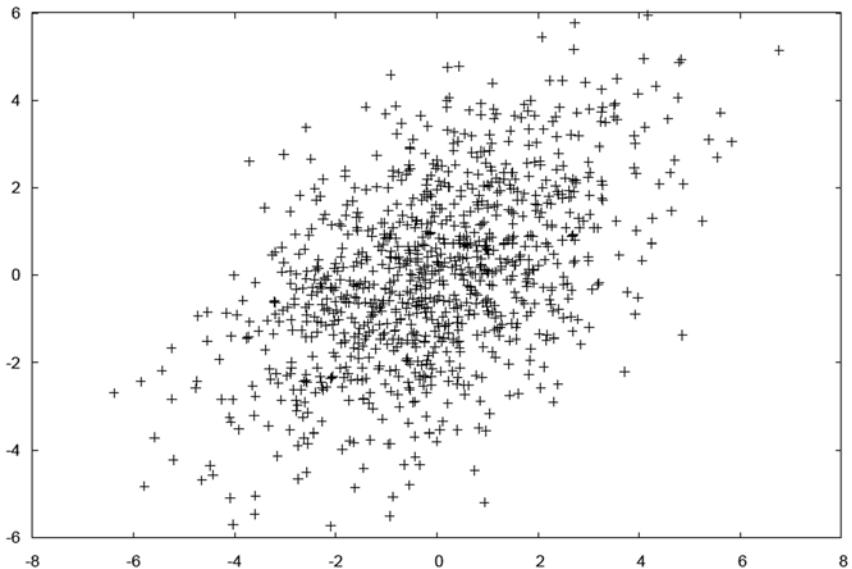
FIGURE 6.4
Scattergram of jointly Gaussian random variables.

SUMMARY

- The joint statistical behavior of a vector of random variables **X** is specified by the joint cumulative distribution function, the joint probability mass function, or the joint probability density function. The probability of any event involving the joint behavior of these random variables can be computed from these functions.

- The statistical behavior of subsets of random variables from a vector **X** is specified by the marginal cdf, marginal pdf, or marginal pmf that can be obtained from the joint cdf, joint pdf, or joint pmf of **X**.

- A set of random variables is independent if the probability of a product-form event is equal to the product of the probabilities of the component events. Equivalent conditions for the independence of a set of random variables are that the joint cdf, joint pdf, or joint pmf factors into the product of the corresponding marginal functions.

- The statistical behavior of a subset of random variables from a vector **X**, given the exact values of the other random variables in the vector, is specified by the conditional cdf, conditional pmf, or conditional pdf. Many problems naturally lend themselves to a solution that involves conditioning on the values of some of the random variables. In these problems, the expected value of random variables can be obtained through the use of conditional expectation.

- The mean vector and the covariance matrix provide summary information about a vector random variable. The joint characteristic function contains all of the information provided by the joint pdf.

- Transformations of vector random variables generate other vector random variables. Standard methods are available for finding the joint distributions of the new random vectors.

- The orthogonality condition provides a set of linear equations for finding the minimum mean square linear estimate. The best mean square estimator is given by the conditional expected value.

- The joint pdf of a vector **X** of jointly Gaussian random variables is determined by the vector of the means and by the covariance matrix. All marginal pdf's and conditional pdf's of subsets of **X** have Gaussian pdf's. Any linear function or linear transformation of jointly Gaussian random variables will result in a set of jointly Gaussian random variables.

- A vector of random variables with an arbitrary covariance matrix can be generated by taking a linear transformation of a vector of unit-variance, uncorrelated random variables. A vector of Gaussian random variables with an arbitrary covariance matrix can be generated by taking a linear transformation of a vector of independent, unit-variance jointly Gaussian random variables.

CHECKLIST OF IMPORTANT TERMS

Conditional cdf
Conditional expectation
Conditional pdf
Conditional pmf
Correlation matrix
Covariance matrix
Independent random variables
Jacobian of a transformation
Joint cdf
Joint characteristic function
Joint pdf
Joint pmf
Jointly continuous random variables
Jointly Gaussian random variables

Karhunen-Loeve expansion
MAP estimator
Marginal cdf
Marginal pdf
Marginal pmf
Maximum likelihood estimator
Mean square error
Mean vector
MMSE linear estimator
Orthogonality condition
Product-form event
Regression curve
Vector random variables

ANNOTATED REFERENCES

Reference [3] provides excellent coverage on linear transformation and jointly Gaussian random variables. Reference [5] provides excellent coverage of vector random variables. The book by Anton [6] provides an accessible introduction to linear algebra.

1. A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 2002.

2. N. Johnson et al., *Continuous Multivariate Distributions*, Wiley, New York, 2000.

3. H. Cramer, *Mathematical Methods of Statistics*, Princeton Press, 1999.

4. R. Gray and L.D. Davisson, *An Introduction to Statistical Signal Processing*, Cambridge Univ. Press, Cambridge, UK, 2005.

5. H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice Hall, Englewood Cliffs, N.J., 1986.

6. H. Anton, *Elementary Linear Algebra*, 9th ed., Wiley, New York, 2005.

7. C. H. Edwards, Jr., and D. E. Penney, *Calculus and Analytic Geometry*, 4th ed., Prentice Hall, Englewood Cliffs, N.J., 1984.

PROBLEMS

Section 6.1: Vector Random Variables

6.1. The point $\mathbf{X} = (X, Y, Z)$ is uniformly distributed inside a sphere of radius 1 about the origin. Find the probability of the following events:
   (a) $\mathbf{X}$ is inside a sphere of radius r, $r > 0$.
   (b) $\mathbf{X}$ is inside a cube of length $2/\sqrt{3}$ centered about the origin.
   (c) All components of $\mathbf{X}$ are positive.
   (d) Z is negative.

6.2. A random sinusoid signal is given by $X(t) = A \sin(t)$ where A is a uniform random variable in the interval $[0, 1]$. Let $\mathbf{X} = (X(t_1), X(t_2), X(t_3))$ be samples of the signal taken at times t_1, t_2, and t_3.
   (a) Find the joint cdf of $\mathbf{X}$ in terms of the cdf of A if $t_1 = 0$, $t_2 = \pi/2$, and $t_3 = \pi$. Are $X(t_1)$, $X(t_2)$, $X(t_3)$ independent random variables?
   (b) Find the joint cdf of $\mathbf{X}$ for t_1, $t_2 = t_1 + \pi/2$, and $t_3 = t_1 + \pi$. Let $t_1 = \pi/6$.

6.3. Let the random variables X, Y, and Z be independent random variables. Find the following probabilities in terms of $F_X(x)$, $F_Y(y)$, and $F_Z(z)$.
   (a) $P[|X| < 5, Y < 4, Z^3 > 8]$.
   (b) $P[X = 5, Y < 0, Z > 1]$.
   (c) $P[\min(X, Y, Z) < 2]$.
   (d) $P[\max(X, Y, Z) > 6]$.

6.4. A radio transmitter sends a signal $s > 0$ to a receiver using three paths. The signals that arrive at the receiver along each path are:

$$X_1 = s + N_1, X_2 = s + N_2, \text{and } X_3 = s + N_3,$$

where N_1, N_2, and N_3 are independent Gaussian random variables with zero mean and unit variance.
   (a) Find the joint pdf of $\mathbf{X} = (X_1, X_2, X_3)$. Are X_1, X_2, and X_3 independent random variables?
   (b) Find the probability that the minimum of all three signals is positive.
   (c) Find the probability that a majority of the signals are positive.

6.5. An urn contains one black ball and two white balls. Three balls are drawn from the urn. Let $I_k = 1$ if the outcome of the kth draw is the black ball and let $I_k = 0$ otherwise. Define the following three random variables:

$$X = I_1 + I_2 + I_3,$$
$$Y = \min\{I_1, I_2, I_3\},$$
$$Z = \max\{I_1, I_2, I_3\}.$$

   (a) Specify the range of values of the triplet (X, Y, Z) if each ball is put back into the urn after each draw; find the joint pmf for (X, Y, Z).
   (b) In part a, are X, Y, and Z independent? Are X and Y independent?
   (c) Repeat part a if each ball is not put back into the urn after each draw.

6.6. Consider the packet switch in Example 6.1. Suppose that each input has one packet with probability p and no packets with probability $1 - p$. Packets are equally likely to be

destined to each of the outputs. Let X_1, X_2 and X_3 be the number of packet arrivals destined for output 1, 2, and 3, respectively.

   (a) Find the joint pmf of X_1, X_2, and X_3 Hint: Imagine that every input has a *packet* go to a fictional port 4 with probality $1 - p$.

   (b) Find the joint pmf of X_1 and X_2.

   (c) Find the pmf of X_2.

   (d) Are X_1, X_2, and X_3 independent random variables?

   (e) Suppose that each output will accept at most one packet and discard all additional packets destined to it. Find the average number of packets discarded by the module in each T-second period.

6.7. Let X, Y, Z have joint pdf

$$f_{X,Y,Z}(x, y, z) = k(x + y + z) \quad \text{for} \quad 0 \le x \le 1, 0 \le y \le 1, 0 \le z \le 1.$$

   (a) Find k.

   (b) Find $f_X(x \mid y, z)$ and $f_Z(z \mid x, y)$.

   (c) Find $f_X(x)$, $f_Y(y)$, and $f_Z(z)$.

6.8. A point $\mathbf{X} = (X, Y, Z)$ is selected at random inside the unit sphere.

   (a) Find the marginal joint pdf of Y and Z.

   (b) Find the marginal pdf of Y.

   (c) Find the conditional joint pdf of X and Y given Z.

   (d) Are X, Y, and Z independent random variables?

   (e) Find the joint pdf of $\mathbf{X}$ given that the distance from $\mathbf{X}$ to the origin is greater than 1/2 and all the components of $\mathbf{X}$ are positive.

6.9. Show that $p_{X_1, X_2, X_3}(x_1, x_2, x_3) = p_{X_3}(x_3 \mid x_1, x_2) p_{X_2}(x_2 \mid x_1) p_{X_1}(x_1)$.

6.10. Let $X_1, X_2, \ldots, X_n$ be binary random variables taking on values 0 or 1 to denote whether a speaker is silent (0) or active (1). A silent speaker remains idle at the next time slot with probability 3/4, and an active speaker remains active with probability 1/2. Find the joint pmf for X_1, X_2, X_3, and the marginal pmf of X_3. Assume that the speaker begins in the silent state.

6.11. Show that $f_{X,Y,Z}(x, y, z) = f_Z(z \mid x, y) f_Y(y \mid x) f_X(x)$.

6.12. Let U_1, U_2, and U_3 be independent random variables and let $X = U_1, Y = U_1 + U_2$, and $Z = U_1 + U_2 + U_3$.

   (a) Use the result in Problem 6.11 to find the joint pdf of X, Y, and Z.

   (b) Let the U_i be independent uniform random variables in the interval $[0, 1]$. Find the marginal joint pdf of Y and Z. Find the marginal pdf of Z.

   (c) Let the U_i be independent zero-mean, unit-variance Gaussian random variables. Find the marginal pdf of Y and Z. Find the marginal pdf of Z.

6.13. Let X_1, X_2, and X_3 be the multiplicative sequence in Example 6.7.

   (a) Find, plot, and compare the marginal pdfs of X_1, X_2, and X_3.

   (b) Find the conditional pdf of X_3 given $X_1 = x$.

   (c) Find the conditional pdf of X_1 given $X_3 = z$.

6.14. Requests at an online music site are categorized as follows: Requests for most popular title with $p_1 = 1/2$; second most popular title with $p_2 = 1/4$; third most popular title with $p_3 = 1/8$; and other $p_4 = 1 - p_1 - p_2 - p_3 = 1/8$. Suppose there are a total number of

n requests in T seconds. Let X_k be the number of times category k occurs.

(a)   Find the joint pmf of (X_1, X_2, X_3).

(b)   Find the marginal pmf of (X_1, X_2). *Hint*: Use the binomial theorem.

(c)   Find the marginal pmf of X_1.

(d)   Find the conditional joint pmf of (X_2, X_3) given $X_1 = m$, where $0 \leq m \leq n$.

6.15. The number N of requests at the online music site in Problem 6.14 is a Poisson random variable with mean α customers per second. Let X_k be the number of type k requests in T seconds. Find the joint pmf of (X_1, X_2, X_3, X_4).

6.16. A random experiment has four possible outcomes. Suppose that the experiment is repeated n independent times and let X_k be the number of times outcome k occurs. The joint pmf of (X_1, X_2, X_3) is given by

$$p(k_1, k_2, k_3) = \frac{n! \, 3!}{(n+3)!} = \binom{n+3}{3}^{-1} \quad \text{for} \quad 0 \leq k_i \text{ and } k_1 + k_2 + k_3 \leq n.$$

(a)   Find the marginal pmf of (X_1, X_2).

(b)   Find the marginal pmf of X_1.

(c)   Find the conditional joint pmf of (X_2, X_3) given $X_1 = m$, where $0 \leq m \leq n$.

6.17. The number of requests of types 1, 2, and 3, respectively, arriving at a service station in t seconds are independent Poisson random variables with means $\lambda_1 t$, $\lambda_2 t$, and $\lambda_3 t$. Let N_1, N_2, and N_3 be the number of requests that arrive during an exponentially distributed time T with mean αt.

(a)   Find the joint pmf of N_1, N_2, and N_3.

(b)   Find the marginal pmf of N_1.

(c)   Find the conditional pmf of N_1 and N_2, given N_3.

Section 6.2: Functions of Several Random Variables

6.18. N devices are installed at the same time. Let Y be the time until the first device fails.

(a)   Find the pdf of Y if the lifetimes of the devices are independent and have the same Pareto distribution.

(b)   Repeat part a if the device lifetimes have a Weibull distribution.

6.19. In Problem 6.18 let $I_k(t)$ be the indicator function for the event "kth device is still working at time t." Let $N(t)$ be the number of devices still working at time t: $N(t) = I_1(t) + I_2(t) + \cdots + I_N(t)$. Find the pmf of $N(t)$ as well as its mean and variance.

6.20. A diversity receiver receives N independent versions of a signal. Each signal version has an amplitude X_k that is Rayleigh distributed. The receiver selects that signal with the largest amplitude X_k^2. A signal is not useful if the squared amplitude falls below a threshold γ. Find the probability that all N signals are below the threshold.

6.21. (Haykin) A receiver in a multiuser communication system accepts K binary signals from K independent transmitters: $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_K)$, where Y_k is the received signal from the kth transmitter. In an ideal system the received vector is given by:

$$\mathbf{Y} = \mathbf{Ab} + \mathbf{N}$$

where $\mathbf{A} = [\alpha_k]$ is a diagonal matrix of positive channel gains, $\mathbf{b} = (b_1, b_2, \ldots, b_K)$ is the vector of bits from each of the transmitters where $b_k = \pm 1$, and $\mathbf{N}$ is a vector of $\mathbf{K}$

independent zero-mean, unit-variance Gaussian random variables.

(a) Find the joint pdf of **Y**.

(b) Suppose $\mathbf{b} = (1, 1, \ldots, 1)$, find the probability that all components of **Y** are positive.

6.22. (a) Find the joint pdf of $U = X_1$, $V = X_1 + X_2$, and $W = X_1 + X_2 + X_3$.

(b) Evaluate the joint pdf of (U, V, W) if the X_i are independent zero-mean, unit variance Gaussian random variables.

(c) Find the marginal pdf of V and of W.

6.23. (a) Find the joint pdf of the sample mean and variance of two random variables:

$$M = \frac{X_1 + X_2}{2} \qquad V = \frac{(X_1 - M)^2 + (X_2 - M)^2}{2}$$

in terms of the joint pdf of X_1 and X_2.

(b) Evaluate the joint pdf if X_1 and X_2 are independent Gaussian random variables with the same mean 1 and variance 1.

(c) Evaluate the joint pdf if X_1 and X_2 are independent exponential random variables with the same parameter 1.

6.24. (a) Use the auxiliary variable method to find the pdf of

$$Z = \frac{X}{X + Y}.$$

(b) Find the pdf of Z if X and Y are independent exponential random variables with the parameter 1.

(c) Repeat part b if X and Y are independent Pareto random variables with parameters $k = 2$ and $x_m = 1$.

6.25. Repeat Problem 6.24 parts a and b for $Z = X/Y$.

6.26. Let X and Y be zero-mean, unit-variance Gaussian random variables with correlation coefficient 1/2. Find the joint pdf of $U = X^2$ and $V = Y^4$.

6.27. Use auxilliary variables to find the pdf of $Z = X_1 X_2 X_3$ where the X_i are independent random variables that are uniformly distributed in $[0, 1]$.

6.28. Let X, Y, and Z be independent zero-mean, unit-variance Gaussian random variables.

(a) Find the pdf of $R = (X^2 + Y^2 + Z^2)^{1/2}$.

(b) Find the pdf of $R^2 = X^2 + Y^2 + Z^2$.

6.29. Let X_1, X_2, X_3, X_4 be processed as follows:

$$Y_1 = X_1, Y_2 = X_1 + X_2, Y_3 = X_2 + X_3, Y_4 = X_3 + X_4.$$

(a) Find an expression for the joint pdf of $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$ in terms of the joint pdf of $\mathbf{X} = (X_1, X_2, X_3, X_4)$.

(b) Find the joint pdf of **Y** if X_1, X_2, X_3, X_4 are independent zero-mean, unit-variance Gaussian random variables.

Section 6.3: Expected Values of Vector Random Variables

6.30. Find $E[M], E[V]$, and $E[MV]$ in Problem 6.23c.

6.31. Compute $E[Z]$ in Problem 6.27 in two ways:

(a) by integrating over $f_Z(z)$;

(b) by integrating over the joint pdf of (X_1, X_2, X_3).

6.32. Find the mean vector and covariance matrix for three multipath signals $\mathbf{X} = (X_1, X_2, X_3)$ in Problem 6.4.

6.33. Find the mean vector and covariance matrix for the samples of the sinusoidal signals $\mathbf{X} = (X(t_1), X(t_2), X(t_3))$ in Problem 6.2.

6.34. **(a)** Find the mean vector and covariance matrix for (X, Y, Z) in Problem 6.5a.

   (b) Repeat part a for Problem 6.5c.

6.35. Find the mean vector and covariance matrix for (X, Y, Z) in Problem 6.7.

6.36. Find the mean vector and covariance matrix for the point (X, Y, Z) inside the unit sphere in Problem 6.8.

6.37. **(a)** Use the results of Problem 6.6c to find the mean vector for the packet arrivals X_1, X_2, and X_3 in Example 6.5.

   (b) Use the results of Problem 6.6b to find the covariance matrix.

   (c) Explain why X_1, X_2, and X_3 are correlated.

6.38. Find the mean vector and covariance matrix for the joint number of packet arrivals in a random time N_1, N_2, and N_3 in Problem 6.17. *Hint:* Use conditional expectation.

6.39. **(a)** Find the mean vector and covariance matrix (U, V, W) in terms of (X_1, X_2, X_3) in Problem 6.22b.

   (b) Find the cross-covariance matrix between (U, V, W) and (X_1, X_2, X_3).

6.40. **(a)** Find the mean vector and covariance matrix of $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$ in terms of those of $\mathbf{X} = (X_1, X_2, X_3, X_4)$ in Problem 6.29.

   (b) Find the cross-covariance matrix between $\mathbf{Y}$ and $\mathbf{X}$.

   (c) Evaluate the mean vector, covariance, and cross-covariance matrices if X_1, X_2, X_3, X_4 are independent random variables.

   (d) Generalize the results in part c to $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n-1}, Y_n)$.

6.41. Let $\mathbf{X} = (X_1, X_2, X_3, X_4)$ consist of equal mean, independent, unit-variance random variables. Find the mean vector, covariance, and cross-covariance matrices of $\mathbf{Y} = \mathbf{A}\mathbf{X}$:

(a) $\mathbf{A} = \begin{bmatrix} 1 & 1/2 & 1/4 & 1/8 \\ 0 & 1 & 1/2 & 1/4 \\ 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(b) $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$.

6.42. Let $W = aX + bY + c$, where X and Y are random variables.

   (a) Find the characteristic function of W in terms of the joint characteristic function of X and Y.

   (b) Find the characteristic function of W if X and Y are the random variables discussed in Example 6.19. Find the pdf of W.

6.43. (a) Find the joint characteristic function of the jointly Gaussian random variables X and Y introduced in Example 5.45. *Hint:* Consider X and Y as a transformation of the independent Gaussian random variables V and W.

  (b) Find $E[X^2Y]$.

  (c) Find the joint characteristic function of $X' = X + a$ and $Y' = Y + b$.

6.44. Let $X = aU + bV$ and $y = cU + dV$, where $|ad - bc| \neq 0$.

  (a) Find the joint characteristic function of X and Y in terms of the joint characteristic function of U and V.

  (b) Find an expression for $E[XY]$ in terms of joint moments of U and V.

6.45. Let X and Y be nonnegative, integer-valued random variables. The joint probability generating function is defined by

$$G_{X,Y}(z_1, z_2) = E[z_1^X z_2^Y] = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} z_1^j z_2^k P[X = j, Y = k].$$

  (a) Find the joint pgf for two independent Poisson random variables with parameters α_1 and α_2.

  (b) Find the joint pgf for two independent binomial random variables with parameters (n, p) and (m, p).

6.46. Suppose that X and Y have joint pgf

$$G_{X,Y}(z_1, z_2) = e^{\alpha_1(z_1-1)+\alpha_2(z_2-1)+\beta(z_1z_2-1)}.$$

  (a) Use the marginal pgf's to show that X and Y are Poisson random variables.

  (b) Find the pgf of $Z = X + Y$. Is Z a Poisson random variable?

6.47. Let X and Y be trinomial random variables with joint pmf

$$P[X = j, Y = k] = \frac{n!\, p_1^j p_2^k (1 - p_1 - p_2)^{n-j-k}}{j!\, k!\, (n - j - k)!} \quad \text{for} \quad 0 \leq j, k \text{ and } j + k \leq n.$$

  (a) Find the joint pgf of X and Y.

  (b) Find the correlation and covariance of X and Y.

6.48. Find the mean vector and covariance matrix for (X, Y) in Problem 6.46.

6.49. Find the mean vector and covariance matrix for (X, Y) in Problem 6.47.

6.50. Let $\mathbf{X} = (X_1, X_2)$ have covariance matrix:

$$\mathbf{K_X} = \begin{bmatrix} 1 & 1/4 \\ 1/4 & 1 \end{bmatrix}.$$

  (a) Find the eigenvalues and eigenvectors of $\mathbf{K_X}$.

  (b) Find the orthogonal matrix $\mathbf{P}$ that diagonalizes $\mathbf{K_X}$. Verify that $\mathbf{P}$ is orthogonal and that $\mathbf{P}^T\mathbf{K_X}\mathbf{P} = \Lambda$.

  (c) Express $\mathbf{X}$ in terms of the eigenvectors of $\mathbf{K_X}$ using the Karhunen-Loeve expansion.

6.51. Repeat Problem 6.50 for $\mathbf{X} = (X_1, X_2, X_3)$ with covariance matrix:

$$\mathbf{K_X} = \begin{bmatrix} 1 & -1/2 & -1/2 \\ -1/2 & 1 & -1/2 \\ -1/2 & -1/2 & 1 \end{bmatrix}.$$

6.52. A square matrix **A** is said to be nonnegative definite if for any vector $\boldsymbol{a} = (a_1, a_2,$ $\ldots, a_n)^{\mathrm{T}}: \boldsymbol{a}^{\mathrm{T}}\mathbf{A}\,\boldsymbol{a} \geq 0$. Show that the covariance matrix is nonnegative definite. *Hint*: Use the fact that $E[(\boldsymbol{a}^{\mathrm{T}}(\mathbf{X} - \mathbf{m_X}))^2] \geq 0$.

6.53. **A** is positive definite if for any nonzero vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}: \boldsymbol{a}^{\mathrm{T}}\mathbf{A}\,\boldsymbol{a} > 0$.

   (a) Show that if all the eigenvalues are positive, then $\mathbf{K_X}$ is positive definite. *Hint*: Let $\mathbf{b} = \mathbf{P}^{\mathrm{T}}\boldsymbol{a}$.

   (b) Show that if $\mathbf{K_X}$ is positive definite, then all the eigenvalues are positive. *Hint*: Let $\boldsymbol{a}$ be an eigenvector of $\mathbf{K_X}$.

Section 6.4: Jointly Gaussian Random Vectors

6.54. Let $\mathbf{X} = (X_1, X_2)$ be the jointly Gaussian random variables with mean vector and covariance matrix given by:

$$\mathbf{m_X} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \mathbf{K_X} = \begin{bmatrix} 3/2 & -1/2 \\ -1/2 & 3/2 \end{bmatrix}.$$

   (a) Find the pdf of **X** in matrix notation.
   (b) Find the pdf of **X** using the quadratic expression in the exponent.
   (c) Find the marginal pdfs of X_1 and X_2.
   (d) Find a transformation A such that the vector $\mathbf{Y} = A\mathbf{X}$ consists of independent Gaussian random variables.
   (e) Find the joint pdf of **Y**.

6.55. Let $\mathbf{X} = (X_1, X_2, X_3)$ be the jointly Gaussian random variables with mean vector and covariance matrix given by:

$$\mathbf{m_X} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \qquad \mathbf{K_X} = \begin{bmatrix} 3/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 3/2 \end{bmatrix}.$$

   (a) Find the pdf of **X** in matrix notation.
   (b) Find the pdf of **X** using the quadratic expression in the exponent.
   (c) Find the marginal pdfs of X_1, X_2, and X_3.
   (d) Find a transformation A such that the vector $\mathbf{Y} = A\mathbf{X}$ consists of independent Gaussian random variables.
   (e) Find the joint pdf of **Y**.

6.56. Let U_1, U_2, and U_3 be independent zero-mean, unit-variance Gaussian random variables and let $X = U_1, Y = U_1 + U_2$, and $Z = U_1 + U_2 + U_3$.

   (a) Find the covariance matrix of (X, Y, Z).
   (b) Find the joint pdf of (X, Y, Z).
   (c) Find the conditional pdf of Y and Z given X.
   (d) Find the conditional pdf of Z given X and Y.

6.57. Let X_1, X_2, X_3, X_4 be independent zero-mean, unit-variance Gaussian random variables that are processed as follows:

$$Y_1 = X_1 + X_2, Y_2 = X_2 + X_3, Y_3 = X_3 + X_4.$$

   (a) Find the covariance matrix of $\mathbf{Y} = (Y_1, Y_2, Y_3)$.
   (b) Find the joint pdf of **Y**.
   (c) Find the joint pdf of Y_1 and Y_2; Y_1 and Y_3.
   (d) Find a transformation A such that the vector $\mathbf{Z} = A\mathbf{Y}$ consists of independent Gaussian random variables.

6.58. A more realistic model of the receiver in the multiuser communication system in Problem 6.21 has the K received signals $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_K)$ given by:

$$\mathbf{Y} = \mathbf{ARb} + \mathbf{N}$$

where $\mathbf{A} = [\alpha_k]$ is a diagonal matrix of positive channel gains, $\mathbf{R}$ is a symmetric matrix that accounts for the interference between users, and $\mathbf{b} = (b_1, b_2, \ldots, b_K)$ is the vector of bits from each of the transmitters. $\mathbf{N}$ is the vector of K independent zero-mean, unit-variance Gaussian noise random variables.

    (a) Find the joint pdf of $\mathbf{Y}$.

    (b) Suppose that in order to recover $\mathbf{b}$, the receiver computes $\mathbf{Z} = (\mathbf{AR})^{-1}\mathbf{Y}$. Find the joint pdf of $\mathbf{Z}$.

6.59. **(a)** Let $\mathbf{K_3}$ be the covariance matrix in Problem 6.55. Find the corresponding $\mathbf{Q_2}$ and $\mathbf{Q_3}$ in Example 6.23.

    (b) Find the conditional pdf of X_3 given X_1 and X_2.

6.60. In Example 6.23, show that:

$$\tfrac{1}{2}(\mathbf{x}_n - \mathbf{m}_n)^{\mathrm{T}}\mathbf{Q}_n(\mathbf{x}_n - \mathbf{m}_n) - \tfrac{1}{2}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})^{\mathrm{T}}\mathbf{Q}_{n-1}(\mathbf{x}_{n-1} - \mathbf{m}_{n-1})$$
$$= Q_{nn}\{(x_n - m_n) + B\}^2 - Q_{nn}B^2$$

$$\text{where } B = \frac{1}{Q_{nn}}\sum_{j=1}^{n-1}Q_{jk}(x_j - m_j) \quad \text{and} \quad |\mathbf{K}_n|/|\mathbf{K}_{n-1}| = Q_{nn}.$$

6.61. Find the pdf of the sum of Gaussian random variables in the following cases:

    (a) $Z = X_1 + X_2 + X_3$ in Problem 6.55.

    (b) $Z = X + Y + Z$ in Problem 6.56.

    (c) $Z = Y_1 + Y_2 + Y_3$ in Problem 6.57.

6.62. Find the joint characteristic function of the jointly Gaussian random vector $\mathbf{X}$ in Problem 6.54.

6.63. Suppose that a jointly Gaussian random vector $\mathbf{X}$ has zero mean vector and the covariance matrix given in Problem 6.51.

    (a) Find the joint characteristic function.

    (b) Can you obtain an expression for the joint pdf? Explain your answer.

6.64. Let X and Y be jointly Gaussian random variables. Derive the joint characteristic function for X and Y using conditional expectation.

6.65. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be jointly Gaussian random variables. Derive the characteristic function for $\mathbf{X}$ by carrying out the integral in Eq. (6.32). *Hint*: You will need to complete the square as follows:

$$(\mathbf{x} - j\mathbf{K}\omega)^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{x} - j\mathbf{K}\omega) = \mathbf{x}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{x} - 2j\mathbf{x}^{\mathrm{T}}\omega + j^2\omega^{\mathrm{T}}\mathbf{K}\omega.$$

6.66. Find $E[X^2Y^2]$ for jointly Gaussian random variables from the characteristic function.

6.67. Let $\mathbf{X} = (X_1, X_2, X_3, X_4)$ be zero-mean jointly Gaussian random variables. Show that $E[X_1X_2X_3X_4] = E[X_1X_2]E[X_3X_4] + E[X_1X_3]E[X_2X_4] + E[X_1X_4]E[X_2X_3]$.

Section 6.5: Mean Square Estimation

6.68. Let X and Y be discrete random variables with three possible joint pmf's:

	(i)				(ii)				(iii)		
X/Y	−1	0	1	X/Y	−1	0	1	X/Y	−1	0	1
−1	1/6	1/6	0	−1	1/9	1/9	1/9	−1	1/3	0	0
0	0	0	1/3	0	1/9	1/9	1/9	0	0	1/3	0
1	1/6	1/6	0	1	1/9	1/9	1/9	1	0	0	1/3

    (a) Find the minimum mean square error linear estimator for Y given X.

    (b) Find the minimum mean square error estimator for Y given X.

    (c) Find the MAP and ML estimators for Y given X.

    (d) Compare the mean square error of the estimators in parts a, b, and c.

6.69. Repeat Problem 6.68 for the continuous random variables X and Y in Problem 5.26.

6.70. Find the ML estimator for the signal s in Problem 6.4.

6.71. Let N_1 be the number of Web page requests arriving at a server in the period $(0, 100)$ ms and let N_2 be the *total* combined number of Web page requests arriving at a server in the period $(0, 200)$ ms. Assume page requests occur every 1-ms interval according to independent Bernoulli trials with probability of success p.

    (a) Find the minimum linear mean square estimator for N_2 given N_1 and the associated mean square error.

    (b) Find the minimum mean square error estimator for N_2 given N_1 and the associated mean square error.

    (c) Find the maximum a posteriori estimator for N_2 given N_1.

    (d) Repeat parts a, b, and c for the estimation of N_1 given N_2.

6.72. Let $Y = X + N$ where X and N are independent Gaussian random variables with different variances and N is zero mean.

    (a) Plot the correlation coefficient between the "observed signal" Y and the "desired signal" X as a function of the signal-to-noise ratio σ_X/σ_N.

    (b) Find the minimum mean square error estimator for X given Y.

    (c) Find the MAP and ML estimators for X given Y.

    (d) Compare the mean square error of the estimators in parts a, b and c.

6.73. Let X, Y, Z be the random variables in Problem 6.7.

    (a) Find the minimum mean square error linear estimator for Y given X and Z.

    (b) Find the minimum mean square error estimator for Y given X and Z.

    (c) Find the MAP and ML estimators for Y given X and Z.

    (d) Compare the mean square error of the estimators in parts b and c.

6.74. **(a)** Repeat Problem 6.73 for the estimator of X_2, given X_1 and X_3 in Problem 6.13.

    (b) Repeat Problem 6.73 for the estimator of X_3 given X_1 and X_2.

6.75. Consider the ideal multiuser communication system in Problem 6.21. Assume the transmitted bits b_k are independent and equally likely to be $+1$ or -1.

    (a) Find the ML and MAP estimators for **b** given the observation **Y**.

    (b) Find the minimum mean square linear estimator for **b** given the observation **Y**. How can this estimator be used in deciding what were the transmitted bits?

6.76. Repeat Problem 6.75 for the multiuser system in Problem 6.58.

6.77. A second-order predictor for samples of an image predicts the sample E as a linear function of sample D to its left and sample B in the previous line, as shown below:

$$\begin{array}{lcccc}
\text{line } j & \dots & A & B & C \dots \\
\text{line } j + 1 & \dots & D & E & \dots
\end{array}$$

$$\text{Estimate for } E = aD + bB.$$

    (a) Find a and b if all samples have variance σ^2 and if the correlation coefficient between D and E is ρ, between B and E is ρ, and between D and B is ρ^2.

    (b) Find the mean square error of the predictor found in part a, and determine the reduction in the variance of the signal in going from the input to the output of the predictor.

6.78. Show that the mean square error of the two-tap linear predictor is given by Eq. (6.64).

6.79. In "hexagonal sampling" of an image, the samples in consecutive lines are offset relative to each other as shown below:

$$\begin{array}{ccccc} \text{line } j & \ldots & & A & B \\ \text{line } j+1 & \ldots & C & D & \end{array}$$

The covariance between two samples a and b is given by $\rho^{d(a,b)}$ where $d(a, b)$ is the Euclidean distance between the points. In the above samples, the distance between A and B, A and C, A and D, C and D, and B and D is 1. Suppose we wish to use a two-tap linear predictor to predict the sample D. Which two samples from the set $\{A, B, C\}$ should we use in the predictor? What is the resulting mean square error?

*Section 6.6: Generating Correlated Vector Random Variables

6.80. Find a linear transformation that diagonalizes $\mathbf{K}$.

(a) $\mathbf{K} = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$.

(b) $\mathbf{K} = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$.

6.81. Generate and plot the scattergram of 1000 pairs of random variables $\mathbf{Y}$ with the covariance matrices in Problem 6.80 if:

(a) X_1 and X_2 are independent random variables that are each uniform in the unit interval;

(b) X_1 and X_2 are independent zero-mean, unit-variance Gaussian random variables.

6.82. Let $\mathbf{X} = (X_1, X_2, X_3)$ be the jointly Gaussian random variables in Problem 6.55.

(a) Find a linear transformation that diagonalizes the covariance matrix.

(b) Generate 1000 triplets of $\mathbf{Y} = \mathbf{AX}$ and plot the scattergrams for Y_1 and Y_2, Y_1 and Y_3, and Y_2 and Y_3. Confirm that the scattergrams are what is expected.

6.83. Let $\mathbf{X}$ be a jointly Gaussian random vector with mean $\mathbf{m_X}$ and covariance matrix $\mathbf{K_X}$ and let $\mathbf{A}$ be a matrix that diagonalizes $\mathbf{K_X}$. What is the joint pdf of $\mathbf{A}^{-1}(\mathbf{X} - \mathbf{m_X})$?

6.84. Let $X_1, X_2, \ldots, X_n$ be independent zero-mean, unit-variance Gaussian random variables. Let $Y_k = (X_k + X_{k-1})/2$, that is, Y_k is the moving average of pairs of values of X. Assume $X_{-1} = 0 = X_{n+1}$.

(a) Find the covariance matrix of the Y_k's.

(b) Use Octave to generate a sequence of 1000 samples $Y_1, \ldots, Y_n$. How would you check whether the Y_k's have the correct covariances?

6.85. Repeat Problem 6.84 with $Y_k = X_k - X_{k-1}$.

6.86. Let $\mathbf{U}$ be an orthogonal matrix. Show that if $\mathbf{A}$ diagonalizes the covariance matrix $\mathbf{K}$, then $\mathbf{B} = \mathbf{UA}$ also diagonalizes $\mathbf{K}$.

6.87. The transformation in Problem 6.56 is said to be "causal" because each output depends only on "past" inputs.

(a) Find the covariance matrix of X, Y, Z in Problem 6.56.

(b) Find a noncausal transformation that diagonalizes the covariance matrix in part a.

6.88. (a) Find a causal transformation that diagonalizes the covariance matrix in Problem 6.54.

(b) Repeat for the covariance matrix in Problem 6.55.

Problems Requiring Cumulative Knowledge

6.89. Let $U_0, U_1, \ldots$ be a sequence of independent zero-mean, unit-variance Gaussian random variables. A "low-pass filter" takes the sequence U_i and produces the output sequence $X_n = (U_n + U_{n-1})/2$, and a "high-pass filter" produces the output sequence $Y_n = (U_n - U_{n-1})/2$.

  (a) Find the joint pdf of X_{n+1}, X_n, and X_{n-1}; of X_n, X_{n+m}, and X_{n+2m}, $m > 1$.

  (b) Repeat part a for Y_n.

  (c) Find the joint pdf of X_n, X_m, Y_n, and Y_m.

  (d) Find the corresponding joint characteristic functions in parts a, b, and c.

6.90. Let $X_1, X_2, \ldots, X_n$ be the samples of a speech waveform in Example 6.31. Suppose we want to interpolate for the value of a sample in terms of the previous and the next samples, that is, we wish to find the best linear estimate for X_2 in terms of X_1 and X_3.

  (a) Find the coefficients of the best linear estimator (interpolator).

  (b) Find the mean square error of the best linear interpolator and compare it to the mean square error of the two-tap predictor in Example 6.31.

  (c) Suppose that the samples are jointly Gaussian. Find the pdf of the interpolation error.

6.91. Let $X_1, X_2, \ldots, X_n$ be samples from some signal. Suppose that the samples are jointly Gaussian random variables with covariance

$$\text{COV}(X_i, X_j) = \begin{cases} \sigma^2 & \text{for } i = j \\ \rho\sigma^2 & \text{for } |i - j| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose we take blocks of two consecutive samples to form a vector $\mathbf{X}$, which is then linearly transformed to form $\mathbf{Y} = \mathbf{AX}$.

  (a) Find the matrix $\mathbf{A}$ so that the components of $\mathbf{Y}$ are independent random variables.

  (b) Let $\mathbf{X}_i$ and $\mathbf{X}_{i+1}$ be two consecutive blocks and let $\mathbf{Y}_i$ and $\mathbf{Y}_{i+1}$ be the corresponding transformed variables. Are the components of $\mathbf{Y}_i$ and $\mathbf{Y}_{i+1}$ independent?

6.92. A multiplexer combines N digital television signals into a common communications line. TV signal n generates X_n bits every 33 milliseconds, where X_n is a Gaussian random variable with mean m and variance σ^2. Suppose that the multiplexer accepts a maximum total of T bits from the combined sources every 33 ms, and that any bits in excess of T are discarded. Assume that the N signals are independent.

  (a) Find the probability that bits are discarded in a given 33-ms period, if we let $T = m_a + t\sigma$, where m_a is the mean total bits generated by the combined sources, and σ is the standard deviation of the total number of bits produced by the combined sources.

  (b) Find the average number of bits discarded per period.

  (c) Find the long-term fraction of bits lost by the multiplexer.

  (d) Find the average number of bits per source allocated in part a, and find the average number of bits lost per source. What happens as N becomes large?

  (e) Suppose we require that t be adjusted with N so that the fraction of bits lost per source is kept constant. Find an equation whose solution yields the desired value of t.

  (f) Do the above results change if the signals have pairwise covariance ρ?

6.93. Consider the estimation of T given N_1 and arrivals in Problem 6.17.

  (a) Find the ML and MAP estimators for T.

  (b) Find the linear mean square estimator for T.

  (c) Repeat parts a and b if N_1 and N_2 are given.

Random Processes

In certain random experiments, the outcome is a function of time or space. For example, in speech recognition systems, decisions are made on the basis of a voltage waveform corresponding to a speech utterance. In an image processing system, the intensity and color of the image varies over a rectangular region. In a peer-to-peer network, the number of peers in the system varies with time. In some situations, two or more functions of time may be of interest. For example, the temperature in a certain city and the demand placed on the local electric power utility vary together in time.

The random time functions in the above examples can be viewed as numerical quantities that evolve randomly in time or space. Thus what we really have is a family of random variables indexed by the time or space variable. In this chapter we begin the study of random processes. We will proceed as follows:

- In Section 9.1 we introduce the notion of a *random process* (or *stochastic process*), which is defined as an *indexed family of random variables*.
- We are interested in specifying the joint behavior of the random variables within a family (i.e., the temperature at two time instants). In Section 9.2 we see that this is done by specifying joint distribution functions, as well as mean and covariance functions.
- In Sections 9.3 to 9.5 we present examples of stochastic processes and show how models of complex processes can be developed from a few simple models.
- In Section 9.6 we introduce the class of stationary random processes that can be viewed as random processes in "steady state."
- In Section 9.7 we investigate the continuity properties of random processes and define their derivatives and integrals.
- In Section 9.8 we examine the properties of time averages of random processes and the problem of estimating the parameters of a random process.
- In Section 9.9 we describe methods for representing random processes by Fourier series and by the Karhunen-Loeve expansion.
- Finally, in Section 9.10 we present methods for generating random processes.

9.1   DEFINITION OF A RANDOM PROCESS

Consider a random experiment specified by the outcomes ζ from some sample space S, by the events defined on S, and by the probabilities on these events. Suppose that to every outcome $\zeta \in S$, we assign a function of time according to some rule:

$$X(t, \zeta) \qquad t \in I.$$

The graph of the function $X(t, \zeta)$ versus t, for ζ fixed, is called a **realization**, **sample path**, or **sample function** of the random process. Thus we can view the outcome of the random experiment as producing an entire function of time as shown in Fig. 9.1. On the other hand, if we fix a time t_k from the index set I, then $X(t_k, \zeta)$ is a random variable (see Fig. 9.1) since we are mapping ζ onto a real number. Thus we have created a family (or ensemble) of random variables indexed by the parameter t, $\{X(t, \zeta), t \in I\}$. This family is called a **random process**. We also refer to random processes as **stochastic processes**. We usually suppress the ζ and use $X(t)$ to denote a random process.

A stochastic process is said to be **discrete-time** if the index set I is a countable set (i.e., the set of integers or the set of nonnegative integers). When dealing with discrete-time processes, we usually use n to denote the time index and X_n to denote the random process. A **continuous-time** stochastic process is one in which I is continuous (i.e., the real line or the nonnegative real line).

The following example shows how we can imagine a stochastic process as resulting from nature selecting ζ at the beginning of time and gradually revealing it in time through $X(t, \zeta)$.
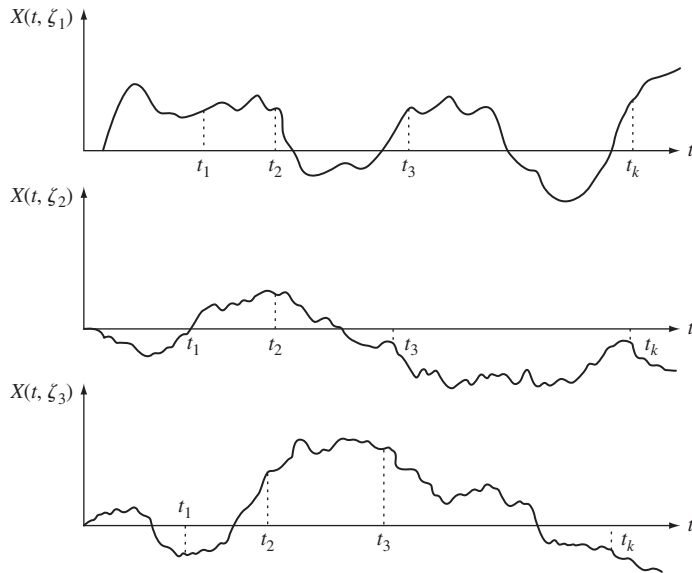

FIGURE 9.1
Several realizations of a random process.

Example 9.1   Random Binary Sequence

Let ζ be a number selected at random from the interval $S = [0, 1]$, and let $b_1 b_2 \ldots$ be the binary expansion of ζ:

$$\zeta = \sum_{i=1}^{\infty} b_i 2^{-i} \qquad \text{where } b_i \in \{0, 1\}.$$

Define the discrete-time random process $X(n, \zeta)$ by

$$X(n, \zeta) = b_n \qquad n = 1, 2, \ldots.$$

The resulting process is sequence of binary numbers, with $X(n, \zeta)$ equal to the nth number in the binary expansion of ζ.

Example 9.2   Random Sinusoids

Let ζ be selected at random from the interval $[-1, 1]$. Define the continuous-time random process $X(t, \zeta)$ by

$$X(t, \zeta) = \zeta \cos(2\pi t) \qquad -\infty < t < \infty.$$

The realizations of this random process are sinusoids with amplitude ζ, as shown in Fig. 9.2(a).

Let ζ be selected at random from the interval $(-\pi, \pi)$ and let $Y(t, \zeta) = \cos(2\pi t + \zeta)$. The realizations of $Y(t, \zeta)$ are phase-shifted versions of $\cos 2\pi t$ as shown in Fig 9.2(b).
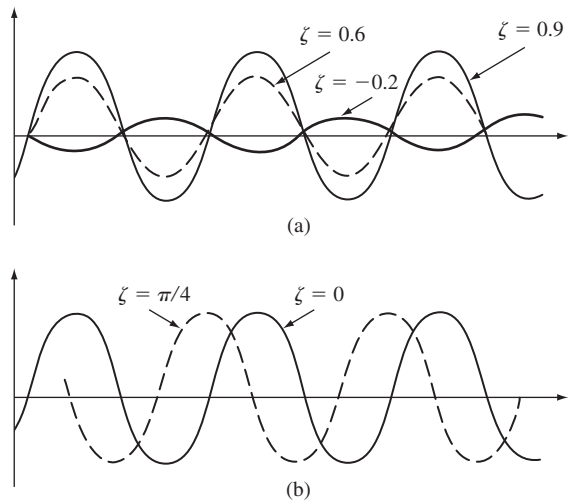
FIGURE 9.2
(a) Sinusoid with random amplitude, (b) Sinusoid with random phase.

The randomness in ζ induces randomness in the observed function $X(t, \zeta)$. In principle, one can deduce the probability of events involving a stochastic process at various instants of time from probabilities involving ζ by using the equivalent-event method introduced in Chapter 4.

Example 9.3

Find the following probabilities for the random process introduced in Example 9.1: $P[X(1, \zeta) = 0]$ and $P[X(1, \zeta) = 0$ and $X(2, \zeta) = 1]$.

The probabilities are obtained by finding the equivalent events in terms of ζ:

$$P[X(1, \zeta) = 0] = P\left[0 \le \zeta < \frac{1}{2}\right] = \frac{1}{2}$$

$$P[X(1, \zeta) = 0 \text{ and } X(2, \zeta) = 1] = P\left[\frac{1}{4} \le \zeta < \frac{1}{2}\right] = \frac{1}{4},$$

since all points in the interval $[0 \le \zeta \le 1]$ begin with $b_1 = 0$ and all points in $[1/4, 1/2]$ begin with $b_1 = 0$ and $b_2 = 1$. Clearly, any sequence of k bits has a corresponding subinterval of length (and hence probability) 2^{-k}.

Example 9.4

Find the pdf of $X_0 = X(t_0, \zeta)$ and $Y(t_0, \zeta)$ in Example 9.2.

If t_0 is such that $\cos(2\pi t_0) = 0$, then $X(t_0, \zeta) = 0$ for all ζ and the pdf of $X(t_0)$ is a delta function of unit weight at $x = 0$. Otherwise, $X(t_0, \zeta)$ is uniformly distributed in the interval $(-\cos 2\pi t_0, \cos 2\pi t_0)$ since ζ is uniformly distributed in $[-1, 1]$ (see Fig. 9.3a). Note that the pdf of $X(t_0, \zeta)$ depends on t_0.

The approach used in Example 4.36 can be used to show that $Y(t_0, \zeta)$ has an arcsine distribution:

$$f_Y(y) = \frac{1}{\pi\sqrt{1 - y^2}}, \qquad |y| < 1$$

(see Fig. 9.3b). Note that the pdf of $Y(t_0, \zeta)$ does not depend on t_0.

Figure 9.3(c) shows a histogram of 1000 samples of the amplitudes $X(t_0, \zeta)$ at $t_0 = 0$, which can be seen to be approximately uniformly distributed in $[-1, 1]$. Figure 9.3(d) shows the histogram for the samples of the sinusoid with random phase. Clearly there is agreement with the arcsine pdf.

In general, the sample paths of a stochastic process can be quite complicated and cannot be described by simple formulas. In addition, it is usually not possible to identify an underlying probability space for the family of observed functions of time. Thus the equivalent-event approach for computing the probability of events involving $X(t, \zeta)$ in terms of the probabilities of events involving ζ does not prove useful in
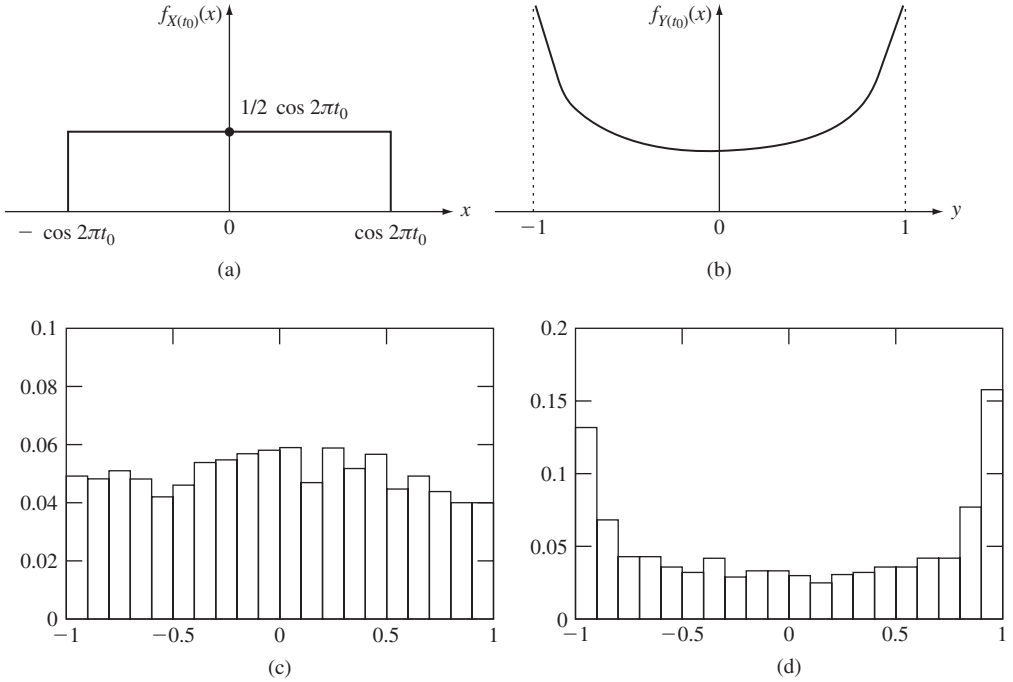
FIGURE 9.3
(a) pdf of sinusoid with random amplitude. (b) pdf of sinusoid with random phase. (c) Histogram of samples from uniform amplitude sinusoid at $t = 0$. (d) Histogram of samples from random phase sinusoid at $t = 0$.

practice. In the next section we show an alternative method for specifying the probabilities of events involving a stochastic process.

9.2   SPECIFYING A RANDOM PROCESS

There are many questions regarding random processes that cannot be answered with just knowledge of the distribution at a single time instant. For example, we may be interested in the temperature at a given locale at two different times. This requires the following information:

$$P[x_1 < X(t_1) \le x_1, x_2 < X(t_2) \le x_2].$$

In another example, the speech compression system in a cellular phone predicts the value of the speech signal at the next sampling time based on the previous k samples. Thus we may be interested in the following probability:

$$P[a < X(t_{k+1}) \le b \,|\, X(t_1) = x_1, X(t_2) = x_2, \ldots, X(t_k) = x_k].$$

It is clear that a general description of a random process should provide probabilities for *vectors* of samples of the process.

9.2.1   Joint Distributions of Time Samples

Let $X_1, X_2, \ldots, X_k$ be the k random variables obtained by sampling the random process $X(t, \zeta)$ at the times $t_1, t_2, \ldots, t_k$:

$$X_1 = X(t_1, \zeta), X_2 = X(t_2, \zeta,), \ldots, X_k = X(t_k, \zeta),$$

as shown in Fig. 9.1. The joint behavior of the random process at these k time instants is specified by the joint cumulative distribution of the *vector* random variable $X_1, X_2, \ldots, X_k$. The probabilities of any event involving the random process at all or some of these time instants can be computed from this cdf using the methods developed for vector random variables in Chapter 6. Thus, *a stochastic process is specified by the collection of kth-order joint cumulative distribution functions*:

$$F_{X_1, \ldots, X_k}(x_1, x_2, \ldots, x_k) = P[X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_k) \leq x_k], \quad (9.1)$$

for any k and any choice of sampling instants $t_1, \ldots, t_k$. Note that the collection of cdf's must be consistent in the sense that lower-order cdf's are obtained as marginals of higher-order cdf's. If the stochastic process is *continuous-valued*, then a collection of probability density functions can be used instead:

$$f_{X_1, \ldots, X_k}(x_1, x_2, \ldots, x_k) \, dx_1 \ldots dx_n$$
$$= P\{x_1 < X(t_1) \leq x_1 + dx_1, \ldots, x_k < X(t_k) \leq x_k + dx_k\}. \quad (9.2)$$

If the stochastic process is *discrete-valued*, then a collection of probability mass functions can be used to specify the stochastic process:

$$p_{X_1, \ldots, X_k}(x_1, x_2, \ldots, x_k) = P[X(t_1) = x_1, X(t_2) = x_2, \ldots, X(t_k) = x_k] \quad (9.3)$$

for any k and any choice of sampling instants $n_1, \ldots, n_k$.

At first glance it does not appear that we have made much progress in specifying random processes because we are now confronted with the task of specifying a vast collection of joint cdf's! However, this approach works because most useful models of stochastic processes are obtained by elaborating on a few simple models, so the methods developed in Chapters 5 and 6 of this book can be used to derive the required cdf's. The following examples give a preview of how we construct complex models from simple models. We develop these important examples more fully in Sections 9.3 to 9.5.

Example 9.5   iid Bernoulli Random Variables

Let X_n be a sequence of independent, identically distributed Bernoulli random variables with $p = 1/2$. The joint pmf for any k time samples is then

$$P[X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k] = P[X_1 = x_1] \ldots P[X_k = x_k] = \left(\frac{1}{2}\right)^k$$

where $x_i \in \{0, 1\}$ for all i. This binary random process is equivalent to the one discussed in Example 9.1.

Example 9.6   iid Gaussian Random Variables

Let X_n be a sequence of independent, identically distributed Gaussian random variables with zero mean and variance σ_X^2. The joint pdf for any k time samples is then

$$f_{X_1, X_2, \ldots, X_k}(x_1, x_2, \ldots, x_k) = \frac{1}{(2\pi\sigma^2)^{k/2}} e^{-(x_1^2 + x_2^2 + \cdots + x_k^2)/2\sigma^2}.$$

The following two examples show how more complex and interesting processes can be built from iid sequences.

Example 9.7   Binomial Counting Process

Let X_n be a sequence of independent, identically distributed Bernoulli random variables with $p = 1/2$. Let S_n be the number of 1's in the first n trials:

$$S_n = X_1 + X_2 + \cdots + X_n \quad \text{for} \quad n = 0, 1, \ldots.$$

S_n is an integer-valued nondecreasing function of n that grows by unit steps after a random number of time instants. From previous chapters we know that S_n is a binomial random variable with parameters n and $p = 1/2$. In the next section we show how to find the joint pmf's of S_n using conditional probabilities.

Example 9.8   Filtered Noisy Signal

Let X_j be a sequence of independent, identically distributed observations of a signal voltage μ corrupted by zero-mean Gaussian noise N_j with variance σ^2:

$$X_j = \mu + N_j \quad \text{for} \quad j = 0, 1, \ldots.$$

Consider the signal that results from averaging the sequence of observations:

$$S_n = (X_1 + X_2 + \cdots + X_n)/n \quad \text{for} \quad n = 0, 1, \ldots.$$

From previous chapters we know that S_n is the sample mean of an iid sequence of Gaussian random variables. We know that S_n itself is a Gaussian random variable with mean μ and variance σ^2/n, and so it tends towards the value μ as n increases. In a later section, we show that S_n is an example from the class of Gaussian random processes.

9.2.2   The Mean, Autocorrelation, and Autocovariance Functions

The moments of time samples of a random process can be used to partially specify the random process because they summarize the information contained in the joint cdf's.

The **mean function $m_X(t)$** and the **variance function VAR$[X(t)]$** of a continuous-time random process $X(t)$ are defined by

$$m_X(t) = E[X(t)] = \int_{-\infty}^{\infty} x f_{X(t)}(x)\, dx, \qquad (9.4)$$

and

$$\text{VAR}[X(t)] = \int_{-\infty}^{\infty} (x - m_X(t))^2\, f_{X(t)}(x)\, dx, \qquad (9.5)$$

where $f_{X(t)}(x)$ is the pdf of $X(t)$. Note that $m_X(t)$ and VAR$[X(t)]$ are *deterministic* functions of time. Trends in the behavior of $X(t)$ are reflected in the variation of $m_X(t)$ with time. The variance gives an indication of the spread in the values taken on by $X(t)$ at different time instants.

The **autocorrelation $R_X(t_1, t_2)$** of a random process $X(t)$ is defined as the joint moment of $X(t_1)$ and $X(t_2)$:

$$R_X(t_1, t_2) = E[X(t_1) X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X(t_1), X(t_2)}(x, y)\, dx\, dy, \qquad (9.6)$$

where $f_{X(t_1), X(t_2)}(x, y)$ is the second-order pdf of $X(t)$. In general, the autocorrelation is a function of t_1 and t_2. Note that $R_X(t, t) = E[X^2(t)]$.

The **autocovariance $C_X(t_1, t_2)$** of a random process $X(t)$ is defined as the covariance of $X(t_1)$ and $X(t_2)$:

$$C_X(t_1, t_2) = E[\{X(t_1) - m_X(t_1)\}\{X(t_2) - m_X(t_2)\}]. \qquad (9.7)$$

From Eq. (5.30), the autocovariance can be expressed in terms of the autocorrelation and the means:

$$C_X(t_1, t_2) = R_X(t_1, t_2) - m_X(t_1) m_X(t_2). \qquad (9.8)$$

Note that the variance of $X(t)$ can be obtained from $C_X(t_1, t_2)$:

$$\text{VAR}[X(t)] = E[(X(t) - m_X(t))^2] = C_X(t, t). \qquad (9.9)$$

The *correlation coefficient* of $X(t)$ is defined as the correlation coefficient of $X(t_1)$ and $X(t_2)$ (see Eq. 5.31):

$$\rho_X(t_1, t_2) = \frac{C_X(t_1, t_2)}{\sqrt{C_X(t_1, t_1)}\sqrt{C_X(t_2, t_2)}}. \qquad (9.10)$$

From Eq. (5.32) we have that $|\rho_X(t_1, t_2)| \le 1$. Recall that the correlation coefficient is a measure of the extent to which a random variable can be predicted as a linear function of another. In Chapter 10, we will see that the autocovariance function and the autocorrelation function play a critical role in the design of linear methods for analyzing and processing random signals.

The mean, variance, autocorrelation, and autocovariance functions for discrete-time random processes are defined in the same manner as above. We use a slightly different notation for the time index. The **mean and variance** of a *discrete-time* random process X_n are defined as:

$$m_X(n) = E[X_n] \quad \text{and} \quad \text{VAR}[X_n] = E[(X_n - m_X(n))^2]. \tag{9.11}$$

The **autocorrelation and autocovariance functions** of a discrete-time random process X_n are defined as follows:

$$R_X(n_1, n_2) = E[X(n_1)X(n_2)] \tag{9.12}$$

and

$$
\begin{aligned}
C_X(n_1, n_2) &= E[\{X(n_1) - m_X(n_1)\}\{X(n_2) - m_X(n_2)\}] \\
&= R_X(n_1, n_2) - m_X(n_1)m_X(n_2). \tag{9.13}
\end{aligned}
$$

Before proceeding to examples, we reiterate that the mean, autocorrelation, and autocovariance functions are only partial descriptions of a random process. Thus we will see later in the chapter that it is possible for two quite different random processes to have the same mean, autocorrelation, and autocovariance functions.

Example 9.9   Sinusoid with Random Amplitude

Let $X(t) = A \cos 2\pi t$, where A is some random variable (see Fig. 9.2a). The mean of $X(t)$ is found using Eq. (4.30):

$$m_X(t) = E[A \cos 2\pi t] = E[A] \cos 2\pi t.$$

Note that the mean varies with t. In particular, note that the process is always zero for values of t where $\cos 2\pi t = 0$.

The autocorrelation is

$$
\begin{aligned}
R_X(t_1, t_2) &= E[A \cos 2\pi t_1 \, A \cos 2\pi t_2] \\
&= E[A^2] \cos 2\pi t_1 \cos 2\pi t_2,
\end{aligned}
$$

and the autocovariance is then

$$
\begin{aligned}
C_X(t_1, t_2) &= R_X(t_1, t_2) - m_X(t_1)m_X(t_2) \\
&= \{E[A^2] - E[A]^2\} \cos 2\pi t_1 \cos 2\pi t_2 \\
&= \text{VAR}[A] \cos 2\pi t_1 \cos 2\pi t_2.
\end{aligned}
$$

Example 9.10   Sinusoid with Random Phase

Let $X(t) = \cos(\omega t + \Theta)$, where Θ is uniformly distributed in the interval $(-\pi, \pi)$ (see Fig. 9.2b). The mean of $X(t)$ is found using Eq. (4.30):

$$m_X(t) = E[\cos(\omega t + \Theta)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(\omega t + \theta) \, d\theta = 0.$$

The autocorrelation and autocovariance are then

$$C_X(t_1, t_2) = R_X(t_1, t_2) = E[\cos(\omega t_1 + \Theta) \cos(\omega t_2 + \Theta)]$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \{\cos(\omega(t_1 - t_2)) + \cos(\omega(t_1 + t_2) + 2\theta)\} \, d\theta$$

$$= \frac{1}{2} \cos(\omega(t_1 - t_2)),$$

where we used the identity $\cos(a) \cos(b) = 1/2 \cos(a + b) + 1/2 \cos(a - b)$. Note that $m_X(t)$ is a constant and that $C_X(t_1, t_2)$ depends only on $|t_1 - t_2|$. Note as well that the samples at time t_1 and t_2 are uncorrelated if $\omega(t_1 - t_2) = k\pi$ where k is any integer.

9.2.3   Multiple Random Processes

In most situations we deal with more than one random process at a time. For example, we may be interested in the temperatures at city a, $X(t)$, and city b, $Y(t)$. Another very common example involves a random process $X(t)$ that is the "input" to a system and another random process $Y(t)$ that is the "output" of the system. Naturally, we are interested in the interplay between $X(t)$ and $Y(t)$.

The joint behavior of two or more random processes is specified by the collection of joint distributions for all possible choices of time samples of the processes. Thus for a pair of continuous-valued random processes $X(t)$ and $Y(t)$ we must specify all possible joint density functions of $X(t_1), \ldots, X(t_k)$ and $Y(t_1'), \ldots, Y(t_j')$ for all k, j, and all choices of $t_1, \ldots, t_k$ and $t_1', \ldots, t_j'$. For example, the simplest joint pdf would be:

$$f_{X(t_1), Y(t_2)}(x, y) \, dx \, dy = P\{x < X(t_1) \le x + dx, y < Y(t_2) \le y + dy\}.$$

Note that the time indices of $X(t)$ and $Y(t)$ need not be the same. For example, we may be interested in the input at time t_1 and the output at a later time t_2.

The random processes $X(t)$ and $Y(t)$ are said to be **independent random processes** if the vector random variables $\mathbf{X} = (X(t_1), \ldots, X(t_k))$ and $\mathbf{Y} = (Y(t_1'), \ldots, Y(t_j'))$ are independent for *all* k, j, and *all* choices of $t_1, \ldots, t_k$ and $t_1', \ldots, t_j'$:

$$F_{\mathbf{X}, \mathbf{Y}}(x_1, \ldots, x_k, y_1, \ldots, y_j) = F_{\mathbf{X}}(X_1, \ldots, X_k) F_{\mathbf{Y}}(y_1, \ldots, y_j).$$

The **cross-correlation** $R_{X,Y}(t_1, t_2)$ of $X(t)$ and $Y(t)$ is defined by

$$R_{X,Y}(t_1, t_2) = E[X(t_1) Y(t_2)]. \tag{9.14}$$

The processes $X(t)$ and $Y(t)$ are said to be **orthogonal random processes** if

$$R_{X,Y}(t_1, t_2) = 0 \quad \text{for all } t_1 \text{ and } t_2. \tag{9.15}$$

The **cross-covariance $C_{X,Y}(t_1, t_2)$** of $X(t)$ and $Y(t)$ is defined by

$$C_{X,Y}(t_1, t_2) = E[\{X(t_1) - m_X(t_1)\}\{Y(t_2) - m_X(t_2)\}]$$
$$= R_{X,Y}(t_1, t_2) - m_X(t_1)m_X(t_2). \qquad (9.16)$$

The processes $X(t)$ and $Y(t)$ are said to be **uncorrelated random processes** if

$$C_{X,Y}(t_1, t_2) = 0 \quad \text{for all } t_1 \text{ and } t_2. \qquad (9.17)$$

Example 9.11

Let $X(t) = \cos(\omega t + \Theta)$ and $Y(t) = \sin(\omega t + \Theta)$, where Θ is a random variable uniformly distributed in $[-\pi, \pi]$. Find the cross-covariance of $X(t)$ and $Y(t)$.

From Example 9.10 we know that $X(t)$ and $Y(t)$ are zero mean. From Eq. (9.16), the cross-covariance is then equal to the cross-correlation:

$$C_{X,Y}(t_1, t_2) = R_{X,Y}(t_1, t_2) = E[\cos(\omega t_1 + \Theta)\sin(\omega t_2 + \Theta)]$$
$$= E\left[-\frac{1}{2}\sin(\omega(t_1 - t_2)) + \frac{1}{2}\sin(\omega(t_1 + t_2) + 2\Theta)\right]$$
$$= -\frac{1}{2}\sin(\omega(t_1 - t_2)),$$

since $E[\sin(\omega(t_1 + t_2) + 2\Theta)] = 0$. $X(t)$ and $Y(t)$ are not uncorrelated random processes because the cross-covariance is not equal to zero for all choices of time samples. Note, however, that $X(t_1)$ and $Y(t_2)$ are *uncorrelated random variables* for t_1 and t_2 such that $\omega(t_1 - t_2) = k\pi$ where k is any integer.

Example 9.12   Signal Plus Noise

Suppose process $Y(t)$ consists of a desired signal $X(t)$ plus noise $N(t)$:

$$Y(t) = X(t) + N(t).$$

Find the cross-correlation between the observed signal and the desired signal assuming that $X(t)$ and $N(t)$ are independent random processes.

From Eq. (8.14), we have

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)]$$
$$= E[X(t_1)\{X(t_2) + N(t_2)\}]$$
$$= R_X(t_1, t_2) + E[X(t_1)]E[N(t_2)]$$
$$= R_X(t_1, t_2) + m_X(t_l)m_N(t_2),$$

where the third equality followed from the fact that $X(t)$ and $N(t)$ are independent.

9.3   DISCRETE-TIME PROCESSES: SUM PROCESS, BINOMIAL COUNTING PROCESS, AND RANDOM WALK

In this section we introduce several important discrete-time random processes. We begin with the simplest class of random processes—independent, identically distributed sequences—and then consider the sum process that results from adding an iid sequence. We show that the sum process satisfies the independent increments property as well as the Markov property. Both of these properties greatly facilitate the calculation of joint probabilities. We also introduce the binomial counting process and the random walk process as special cases of sum processes.

9.3.1   iid Random Process

Let X_n be a discrete-time random process consisting of a sequence of independent, identically distributed (iid) random variables with common cdf $F_X(x)$, mean m, and variance σ^2. The sequence X_n is called the **iid random process**.

The joint cdf for any time instants $n_1, \ldots, n_k$ is given by

$$F_{X_1, \ldots, X_k}(x_1, x_2, \ldots, x_k) = P[X_1 \le x_1, X_2 \le x_2, \ldots, X_k \le x_k]$$

$$= F_X(x_1)F_X(x_2) \ldots F_X(x_k), \tag{9.18}$$

where, for simplicity, X_k denotes X_{n_k}. Equation (9.18) implies that if X_n is discrete-valued, the joint pmf factors into the product of individual pmf's, and if X_n is continuous-valued, the joint pdf factors into the product of the individual pdf's.

The *mean of an iid process* is obtained from Eq. (9.4):

$$m_X(n) = E[X_n] = m \qquad \text{for all } n. \tag{9.19}$$

Thus, the mean is constant.

The autocovariance function is obtained from Eq. (9.6) as follows. If $n_1 \ne n_2$, then

$$C_X(n_1, n_2) = E[(X_{n_1} - m)(X_{n_2} - m)]$$
$$= E[(X_{n_1} - m)]E[(X_{n_2} - m)] = 0,$$

since X_{n_1} and X_{n_2} are independent random variables. If $n_1 = n_2 = n$, then

$$C_X(n_1, n_2) = E[(X_n - m)^2] = \sigma^2.$$

We can express the *autocovariance of the iid process* in compact form as follows:

$$C_X(n_1, n_2) = \sigma^2 \delta_{n_1 n_2}, \tag{9.20}$$

where $\delta_{n_1 n_2} = 1$ if $n_1 = n_2$, and 0 otherwise. Therefore the autocovariance function is zero everywhere except for $n_1 = n_2$. The *autocorrelation function of the iid process* is found from Eq. (9.7):

$$R_X(n_1, n_2) = C_X(n_1, n_2) + m^2. \tag{9.21}$$
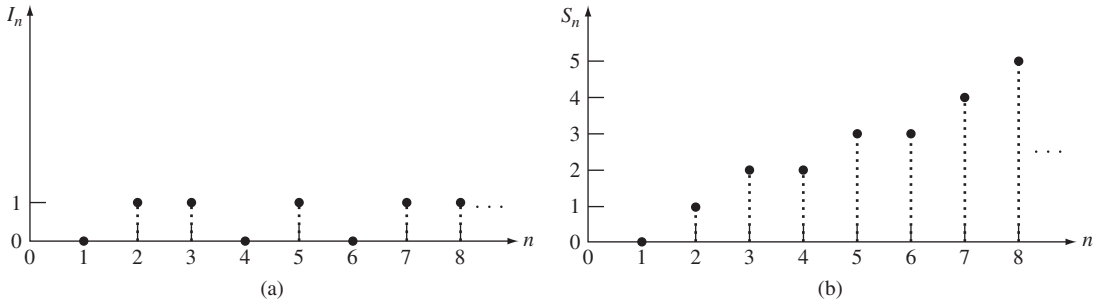
FIGURE 9.4

(a) Realization of a Bernoulli process. $I_n = 1$ indicates that a light bulb fails and is replaced on day n. (b) Realization of a binomial process. S_n denotes the number of light bulbs that have failed up to time n.

Example 9.13   Bernoulli Random Process

Let I_n be a sequence of independent Bernoulli random variables. I_n is then an iid random process taking on values from the set $\{0, 1\}$. A realization of such a process is shown in Fig. 9.4(a). For example, I_n could be an indicator function for the event "a light bulb fails and is replaced on day n."

Since I_n is a Bernoulli random variable, it has mean and variance

$$m_I(n) = p \qquad \text{VAR}[I_n] = p(1 - p).$$

The independence of the I_n's makes probabilities easy to compute. For example, the probability that the first four bits in the sequence are 1001 is

$$P[I_1 = 1, I_2 = 0, I_3 = 0, I_4 = 1]$$
$$= P[I_1 = 1]P[I_2 = 0]P[I_3 = 0]P[I_4 = 1]$$
$$= p^2(1 - p)^2.$$

Similarly, the probability that the second bit is 0 and the seventh is 1 is

$$P[I_2 = 0, I_7 = 1] = P[I_2 = 0]P[I_7 = 1] = p(1 - p).$$

Example 9.14   Random Step Process

An up-down counter is driven by $+1$ or -1 pulses. Let the input to the counter be given by $D_n = 2I_n - 1$, where I_n is the Bernoulli random process, then

$$D_n = \begin{cases} +1 & \text{if } I_n = 1 \\ -1 & \text{if } I_n = 0. \end{cases}$$

For example, D_n might represent the change in position of a particle that moves along a straight line in jumps of ± 1 every time unit. A realization of D_n is shown in Fig. 9.5(a).
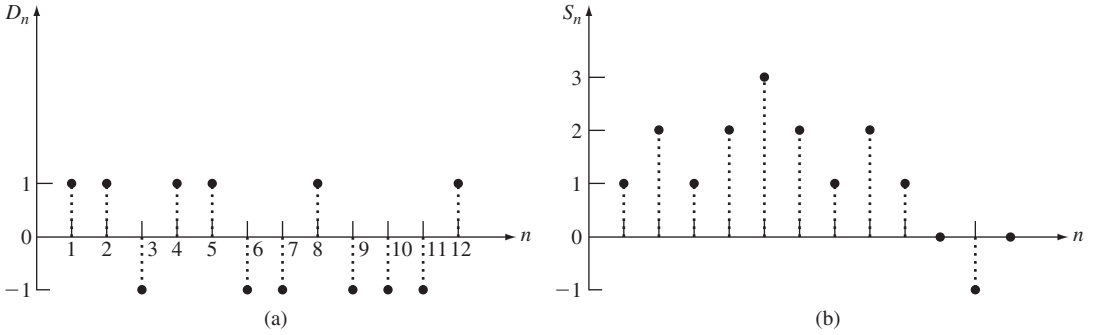
FIGURE 9.5
(a) Realization of a random step process. $D_n = 1$ implies that the particle moves one step to the right at time n. (b) Realization of a random walk process. S_n denotes the position of a particle at time n.

The mean of D_n is

$$m_D(n) = E[D_n] = E[2I_n - 1] = 2E[I_n] - 1 = 2p - 1.$$

The variance of D_n is found from Eqs. (4.37) and (4.38):

$$\text{VAR}[D_n] = \text{VAR}[2I_n - 1] = 2^2 \, \text{VAR}[I_n] = 4p(1 - p).$$

The probabilities of events involving D_n are computed as in Example 9.13.

9.3.2 Independent Increments and Markov Properties of Random Processes

Before proceeding to build random processes from iid processes, we present two very useful properties of random processes. Let $X(t)$ be a random process and consider two time instants, $t_1 < t_2$. The increment of the random process in the interval $t_1 < t \le t_2$ is defined as $X(t_2) - X(t_1)$. *A random process $X(t)$ is said to have **independent increments** if the increments in disjoint intervals are independent random variables*, that is, for any k and any choice of sampling instants $t_1 < t_2 < \cdots < t_k$, the associated increments

$$X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_k) - X(t_{k-1})$$

are independent random variables. In the next subsection, we show that the joint pdf (pmf) of $X(t_1), X(t_2), \ldots, X(t_k)$ is given by the product of the pdf (pmf) of $X(t_1)$ and the marginal pdf's (pmf's) of the individual increments.

Another useful property of random processes that allows us to readily obtain the joint probabilities is the Markov property. *A random process $X(t)$ is said to be **Markov** if the future of the process given the present is independent of the past;* that is, for any k and any choice of sampling instants $t_1 < t_2 < \cdots < t_k$ and for any $x_1, x_2, \ldots, x_k$,

$$f_{X(t_k)}(x_k \,|\, X(t_{k-1}) = x_{k-1}, \ldots, X(t_1) = x_1)$$
$$= f_{X(t_k)}(x_k \,|\, X(t_{k-1}) = x_{k-1}) \tag{9.22}$$
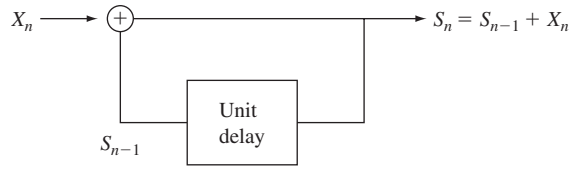
FIGURE 9.6
The sum process $S_n = X_1 + \cdots + X_n$, $S_0 = 0$, can be generated in this way.

if $X(t)$ is continuous-valued, and

$$P[X(t_k) = x_k \mid X(t_{k-1}) = x_{k-1}, \dots, X(t_1) = x_1]$$
$$= P[X(t_k) = x_k \mid X(t_{k-1}) = x_{k-1}] \qquad (9.23)$$

if $X(t)$ is discrete-valued. The expressions on the right-hand side of the above two equations are called the transition pdf and transition pmf, respectively. In the next sections we encounter several processes that satisfy the Markov property. Chapter 11 is entirely devoted to random processes that satisfy this property.

It is easy to show that a random process that has independent increments is also a Markov process. The converse is not true; that is, the Markov property does not imply independent increments.

9.3.3   Sum Processes: The Binomial Counting and Random Walk Processes

Many interesting random processes are obtained as the sum of a sequence of iid random variables, $X_1, X_2, \dots$:

$$S_n = X_1 + X_2 + \cdots + X_n \qquad n = 1, 2, \dots$$
$$= S_{n-1} + X_n, \qquad (9.24)$$

where $S_0 = 0$. We call S_n the **sum process**. The pdf or pmf of S_n is found using the convolution or characteristic-equation methods presented in Section 7.1. Note that S_n depends on the "past," $S_1, \dots, S_{n-1}$, only through S_{n-1}, that is, S_n is independent of the past when S_{n-1} is known. This can be seen clearly from Fig. 9.6, which shows a recursive procedure for computing S_n in terms of S_{n-1} and the increment X_n. Thus S_n *is a Markov process*.

Example 9.15   Binomial Counting Process

Let the I_i be the sequence of independent Bernoulli random variables in Example 9.13, and let S_n be the corresponding sum process. S_n is then the *counting process* that gives the number of successes in the first n Bernoulli trials. The sample function for S_n corresponding to a particular sequence of I_i's is shown in Fig. 9.4(b). Note that the counting process can only increase over time. Note as well that the binomial process can increase by at most one unit at a time. If I_n indicates that a light bulb fails and is replaced on day n, then S_n denotes the number of light bulbs that have failed up to day n.

Since S_n is the sum of n independent Bernoulli random variables, S_n is a binomial random variable with parameters n and $p = P[I = 1]$:

$$P[S_n = j] = \binom{n}{j} p^j (1 - p)^{n-j} \qquad \text{for } 0 \le j \le n,$$

and zero otherwise. Thus S_n has mean np and variance $np(1 - p)$. Note that the mean and variance of this process grow linearly with time. This reflects the fact that as time progresses, that is, as n grows, the range of values that can be assumed by the process increases. If $p > 0$ then we also know that S_n has a tendency to grow steadily without bound over time.

The Markov property of the binomial counting process is easy to deduce. Given that the current value of the process at time $n - 1$ is $S_{n-1} = k$, the process at the next time instant will be k with probability $1 - p$ or $k + 1$ with probability p. Once we know the value of the process at time $n - 1$, the values of the random process prior to time $n - 1$ are irrelevant.

Example 9.16   One-Dimensional Random Walk

Let D_n be the iid process of ± 1 random variables in Example 9.14, and let S_n be the corresponding sum process. S_n can represent the position of a particle at time n. The random process S_n is an example of a **one-dimensional random walk**. A sample function of S_n is shown in Fig. 9.5(b). Unlike the binomial process, the random walk can increase or decrease over time. The random walk process changes by one unit at a time.

The pmf of S_n is found as follows. If there are k "+1"s in the first n trials, then there are $n - k$ "−1"s, and $S_n = k - (n - k) = 2k - n$. Conversely, $S_n = j$ if the number of +1's is $k = (j + n)/2$. If $(j + n)/2$ is not an integer, then S_n cannot equal j. Thus

$$P[S_n = 2k - n] = \binom{n}{k} p^k (1 - p)^{n-k} \qquad \text{for } k \in \{0, 1, \dots, n\}.$$

Since k is the number of successes in n Bernoulli trials, the mean of the random walk is:

$$E[S_n] = 2np - n = n(2p - 1).$$

As time progresses, the random walk can fluctuate over an increasingly broader range of positive and negative values. S_n has a tendency to either grow if $p > 1/2$, or to decrease if $p < 1/2$. The case $p = 1/2$ provides a precarious balance, and we will see later, in Chapter 12, very interesting dynamics. Figure 9.7(a) shows the first 100 steps from a sample function of the random walk with $p = 1/2$. Figure 9.7(b) shows four sample functions of the random walk process with $p = 1/2$ for 1000 steps. Figure 9.7(c) shows four sample functions in the asymmetric case where $p = 3/4$. Note the strong linear growth trend in the process.

The sum process S_n has **independent increments** in nonoverlapping time intervals. To see this consider two time intervals: $n_0 < n \le n_1$ and $n_2 < n \le n_3$, where $n_1 \le n_2$. The increments of S_n in these disjoint time intervals are given by

$$S_{n_1} - S_{n_0} = X_{n_0+1} + \cdots + X_{n_1}$$
$$S_{n_3} - S_{n_2} = X_{n_2+1} + \cdots + X_{n_3}. \qquad (9.25)$$
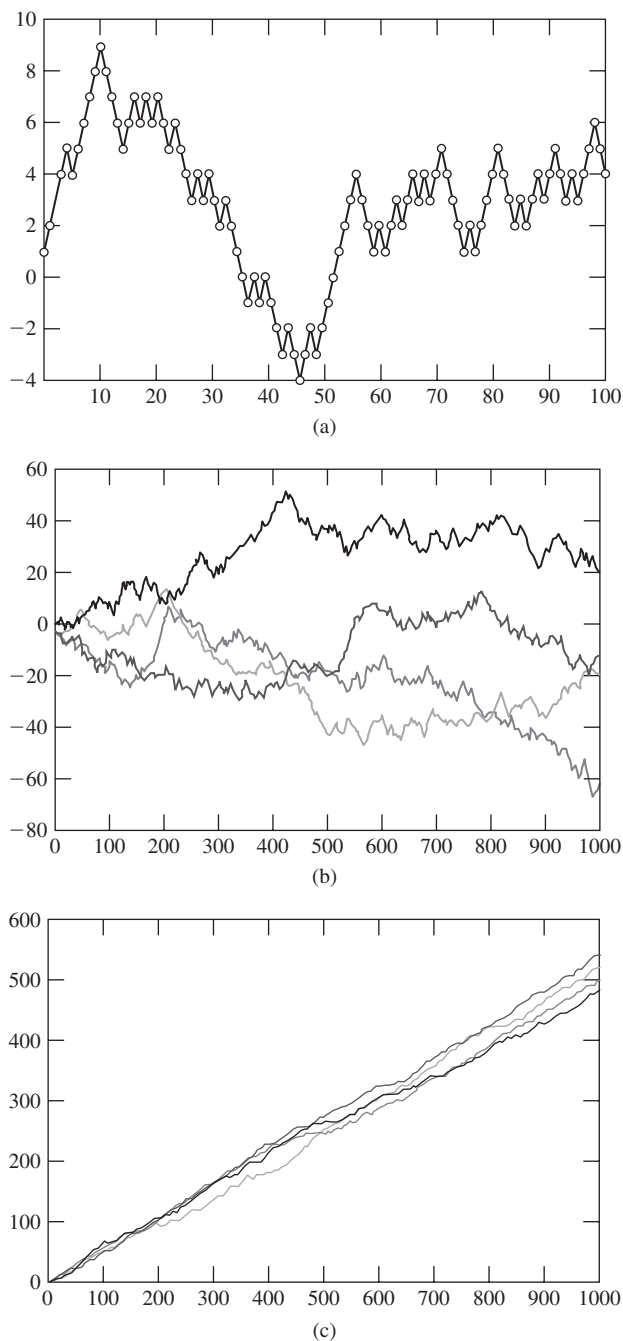
FIGURE 9.7
(a) Random walk process with $p = 1/2$. (b) Four sample functions of symmetric random walk process with $p = 1/2$. (c) Four sample functions of asymmetric random walk with $p = 3/4$.

The above increments do not have any of the X_n's in common, so the independence of the X_n's implies that the increments $(S_{n_1} - S_{n_0})$ and $(S_{n_3} - S_{n_2})$ are independent random variables.

For $n' > n$, the increment $S_{n'} - S_n$ is the sum of $n' - n$ iid random variables, so it has the same distribution as $S_{n'-n}$, the sum of the first $n' - n$ X's, that is,

$$P[S_{n'} - S_n = y] = P[S_{n'-n} = y]. \tag{9.26}$$

Thus increments in intervals of the same length have the same distribution regardless of when the interval begins. For this reason, we also say that S_n has **stationary increments**.

Example 9.17   Independent and Stationary Increments of Binomial Process and Random Walk

The independent and stationary increments property is particularly easy to see for the binomial process since the increments in an interval are the number of successes in the corresponding Bernoulli trials. The independent increment property follows from the fact that the numbers of successes in disjoint time intervals are independent. The stationary increments property follows from the fact that the pmf for the increment in a time interval is the binomial pmf with the corresponding number of trials.

The increment in a random walk process is determined by the same number of successes as a binomial process. It then follows that the random walk also has independent and stationary increments.

The independent and stationary increments property of the sum process S_n makes it easy to compute the joint pmf/pdf for any number of time instants. For simplicity, suppose that the X_n are integer-valued, so S_n is also integer-valued. We compute the joint pmf of S_n at times n_1, n_2, and n_3:

$$P[S_{n_1} = y_1, S_{n_2} = y_2, S_{n_3} = y_3]$$
$$= P[S_{n_1} = y_1, S_{n_2} - S_{n_1} = y_2 - y_1, S_{n_3} - S_{n_2} = y_3 - y_2], \tag{9.27}$$

since the process is equal to y_1, y_2, and y_3 at times n_1, n_2, and n_3, if and only if it is equal to y_1 at time n_1, and the subsequent increments are $y_2 - y_1$, and $y_3 - y_2$. The independent increments property then implies that

$$P[S_{n_1} = y_1, S_{n_2} = y_2, S_{n_3} = y_3]$$
$$= P[S_{n_1} = y_1]P[S_{n_2} - S_{n_1} = y_2 - y_1]P[S_{n_3} - S_{n_2} = y_3 - y_2]. \tag{9.28}$$

Finally, the stationary increments property implies that the *joint pmf of S_n is given by*:

$$P[S_{n_1} = y_1, S_{n_2} = y_2, S_{n_3} = y_3]$$
$$= P[S_{n_1} = y_1]P[S_{n_2-n_1} = y_2 - y_1]P[S_{n_3-n_2} = y_3 - y_2].$$

Clearly, we can use this procedure to write the joint pmf of S_n at any time instants $n_1 < n_2 < \cdots < n_k$ in terms of the pmf at the initial time instant and the pmf's of the subsequent increments:

$$P[S_{n_1} = y_1, S_{n_2} = y_2, \ldots, S_{n_k} = y_k]$$
$$= P[S_{n_1} = y_1]P[S_{n_2-n_1} = y_2 - y_1] \ldots P[S_{n_k-n_{k-1}} = y_k - y_{k-1}]. \qquad (9.29)$$

If the X_n are continuous-valued random variables, then it can be shown that *the joint density of S_n at times $n_1, n_2, \ldots, n_k$ is*:

$$f_{S_{n_1}, S_{n_2}, \ldots, S_{n_k}}(y_1, y_2, \ldots, y_k) = f_{S_{n_1}}(y_1)f_{S_{n_2-n_1}}(y_2 - y_1) \ldots f_{S_{n_k-n_{k-1}}}(y_k - y_{k-1}). \qquad (9.30)$$

Example 9.18   Joint pmf of Binomial Counting Process

Find the joint pmf for the binomial counting process at times n_1 and n_2. Find the probability that $P[S_{n_1} = 0, S_{n_2} = n_2 - n_1]$, that is, the first n_1 trials are failures and the remaining trials are all successes.

Following the above approach we have

$$P[S_{n_1} = y_1, S_{n_2} = y_2] = P[S_{n_1} = y_1]P[S_{n_2} - S_{n_1} = y_2 - y_1]$$

$$= \binom{n_2 - n_1}{y_2 - y_1}p^{y_2-y_1}(1 - p)^{n_2-n_1-y_2+y_1}\binom{n_1}{y_1}p^{y_1}(1 - p)^{n_1-y_1}$$

$$= \binom{n_2 - n_1}{y_2 - y_1}\binom{n_1}{y_1}p^{y_2}(1 - p)^{n_2-y_2}.$$

The requested probability is then:

$$P[S_{n_1} = 0, S_{n_2} = n_2 - n_1] = \binom{n_2 - n_1}{n_2 - n_1}\binom{n_1}{0}p^{n_2-n_1}(1 - p)^{n_1} = p^{n_2-n_1}(1 - p)^{n_1}$$

which is what we would obtain from a direct calculation for Bernoulli trials.

Example 9.19   Joint pdf of Sum of iid Gaussian Sequence

Let X_n be a sequence of iid Gaussian random variables with zero mean and variance σ^2. Find the joint pdf of the corresponding sum process at times n_1 and n_2.

From Example 7.3, we know that S_n is a Gaussian random variable with mean zero and variance $n\sigma^2$. The joint pdf of S_n at times n_j and n_2 is given by

$$f_{S_{n_1}, S_{n_2}}(y_1, y_2) = f_{S_{n_2-n_1}}(y_2 - y_1)f_{S_{n_1}}(y_1)$$

$$= \frac{1}{\sqrt{2\pi(n_2 - n_1)\sigma^2}}e^{-(y_2-y_1)^2/[2(n_2-n_1)\sigma^2]}\frac{1}{\sqrt{2\pi n_1\sigma^2}}e^{-y_1^2/2n_1\sigma^2}.$$

Since the sum process S_n is the sum of n iid random variables, it has mean and variance:

$$m_S(n) = E[S_n] = nE[X] = nm \qquad (9.31)$$

$$\text{VAR}[S_n] = n\,\text{VAR}[X] = n\sigma^2. \qquad (9.32)$$

The property of independent increments allows us to compute the autocovariance in an interesting way. Suppose $n \leq k$ so $n = \min(n, k)$, then

$$
\begin{aligned}
C_S(n, k) &= E[(S_n - nm)(S_k - km)] \\
&= E[(S_n - nm)\{(S_n - nm) + (S_k - km) - (S_n - nm)\}] \\
&= E[(S_n - nm)^2] + E[(S_n - nm)(S_k - S_n - (k - n)m)].
\end{aligned}
$$

Since S_n and the increment $S_k - S_n$ are independent,

$$
\begin{aligned}
C_S(n, k) &= E[(S_n - nm)^2] + E[(S_n - nm)]E[(S_k - S_n - (k - n)m)] \\
&= E[(S_n - nm)^2] \\
&= \text{VAR}[S_n] = n\sigma^2,
\end{aligned}
$$

since $E[S_n - nm] = 0$. Similarly, if $k = \min(n, k)$, we would have obtained $k\sigma^2$. Therefore the *autocovariance of the sum process* is

$$
C_S(n, k) = \min(n, k)\sigma^2. \tag{9.33}
$$

Example 9.20   Autocovariance of Random Walk

Find the autocovariance of the one-dimensional random walk.

From Example 9.14 and Eqs. (9.32) and (9.33), S_n has mean $n(2p - 1)$ and variance $4np(1 - p)$. Thus its autocovariance is given by

$$
C_s(n, k) = \min(n, k)4p(1 - p).
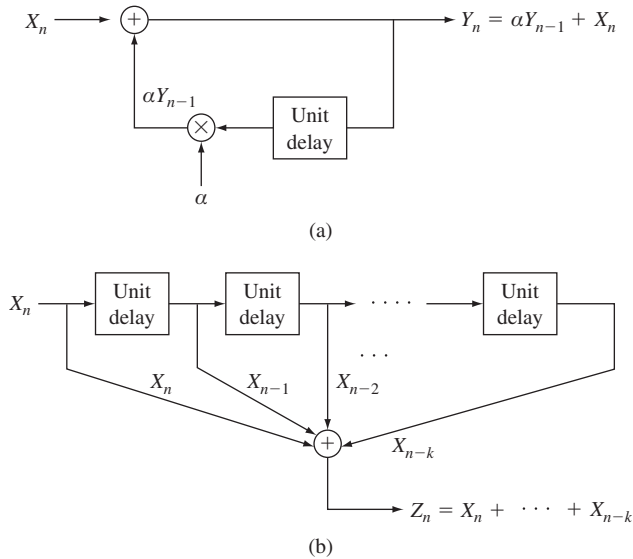$$

(a)

(b)

FIGURE 9.8
(a) First-order autoregressive process; (b) Moving average process.

The sum process can be generalized in a number of ways. For example, the recursive structure in Fig. 9.6 can be modified as shown in Fig. 9.8(a). We then obtain first-order *autoregressive random processes*, which are of interest in time series analysis and in digital signal processing. If instead we use the structure shown in Fig. 9.8(b), we obtain an example of a *moving average process*. We investigate these processes in Chapter 10.

9.4   POISSON AND ASSOCIATED RANDOM PROCESSES

In this section we develop the Poisson random process, which plays an important role in models that involve counting of events and that find application in areas such as queueing systems and reliability analysis. We show how the continuous-time Poisson random process can be obtained as the limit of a discrete-time process. We also introduce several random processes that are derived from the Poisson process.

9.4.1   Poisson Process

Consider a situation in which events occur at random instants of time at an average rate of λ events per second. For example, an event could represent the arrival of a customer to a service station or the breakdown of a component in some system. Let $N(t)$ be the number of event occurrences in the time interval $[0, t]$. $N(t)$ is then a nondecreasing, integer-valued, continuous-time random process as shown in Fig. 9.9.
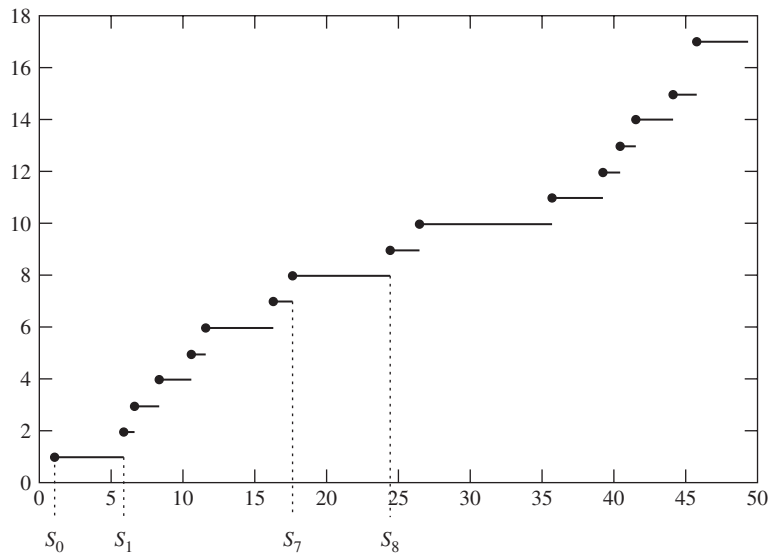

FIGURE 9.9
A sample path of the Poisson counting process. The event occurrence times are denoted by $S_1, S_2, \ldots$. The jth interevent time is denoted by $X_j = S_j - S_{j-1}$.

Suppose that the interval $[0, t]$ is divided into n subintervals of very short duration $\delta = t/n$. Assume that the following two conditions hold:

1. The probability of more than one event occurrence in a subinterval is negligible compared to the probability of observing one or zero events.
2. Whether or not an event occurs in a subinterval is independent of the outcomes in other subintervals.

The first assumption implies that the outcome in each subinterval can be viewed as a Bernoulli trial. The second assumption implies that these Bernoulli trials are independent. The two assumptions together imply that the counting process $N(t)$ *can be approximated by the binomial counting process* discussed in the previous section.

If the probability of an event occurrence in each subinterval is p, then the expected number of event occurrences in the interval $[0, t]$ is np. Since events occur at a rate of λ events per second, the average number of events in the interval $[0, t]$ is λt. Thus we must have that

$$\lambda t = np.$$

If we now let $n \to \infty$ (i.e., $\delta = t/n \to 0$) and $p \to 0$ while $np = \lambda t$ remains fixed, then from Eq. (3.40) the binomial distribution approaches a Poisson distribution with parameter λt. We therefore conclude that the number of event occurrences $N(t)$ in the interval $[0, t]$ has a Poisson distribution with mean λt:

$$P[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad \text{for } k = 0, 1, \ldots. \tag{9.34a}$$

For this reason $N(t)$ is called the **Poisson process**. The mean function and the variance function of the Poisson process are given by:

$$m_N(t) = E[N(t) = k] = \lambda t \quad \text{and} \quad \text{VAR}[N(t)] = \lambda t. \tag{9.34b}$$

In Section 11.3 we rederive the Poisson process using results from Markov chain theory.

The process $N(t)$ inherits the property of independent and stationary increments from the underlying binomial process. First, *the distribution for the number of event occurrences in any interval of length t is given by Eq. (9.34a)*. Next, the independent and stationary increments property allows us to write the joint pmf for $N(t)$ at any number of points. For example, for $t_1 < t_2$,

$$\begin{aligned}
P[N(t_1) = i, N(t_2) = j] &= P[N(t_1) = i]P[N(t_2) - N(t_1) = j - i] \\
&= P[N(t_1) = i]P[N(t_2 - t_1) = j - i] \\
&= \frac{(\lambda t_1)^i e^{-\lambda t_1}}{i!} \frac{(\lambda(t_2 - t_1))^j e^{-\lambda(t_2 - t_1)}}{(j - i)!}.
\end{aligned} \tag{9.35a}$$

The independent increments property also allows us to calculate the autocovariance of $N(t)$. For $t_1 \le t_2$:

$$C_N(t_1, t_2) = E[(N(t_1) - \lambda t_1)(N(t_2) - \lambda t_2)]$$
$$= E[(N(t_1) - \lambda t_1)\{N(t_2) - N(t_1) - \lambda t_2 + \lambda t_1 + (N(t_1) - \lambda t_1)\}]$$
$$= E[(N(t_1) - \lambda t_1)]E[(N(t_2) - N(t_1) - \lambda(t_2 - t_1)] + \text{VAR}[N(t_1)]$$
$$= \text{VAR}[N(t_1)] = \lambda t_1. \tag{9.35b}$$

Example 9.21

Inquiries arrive at a recorded message device according to a Poisson process of rate 15 inquiries per minute. Find the probability that in a 1-minute period, 3 inquiries arrive during the first 10 seconds and 2 inquiries arrive during the last 15 seconds.

   The arrival rate in seconds is $\lambda = 15/60 = 1/4$ inquiries per second. Writing time in seconds, the probability of interest is

$$P[N(10) = 3 \text{ and } N(60) - N(45) = 2].$$

By applying first the independent increments property, and then the stationary increments property, we obtain

$$P[N(10) = 3 \text{ and } N(60) - N(45) = 2]$$
$$= P[N(10) = 3]P[N(60) - N(45) = 2]$$
$$= P[N(10) = 3]P[N(60 - 45) = 2]$$
$$= \frac{(10/4)^3 e^{-10/4}}{3!} \frac{(15/4)^2 e^{-15/4}}{2!}.$$

Consider the time T between event occurrences in a Poisson process. Again suppose that the time interval $[0, t]$ is divided into n subintervals of length $\delta = t/n$. The probability that the interevent time T exceeds t seconds is equivalent to no event occurring in t seconds (or in n Bernoulli trials):

$$P[T > t] = P[\text{no events in } t \text{ seconds}]$$
$$= (1 - p)^n$$
$$= \left(1 - \frac{\lambda t}{n}\right)^n$$
$$\rightarrow e^{-\lambda t} \quad \text{as } n \rightarrow \infty. \tag{9.36}$$

Equation (9.36) implies that T is an exponential random variable with parameter λ. Since the times between event occurrences in the underlying binomial process are independent geometric random variables, it follows that the sequence of interevent times in a Poisson process is composed of independent random variables. We therefore conclude that the *interevent times in a Poisson process form an iid sequence of exponential random variables with mean* $1/\lambda$.

Another quantity of interest is the time S_n at which the nth event occurs in a Poisson process. Let T_j denote the iid exponential interarrival times, then

$$S_n = T_1 + T_2 + \cdots + T_n.$$

In Example 7.5, we saw that the sum of n iid exponential random variables has an Erlang distribution. Thus the pdf of S_n *is an Erlang random variable*:

$$f_{S_n}(y) = \frac{(\lambda y)^{n-1}}{(n-1)!} \lambda e^{-\lambda y} \qquad \text{for } y \geq 0. \tag{9.37}$$

Example 9.22

Find the mean and variance of the time until the tenth inquiry in Example 9.20.

The arrival rate is $\lambda = 1/4$ inquiries per second, so the interarrival times are exponential random variables with parameter λ. From Table 4.1, the mean and variance of exponential interarrival times then $1/\lambda$ and $1/\lambda^2$, respectively. The time of the tenth arrival is the sum of ten such iid random variables, thus

$$E[S_{10}] = 10E[T] = \frac{10}{\lambda} = 40 \text{ sec}$$

$$\text{VAR}[S_{10}] = 10\,\text{VAR}[T] = \frac{10}{\lambda^2} = 160 \text{ sec}^2.$$

In applications where the Poisson process models customer interarrival times, it is customary to say that arrivals occur "at random." We now explain what is meant by this statement. Suppose that we are *given* that only one arrival occurred in an interval $[0, t]$ and we let X be the arrival time of the single customer. For $0 < x < t$, $N(x)$ is the number of events up to time x, and $N(t) - N(x)$ is the increment in the interval $(x, t]$, then:

$$
\begin{aligned}
P[X \leq x] &= P[N(x) = 1 \mid N(t) = 1] \\
&= \frac{P[N(x) = 1 \text{ and } N(t) = 1]}{P[N(t) = 1]} \\
&= \frac{P[N(x) = 1 \text{ and } N(t) - N(x) = 0]}{P[N(t) = 1]} \\
&= \frac{P[N(x) = 1]P[N(t) - N(x) = 0]}{P[N(t) = 1]} \\
&= \frac{\lambda x e^{-\lambda x} e^{-\lambda(t-x)}}{\lambda t e^{-\lambda t}} \\
&= \frac{x}{t}. \tag{9.38}
\end{aligned}
$$

Equation (9.38) implies that given that one arrival has occurred in the interval $[0, \ t]$, then the customer arrival time is uniformly distributed in the interval $[0, t]$. It is in this sense that customer arrival times occur "at random." It can be shown that *if the number of amvals in the interval $[0, t]$ is k, then the individual arrival times are distributed independently and uniformly in the interval.*

Example 9.23

Suppose two customers arrive at a shop during a two-minute period. Find the probability that both customers arrived during the first minute.

The arrival times of the customers are independent and uniformly distributed in the two-minute interval. Each customer arrives during the first minute with probability 1/2. Thus the probability that both arrive during the first minute is $(1/2)^2 = 1/4$. This answer can be verified by showing that $P[N(1) = 2 \,|\, N(2) = 2] = 1/4$.

9.4.2   Random Telegraph Signal and Other Processes Derived from the Poisson Process

Many processes are derived from the Poisson process. In this section, we present two examples of such random processes.

Example 9.24   Random Telegraph Signal

Consider a random process $X(t)$ that assumes the values ± 1. Suppose that $X(0) = +1$ or -1 with probability 1/2, and suppose that $X(t)$ changes polarity with each occurrence of an event in a Poisson process of rate α. Figure 9.10 shows a sample function of $X(t)$.

The pmf of $X(t)$ is given by

$$P[X(t) = \pm 1] = P[X(t) = \pm 1 \,|\, X(0) = 1]P[X(0) = 1]$$
$$+ P[X(t) = \pm 1 \,|\, X(0) = -1]P[X(0) = -1]. \quad (9.39)$$

The conditional pmf's are found by noting that $X(t)$ will have the same polarity as $X(0)$ only when an even number of events occur in the interval $(0, t]$. Thus

$$P[X(t) = \pm 1 \,|\, X(0) = \pm 1] = P[N(t) = \text{even integer}]$$
$$= \sum_{j=0}^{\infty} \frac{(\alpha t)^{2j}}{(2j)!} e^{-\alpha t}$$
$$= e^{-\alpha t} \frac{1}{2} \{e^{\alpha t} + e^{-\alpha t}\}$$
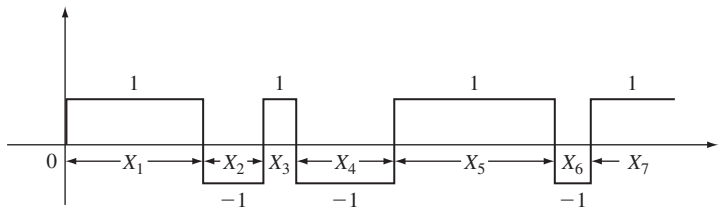$$= \frac{1}{2}(1 + e^{-2\alpha t}). \quad (9.40)$$

FIGURE 9.10
Sample path of a random telegraph signal. The times between transitions X_j are iid exponential random variables.

$X(t)$ and $X(0)$ will differ in sign if the number of events in t is odd:

$$P[X(t) = \pm 1 \mid X(0) = \mp 1] = \sum_{j=0}^{\infty} \frac{(\alpha t)^{2j+1}}{(2j+1)!} e^{-\alpha t}$$

$$= e^{-\alpha t} \frac{1}{2} \{ e^{\alpha t} - e^{-\alpha t} \}$$

$$= \frac{1}{2}(1 - e^{-2\alpha t}). \tag{9.41}$$

We obtain the pmf for $X(t)$ by substituting into Eq. (9.40):

$$P[X(t) = 1] = \frac{1}{2}\frac{1}{2}\{1 + e^{-2\alpha t}\} + \frac{1}{2}\frac{1}{2}\{1 - e^{-2\alpha t}\} = \frac{1}{2}$$

$$P[X(t) = -1] = 1 - P[X(t) = 1] = \frac{1}{2}. \tag{9.42}$$

Thus the random telegraph signal is equally likely to be ± 1 at any time $t > 0$.

The mean and variance of $X(t)$ are

$$m_X(t) = 1P[X(t) = 1] + (-1)P[X(t) = -1] = 0$$

$$\text{VAR}[X(t)] = E[X(t)^2] = (1^2)P[X(t) = 1] + (-1)^2 P[X(t) = -1] = 1. \tag{9.43}$$

The autocovariance of $X(t)$ is found as follows:

$$C_X(t_1, t_2) = E[X(t_1)X(t_2)]$$

$$= 1P[X(t_1) = X(t_2)] + (-1)P[X(t_1) \neq X(t_2)]$$

$$= \frac{1}{2}\{1 + e^{-2\alpha|t_2 - t_1|}\} - \frac{1}{2}\{1 - e^{-2\alpha|t_2 - t_1|}\}$$

$$= e^{-2\alpha|t_2 - t_1|}. \tag{9.44}$$

Thus time samples of $X(t)$ become less and less correlated as the time between them increases.

The Poisson process and the random telegraph processes are examples of the continuous-time Markov chain processes that are discussed in Chapter 11.
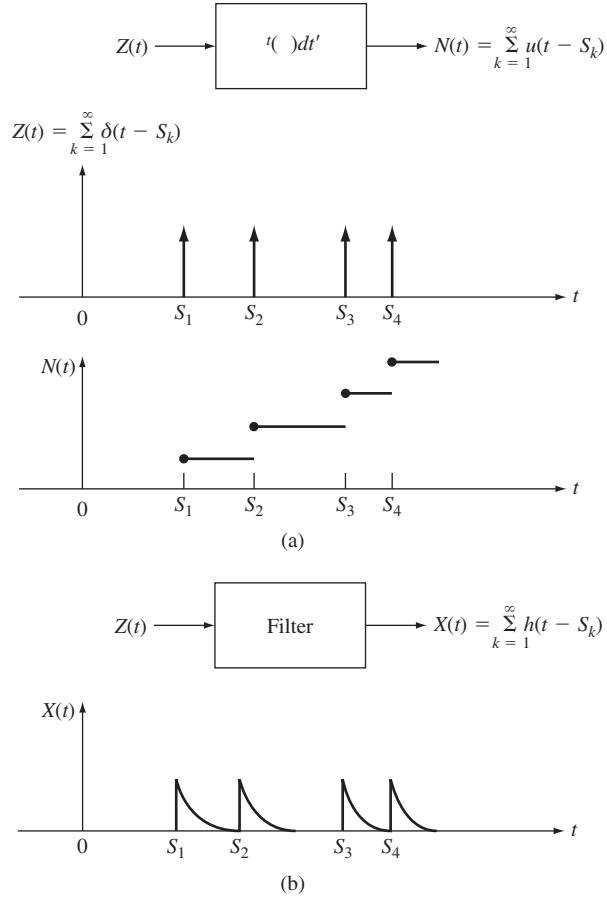

Example 9.25   Filtered Poisson Impulse Train

The Poisson process is zero at $t = 0$ and increases by one unit at the random arrival times $S_j, j = 1, 2, \ldots$. Thus the Poisson process can be expressed as the sum of randomly shifted step functions:

$$N(t) = \sum_{i=1}^{\infty} u(t - S_i) \qquad N(0) = 0,$$

where the S_i are the arrival times.

Since the integral of a delta function $\delta(t - S)$ is a step function $u(t - S)$, we can view $N(t)$ as the result of integrating a train of delta functions that occur at times S_n, as shown in Fig. 9.11(a):

$$Z(t) \longrightarrow \boxed{\int^t(\ )dt'} \longrightarrow N(t) = \sum_{k=1}^{\infty} u(t - S_k)$$

$$Z(t) = \sum_{k=1}^{\infty} \delta(t - S_k)$$

(a)

$$Z(t) \longrightarrow \boxed{\text{Filter}} \longrightarrow X(t) = \sum_{k=1}^{\infty} h(t - S_k)$$

(b)

FIGURE 9.11
(a) Poisson process as integral of train of delta functions. (b) Filtered
train of delta functions.

$$Z(t) = \sum_{i=1}^{\infty} \delta(t - S_i).$$

We can obtain other continuous-time processes by replacing the step function by another
function $h(t)$,[1] as shown in Fig. 9.11(b):

$$X(t) = \sum_{i=1}^{\infty} h(t - S_i). \tag{9.45}$$

For example, $h(t)$ could represent the current pulse that results when a photoelectron hits a de-
tector. $X(t)$ is then the total current flowing at time t. $X(t)$ is called a **shot noise** process.

---

[1]This is equivalent to passing $Z(t)$ through a linear system whose response to a delta function is $h(t)$.

The following example shows how the properties of the Poisson process can be used to evaluate averages involving the filtered process.

Example 9.26 Mean of Shot Noise Process

Find the expected value of the shot noise process $X(t)$.

We condition on $N(t)$, the number of impulses that have occurred up to time t:

$$E[X(t)] = E[E[X(t)|N(t)]].$$

Suppose $N(t) = k$, then

$$E[X(t)|N(t) = k] = E\left[\sum_{j=1}^{k} h(t - S_j)\right]$$

$$= \sum_{j=1}^{k} E[h(t - S_j)].$$

Since the arrival times, $S_1, \ldots, S_k$, when the impulses occurred are independent, uniformly distributed in the interval $[0, t]$,

$$E[h(t - S_j)] = \int_0^t h(t - s)\frac{ds}{t} = \frac{1}{t}\int_0^t h(u)\, du.$$

Thus

$$E[X(t)|N(t) = k] = \frac{k}{t}\int_0^t h(u)\, du,$$

and

$$E[X(t)|N(t)] = \frac{N(t)}{t}\int_0^t h(u)\, du.$$

Finally, we obtain

$$E[X(t)] = E[E[X(t)|N(t)]]$$

$$= \frac{E[N(t)]}{t}\int_0^t h(u)\, du$$

$$= \lambda \int_0^t h(u)\, du, \qquad (9.46)$$

where we used the fact that $E[N(t)] = \lambda t$. Note that $E[X(t)]$ approaches a constant value as t becomes large if the above integral is finite.

9.5 GAUSSIAN RANDOM PROCESSES, WIENER PROCESS, AND BROWNIAN MOTION

In this section we continue the introduction of important random processes. First, we introduce the class of Gaussian random processes which find many important applications in electrical engineering. We then develop an example of a Gaussian random process: the Wiener random process which is used to model Brownian motion.

9.5.1   Gaussian Random Processes

A random process $X(t)$ is a **Gaussian random process** if the samples $X_1 = X(t_1)$, $X_2 = X(t_2), \ldots, X_k = X(t_k)$ are jointly Gaussian random variables for all k, and all choices of $t_1, \ldots, t_k$. This definition applies to both discrete-time and continuous-time processes. Recall from Eq. (6.42) that the joint pdf of jointly Gaussian random variables is determined by the vector of means and by the covariance matrix:

$$f_{X_1, X_2, \ldots, X_k}(x_1, x_2, \ldots, x_k) = \frac{e^{-1/2(\mathbf{x}-\mathbf{m})^{\mathrm{T}} K^{-1}(\mathbf{x}-\mathbf{m})}}{(2\pi)^{k/2} |K|^{1/2}}. \tag{9.47a}$$

In the case of Gaussian random processes, the mean vector and the covariance matrix are the values of the mean function and covariance function at the corresponding time instants:

$$\mathbf{m} = \begin{bmatrix} m_X(t_1) \\ \vdots \\ m_X(t_k) \end{bmatrix} \qquad K = \begin{bmatrix} C_X(t_1, t_1) & C_X(t_1, t_2) & \cdots & C_X(t_1, t_k) \\ C_X(t_2, t_1) & C_X(t_2, t_2) & \cdots & C_X(t_2, t_k) \\ \vdots & \vdots & & \vdots \\ C_X(t_k, t_1) & \cdots & & C_X(t_k, t_k) \end{bmatrix}. \tag{9.47b}$$

Gaussian random processes therefore have the very special property that their joint pdf's are completely specified by the mean function of the process $m_X(t)$ and by the covariance function $C_X(t_1, t_2)$. In Chapter 6 we saw that the linear transformations of jointly Gaussian random vectors result in jointly Gaussian random vectors as well. We will see in Chapter 10 that Gaussian random processes also have the property that the linear operations on a Gaussian process (e.g., a sum, derivative, or integral) results in another Gaussian random process. These two properties, combined with the fact that many signal and noise processes are accurately modeled as Gaussian, make Gaussian random processes the most useful model in signal processing.

Example 9.27   iid Discrete-Time Gaussian Random Process

Let the discrete-time random process X_n be a sequence of independent Gaussian random variables with mean m and variance σ^2. The covariance matrix for the times $n_1, \ldots, n_k$ is

$$\{C_X(n_1, n_2)\} = \{\sigma^2 \delta_{ij}\} = \sigma^2 I,$$

where $\delta_{ij} = 1$ when $i = j$ and 0 otherwise, and I is the identity matrix. Thus the joint pdf for the vector $\mathbf{X}_n = (X_{n_1}, \ldots, X_{n_k})$ is

$$f_{\mathbf{X}_n}(x_1, x_2, \ldots, x_k) = \frac{1}{(2\pi\sigma^2)^{k/2}} \exp\left\{ -\sum_{i=1}^{k} (x_i - m)^2/2\sigma^2 \right\}.$$

The Gaussian iid random process has the property that the value at every time instant is independent of the value at all other time instants.

Example 9.28 Continuous-Time Gaussian Random Process

Let $X(t)$ be a continuous-time Gaussian random process with mean function and covariance function given by:

$$m_X(t) = 3t \qquad C_X(t_1, t_2) = 9e^{-2|t_1 - t_2|}.$$

Find $P[X(3) < 6]$ and $P[X(1) + X(2) > 2]$.

The sample $X(3)$ has a Gaussian pdf with mean $m_X(3) = 3(3) = 9$ and variance $\sigma_X^2(3) = C_X(3, 3) = 9e^{-2|3-3|} = 9$. To calculate $P[X(3) < 6]$ we put $X(3)$ in standard form:

$$P[X(3) < 6] = P\left[\frac{X(3) - 9}{\sqrt{9}} < \frac{6 - 9}{\sqrt{9}}\right] = 1 - Q(-1) = Q(1) = 0.16.$$

From Example 6.24 we know that the sum of two Gaussian random variables is also a Gaussian random variable with mean and variance given by Eq. (6.47). Therefore the mean and variance of $X(1) + X(2)$ are given by:

$$E[X(1) + X(2)] = m_X(1) + m_X(2) = 3 + 6 = 9$$

$$\begin{aligned} \text{VAR}[X(1) + X(2)] &= C_X(1, 1) + C_X(1, 2) + C_X(2, 1) + C_X(2, 2) \\ &= 9\{e^{-2|1-1|} + e^{-2|2-1|} + e^{-2|1-2|} + e^{-2|2-2|}\} \\ &= 9\{2 + 2e^{-2}\} = 20.43. \end{aligned}$$

To calculate $P[X(1) + X(2) > 2]$ we put $X(1) + X(2)$ in standard form:

$$P[X(1) + X(2) > 15] = P\left[\frac{X(1) + X(2) - 9}{\sqrt{20.43}} > \frac{15 - 9}{\sqrt{20.43}}\right] = Q(1.327) = 0.0922.$$

9.5.2 Wiener Process and Brownian Motion

We now construct a continuous-time Gaussian random process as a limit of a discrete-time process. Suppose that the symmetric random walk process (i.e., $p = 1/2$) of Example 9.16 takes steps of magnitude $\pm h$ every δ seconds. We obtain a continuous-time process by letting $X_\delta(t)$ be the accumulated sum of the random step process up to time t. $X_\delta(t)$ is a staircase function of time that takes jumps of $\pm h$ every δ seconds. At time t, the process will have taken $n = [t/\delta]$ jumps, so it is equal to

$$X_\delta(t) = h(D_1 + D_2 + \cdots + D_{[t/\delta]}) = hS_n. \tag{9.48}$$

The mean and variance of $X_\delta(t)$ are

$$E[X_\delta(t)] = hE[S_n] = 0$$
$$\text{VAR}[X_\delta(t)] = h^2 n \, \text{VAR}[D_n] = h^2 n,$$

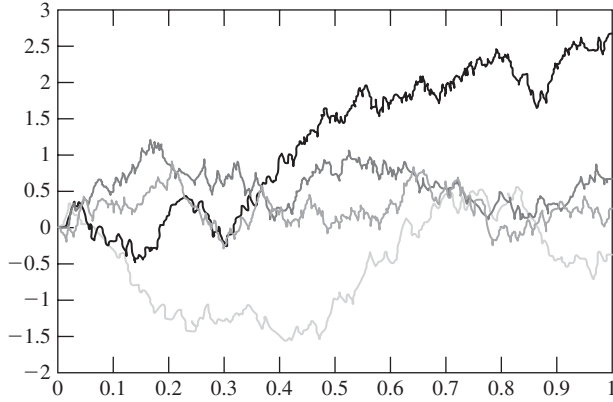where we used the fact that $\text{VAR}[D_n] = 4p(1 - p) = 1$ since $p = 1/2$.

FIGURE 9.12
Four sample functions of the Wiener process.

Suppose that we take a limit where we simultaneously shrink the size of the jumps and the time between jumps. In particular let $\delta \to 0$ and $h \to 0$ with $h = \sqrt{\alpha\delta}$ and let $X(t)$ denote the resulting process.

$X(t)$ then has *mean* and *variance* given by

$$E[X(t)] = 0 \tag{9.49a}$$

$$\text{VAR}[X(t)] = (\sqrt{\alpha\delta})^2 (t/\delta) = \alpha t. \tag{9.49b}$$

Thus we obtain a continuous-time process $X(t)$ that begins at the origin, has zero mean for all time, but has a variance that increases linearly with time. Figure 9.12 shows four sample functions of the process. Note the similarities in fluctuations to the realizations of a symmetric random walk in Fig. 9.7(b). $X(t)$ is called the **Wiener random process**. It is used to model *Brownian motion*, the motion of particles suspended in a fluid that move under the rapid and random impact of neighboring particles.

As $\delta \to 0$, Eq. (9.48) implies that $X(t)$ approaches the sum of an infinite number of random variables since $n = [t/\delta] \to \infty$:

$$X(t) = \lim_{\delta \to 0} h S_n = \lim_{n \to \infty} \sqrt{\alpha t} \frac{S_n}{\sqrt{n}}. \tag{9.50}$$

By the central limit theorem the pdf of $X(t)$ therefore approaches that of a Gaussian random variable with mean zero and variance αt:

$$f_{X(t)}(x) = \frac{1}{\sqrt{2\pi\alpha t}} e^{-x^2/2\alpha t}. \tag{9.51}$$

$X(t)$ inherits the property of independent and stationary increments from the random walk process from which it is derived. As a result, the *joint pdf of $X(t)$* at

several times $t_1, t_2, \ldots, t_k$ can be obtained by using Eq. (9.30):

$$f_{X(t_1),\ldots,X(t_k)}(x_1,\ldots,x_k) = f_{X(t_1)}(x_1)f_{X(t_2-t_1)}(x_2-x_1)\ldots f_{X(t_k-t_{k-1})}(x_k-x_{k-1})$$

$$= \frac{\exp\left\{-\dfrac{1}{2}\left[\dfrac{x_1^2}{\alpha t_1} + \dfrac{(x_2-x_1)^2}{\alpha(t_2-t_1)} + \cdots + \dfrac{(x_k-x_{k-1})^2}{\alpha(t_k-t_{k-1})}\right]\right\}}{\sqrt{(2\pi\alpha)^k t_1(t_2-t_1)\ldots(t_k-t_{k-1})}}. \quad (9.52)$$

The independent increments property and the same sequence of steps that led to Eq. (9.33) can be used to show that the *autocovariance of X(t)* is given by

$$C_X(t_1, t_2) = \alpha \min(t_1, t_2) = \alpha\, t_1 \text{ for } t_1 < t_2. \quad (9.53)$$

By comparing Eq. (9.53) and Eq. (9.35b), we see that the Wiener process and the Poisson process have the same covariance function despite the fact that the two processes have very different sample functions. This underscores the fact that the mean and autocovariance functions are only partial descriptions of a random process.

Example 9.29

Show that the Wiener process is a Gaussian random process.

Equation (9.52) shows that the random variables $X(t_1), X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_k) - X(t_{k-1})$, are independent Gaussian random variables. The random variables $X(t_1), X(t_2), X(t_3), \ldots, X(t_k)$, can be obtained from the $X(t_1)$ and the increments by a linear transformation:

$$X(t_1) = X(t_1)$$
$$X(t_2) = X(t_1) + (X(t_2) - X(t_1))$$
$$X(t_3) = X(t_1) + (X(t_2) - X(t_1)) + (X(t_3) - X(t_2))$$
$$\vdots$$
$$X(t_k) = X(t_1) + (X(t_2) - X(t_1)) + \cdots + (X(t_k) - X(t_{k-1})). \quad (9.54)$$

It then follows (from Eq. 6.45) that $X(t_1), X(t_2), X(t_3), \ldots, X(t_k)$ are jointly Gaussian random variables, and that $X(t)$ is a Gaussian random process.

9.6 STATIONARY RANDOM PROCESSES

Many random processes have the property that the nature of the randomness in the process does not change with time. An observation of the process in the time interval (t_0, t_1) exhibits the same type of random behavior as an observation in some other time interval $(t_0 + \tau, t_1 + \tau)$. This leads us to postulate that the probabilities of samples of the process do not depend on the instant when we begin taking observations, that is, probabilities involving samples taken at times $t_1, \ldots, t_k$ will not differ from those taken at $t_1 + \tau, \ldots, t_k + \tau$.

Example 9.30 Stationarity and Transience

An urn has 6 white balls each with the label "0" and 5 white balls with the label "1". The following sequence of experiments is performed: A ball is selected and the number noted; the first time a white ball is selected it is not put back in the urn, but otherwise balls are always put back in the urn.

The random process that results from this sequence of experiments clearly has a transient phase and a stationary phase. The transient phase consists of a string of n consecutive 1's and it ends with the first occurrence of a "0". During the transient phase $P[I_n = 0] = 6/11$, and the mean duration of the transient phase is geometrically distributed with mean 11/6. After the first occurrence of a "0", the process enters a "stationary" phase where the process is a binary equiprobable iid sequence. The statistical behavior of the process does not change once the stationary phase is reached.

If we are dealing with random processes that began at $t = -\infty$, then the above condition can be stated precisely as follows. *A discrete-time or continuous-time random process $X(t)$ is* **stationary** *if the joint distribution of any set of samples does not depend on the placement of the time origin.* This means that the joint cdf of $X(t_1), X(t_2), \ldots, X(t_k)$ is the same as that of $X(t_1 + \tau), X(t_2 + \tau), \ldots, X(t_k + \tau)$:

$$F_{X(t_1), \ldots, X(t_k)}(x_1, \ldots, x_k) = F_{X(t_1+\tau), \ldots, X(t_k+\tau)}(x_1, \ldots, x_k), \tag{9.55}$$

for all time shifts τ, all k, and all choices of sample times $t_1, \ldots, t_k$. If a process begins at some definite time (i.e., $n = 0$ or $t = 0$), then we say it is stationary if its joint distributions do not change under time shifts to the right.

Two processes $X(t)$ and $Y(t)$ are said to be **jointly stationary** if the joint cdf's of $X(t_1), \ldots, X(t_k)$ and $Y(t_1'), \ldots, Y(t_j')$ do not depend on the placement of the time origin for all k and j and all choices of sampling times $t_1, \ldots, t_k$ and $t_1', \ldots, t_j'$.

The *first-order cdf of a stationary random process must be independent of time*, since by Eq. (9.55),

$$F_{X(t)}(x) = F_{X(t+\tau)}(x) = F_X(x) \qquad \text{all } t, \tau. \tag{9.56}$$

This implies that the mean and variance of $X(t)$ are constant and independent of time:

$$m_X(t) = E[X(t)] = m \qquad \text{for all } t \tag{9.57}$$

$$\text{VAR}[X(t)] = E[(X(t) - m)^2] = \sigma^2 \qquad \text{for all } t. \tag{9.58}$$

The *second-order cdf of a stationary random process can depend only on the time difference between the samples* and not on the particular time of the samples, since by Eq. (9.55),

$$F_{X(t_1), X(t_2)}(x_1, x_2) = F_{X(0), X(t_2-t_1)}(x_1, x_2) \qquad \text{for all } t_1, t_2. \tag{9.59}$$

This implies that the autocorrelation and the autocovariance of $X(t)$ can depend only on $t_2 - t_1$:

$$R_X(t_1, t_2) = R_X(t_2 - t_1) \qquad \text{for all } t_1, t_2 \tag{9.60}$$

$$C_X(t_1, t_2) = C_X(t_2 - t_1) \qquad \text{for all } t_1, t_2. \tag{9.61}$$

Example 9.31   iid Random Process

Show that the iid random process is stationary.

The joint cdf for the samples at any k time instants, $t_1, \ldots, t_k$, is

$$F_{X(t_1),\ldots,X(t_k)}(x_1, x_2,\ldots, x_k) = F_X(x_1)F_X(x_2)\ldots F_X(x_k)$$

$$= F_{X(t_1+\tau),\ldots,X(t_k+\tau)}(x_1,\ldots, x_k),$$

for all k, $t_1,\ldots, t_k$. Thus Eq. (9.55) is satisfied, and so the iid random process is stationary.

Example 9.32

Is the sum process a discrete-time stationary process?

The sum process is defined by $S_n = X_1 + X_2 + \cdots + X_n$, where the X_i are an iid sequence. The process has mean and variance

$$m_S(n) = nm \qquad \text{VAR}[S_n] = n\sigma^2,$$

where m and σ^2 are the mean and variance of the X_n. It can be seen that the mean and variance are not constant but grow linearly with the time index n. Therefore the sum process cannot be a stationary process.

Example 9.33   Random Telegraph Signal

Show that the random telegraph signal discussed in Example 9.24 is a stationary random process when $P[X(0) = \pm 1] = 1/2$. Show that $X(t)$ settles into stationary behavior as $t \to \infty$ even if $P[X(0) = \pm 1] \neq 1/2$.

We need to show that the following two joint pmf's are equal:

$$P[X(t_1) = a_1,\ldots, X(t_k) = a_k] = P[X(t_1 + \tau) = a_1,\ldots, X(t_k + \tau) = a_k],$$

for any k, any $t_1 < \cdots < t_k$, and any $a_j = \pm 1$. The independent increments property of the Poisson process implies that

$$P[X(t_1) = a_1,\ldots, X(t_k) = a_k] = P[X(t_1) = a_1]$$
$$\times P[X(t_2) = a_2 \mid X(t_1) = a_1]\ldots P[X(t_k) = a_k \mid X(t_{k-1}) = a_{k-1}],$$

since the values of the random telegraph at the times $t_1,\ldots, t_k$ are determined by the number of occurrences of events of the Poisson process in the time intervals (t_j, t_{j+1}). Similarly,

$$P[X(t_1 + \tau) = a_1,\ldots, X(t_k + \tau) = a_k]$$
$$= P[X(t_1 + \tau) = a_1]P[X(t_2 + \tau) = a_2 \mid X(t_1 + \tau) = a_1]\ldots$$
$$\times P[X(t_k + \tau) = a_k \mid X(t_{k-1} + \tau) = a_{k-1}].$$

The corresponding transition probabilities in the previous two equations are equal since

$$P[X(t_{j+1}) = a_{j+1} \mid X(t_j) = a_j] = \begin{cases} \dfrac{1}{2}\{1 + e^{-2\alpha(t_{j+1}-t_j)}\} & \text{if } a_j = a_{j+1} \\[2mm] \dfrac{1}{2}\{1 - e^{-2\alpha(t_{j+1}-t_j)}\} & \text{if } a_j \neq a_{j+1} \end{cases}$$

$$= P[X(t_{j+1} + \tau) = a_{j+1} \mid X(t_j + \tau) = a_j].$$

Thus the two joint probabilities differ only in the first term, namely, $P[X(t_1) = a_1]$ and $P[X(t_1 + \tau) = a_1]$.

From Example 9.24 we know that if $P[X(0) = \pm 1] = 1/2$ then $P[X(t) = \pm 1] = 1/2$, for all t. Thus $P[X(t_1) = a_1] = 1/2$, $P[X(t_1 + \tau) = a_1] = 1/2$, and

$$P[X(t_1) = a_1, \ldots, X(t_k) = a_k] = P[X(t_1 + \tau) = a_1, \ldots, X(t_k + \tau) = a_k].$$

Thus we conclude that the process is stationary when $P[X(0) = \pm 1] = 1/2$.

If $P[X(0) = \pm 1] \neq 1/2$, then the two joint pmf's are not equal because $P[X(t_1) = a_1] \neq P[X(t_1 + \tau) = a_1]$. Let's see what happens if we know that the process started at a specific value, say $X(0) = 1$, that is, $P[X(0) = 1] = 1$. The pmf for $X(t)$ is obtained from Eqs. (9.39) through (9.41):

$$P[X(t) = a] = P[X(t) = a \mid X(0) = 1]1$$

$$= \begin{cases} \dfrac{1}{2}\{1 + e^{-2\alpha t}\} & \text{if } a = 1 \\[2mm] \dfrac{1}{2}\{1 - e^{-2\alpha t}\} & \text{if } a = -1. \end{cases}$$

For very small t, the probability that $X(t) = 1$ is close to 1; but as t increases, the probability that $X(t) = 1$ becomes 1/2. Therefore as t_1 becomes large, $P[X(t_1) = a_1] \to 1/2$ and $P[X(t_1 + \tau) = a_1] \to 1/2$ and the two joint pmf's become equal. In other words, the process "forgets" the initial condition and settles down into "steady state," that is, stationary behavior.

9.6.1 Wide-Sense Stationary Random Processes

In many situations we cannot determine whether a random process is stationary, but we can determine whether the mean is a constant:

$$m_X(t) = m \qquad \text{for all } t, \tag{9.62}$$

and whether the autocovariance (or equivalently the autocorrelation) is a function of $t_1 - t_2$ only:

$$C_X(t_1, t_2) = C_X(t_1 - t_2) \qquad \text{for all } t_1, t_2. \tag{9.63}$$

*A discrete-time or continuous-time random process X(t) is **wide-sense stationary** (WSS) if it satisfies Eqs. (9.62) and (9.63).* Similarly, we say that the processes $X(t)$ and $Y(t)$ are **jointly wide-sense stationary** if they are both wide-sense stationary and if their cross-covariance depends only on $t_1 - t_2$. When $X(t)$ is wide-sense stationary, we write

$$C_X(t_1, t_2) = C_X(\tau) \qquad \text{and} \qquad R_X(t_1, t_2) = R_X(\tau),$$

where $\tau = t_1 - t_2$.

All stationary random processes are wide-sense stationary since they satisfy Eqs. (9.62) and (9.63). The following example shows that some wide-sense stationary processes are not stationary.

Example 9.34

Let X_n consist of two interleaved sequences of independent random variables. For n even, X_n assumes the values ± 1 with probability 1/2; for n odd, X_n assumes the values 1/3 and -3 with

probabilities 9/10 and 1/10, respectively. X_n is not stationary since its pmf varies with n. It is easy to show that X_n has mean

$$m_X(n) = 0 \qquad \text{for all } n$$

and covariance function

$$C_X(i, j) = \begin{cases} E[X_i]E[X_j] = 0 & \text{for } i \neq j \\ E[X_i^2] = 1 & \text{for } i = j. \end{cases}$$

X_n is therefore wide-sense stationary.

We will see in Chapter 10 that the autocorrelation function of wide-sense stationary processes plays a crucial role in the design of linear signal processing algorithms. We now develop several results that enable us to deduce properties of a WSS process from properties of its autocorrelation function.

First, *the autocorrelation function at $\tau = 0$ gives the* **average power** (second moment) of the process:

$$R_X(0) = E[X(t)^2] \qquad \text{for all } t. \tag{9.64}$$

Second, *the autocorrelation function is an even function of τ* since

$$R_X(\tau) = E[X(t + \tau)X(t)] = E[X(t)X(t + \tau)] = R_X(-\tau). \tag{9.65}$$

Third, *the autocorrelation function is a measure of the rate of change of a random process* in the following sense. Consider the change in the process from time t to $t + \tau$:

$$\begin{aligned} P[|X(t + \tau) - X(t)| > \varepsilon] &= P[(X(t + \tau) - X(t))^2 > \varepsilon^2] \\ &\leq \frac{E[(X(t + \tau) - X(t))^2]}{\varepsilon^2} \\ &= \frac{2\{R_X(0) - R_X(\tau)\}}{\varepsilon^2}, \end{aligned} \tag{9.66}$$

where we used the Markov inequality, Eq. (4.75), to obtain the upper bound. Equation (9.66) states that if $R_X(0) - R_X(\tau)$ is small, that is, $R_X(\tau)$ drops off slowly, then the probability of a large change in $X(t)$ in τ seconds is small.

Fourth, *the autocorrelation function is maximum at $\tau = 0$.* We use the Cauchy-Schwarz inequality:[2]

$$E[XY]^2 \leq E[X^2]E[Y^2], \tag{9.67}$$

for any two random variables X and Y. If we apply this equation to $X(t + \tau)$ and $X(t)$, we obtain

$$R_X(\tau)^2 = E[X(t + \tau)X(t)]^2 \leq E[X^2(t + \tau)]E[X^2(t)] = R_X(0)^2.$$

Thus

$$|R_X(\tau)| \leq R_X(0). \tag{9.68}$$

---

[2]See Problem 5.74 and Appendix C.

Fifth, *if $R_X(0) = R_X(d)$, then $R_X(\tau)$ is periodic with period d and $X(t)$ is* **mean square periodic**, that is, $E[(X(t + d) - X(t))^2] = 0$. If we apply Eq. (9.67) to $X(t + \tau + d) - X(t + \tau)$ and $X(t)$, we obtain

$$E[(X(t + \tau + d) - X(t + \tau))X(t)]^2$$
$$\leq E[(X(t + \tau + d) - X(t + \tau))^2]E[X^2(t)],$$

which implies that

$$\{R_X(\tau + d) - R_X(\tau)\}^2 \leq 2\{R_X(0) - R_X(d)\}R_X(0).$$

Thus $R_X(d) = R_X(0)$ implies that the right-hand side of the equation is zero, and thus that $R_X(\tau + d) = R_X(\tau)$ for all τ. Repeated applications of this result imply that $R_X(\tau)$ is periodic with period d. The fact that $X(t)$ is mean square periodic follows from

$$E[(X(t + d) - X(t))^2] = 2\{R_X(0) - R_X(d)\} = 0.$$

Sixth, let $X(t) = m + N(t)$, where $N(t)$ is a zero-mean process for which $R_N(\tau) \to 0$ as $\tau \to \infty$, then

$$R_X(\tau) = E[(m + N(t + \tau))(m + N(t))] = m^2 + 2mE[N(t)] + R_N(\tau)$$
$$= m^2 + R_N(\tau) \to m^2 \qquad \text{as } \tau \to \infty.$$

In other words, $R_X(\tau)$ *approaches the square of the mean of $X(t)$ as $\tau \to \infty$*.

In summary, the autocorrelation function can have three types of components: (1) a component that approaches zero as $\tau \to \infty$; (2) a periodic component; and (3) a component due to a nonzero mean.

Example 9.35

Figure 9.13 shows several typical autocorrelation functions. Figure 9.13(a) shows the autocorrelation function for the random telegraph signal $X(t)$ (see Eq. (9.44)):

$$R_X(\tau) = e^{-2\alpha|\tau|} \qquad \text{for all } \tau.$$

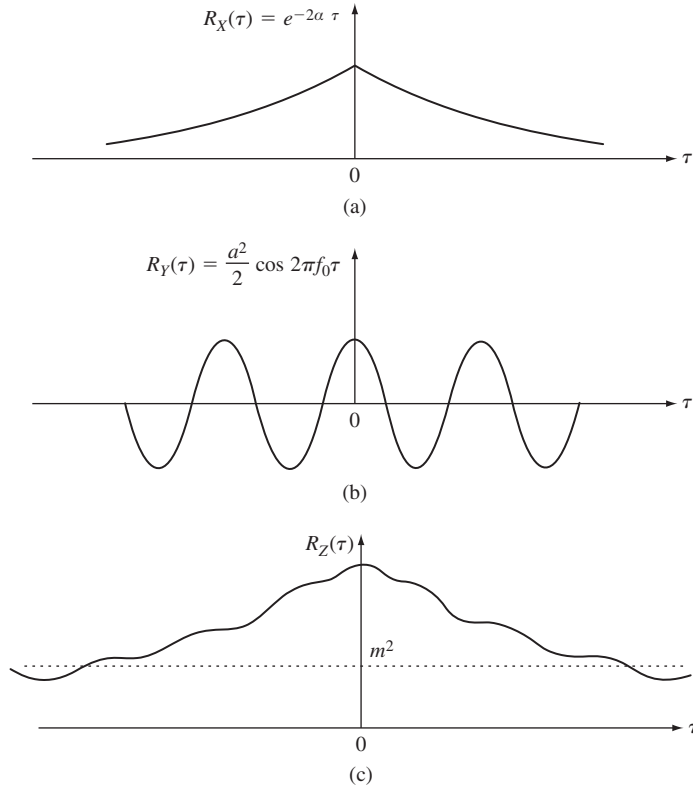$X(t)$ is zero mean and $R_X(\tau) \to 0$ as $|\tau| \to \infty$.

Figure 9.13(b) shows the autocorrelation function for a sinusoid $Y(t)$ with amplitude a and random phase (see Example 9.10):

$$R_Y(\tau) = \frac{a^2}{2}\cos(2\pi f_0 \tau) \qquad \text{for all } \tau.$$

$Y(t)$ is zero mean and $R_Y(\tau)$ is periodic with period $1/f_0$.

Figure 9.13(c) shows the autocorrelation function for the process $Z(t) = X(t) + Y(t) + m$, where $X(t)$ is the random telegraph process, $Y(t)$ is a sinusoid with random phase, and m is a constant. If we assume that $X(t)$ and $Y(t)$ are independent processes, then

$$R_Z(\tau) = E[\{X(t + \tau) + Y(t + \tau) + m\}\{X(t) + Y(t) + m\}]$$
$$= R_X(\tau) + R_Y(\tau) + m^2.$$

$$R_X(\tau) = e^{-2\alpha\,\tau}$$

(a)

$$R_Y(\tau) = \frac{a^2}{2} \cos 2\pi f_0 \tau$$

(b)

$$R_Z(\tau)$$

m^2

(c)

FIGURE 9.13
(a) Autocorrelation function of a random telegraph signal. (b) Autocorrelation function of a sinusoid with random phase. (c) Autocorrelation function of a random process that has nonzero mean, a periodic component, and a "random" component.

9.6.2 Wide-Sense Stationary Gaussian Random Processes

If a Gaussian random process is wide-sense stationary, then it is also stationary. Recall from Section 9.5, Eq. (9.47), that the joint pdf of a Gaussian random process is completely determined by the mean $m_X(t)$ and the autocovariance $C_X(t_1, t_2)$. If $X(t)$ is wide-sense stationary, then its mean is a constant m and its autocovariance depends only on the difference of the sampling times, $t_i - t_j$. It then follows that the joint pdf of $X(t)$ depends only on this set of differences, and hence it is invariant with respect to time shifts. Thus the process is also stationary.

The above result makes WSS Gaussian random processes particularly easy to work with since all the information required to specify the joint pdf is contained in m and $C_X(\tau)$.

Example 9.36 A Gaussian Moving Average Process

Let X_n be an iid sequence of Gaussian random variables with zero mean and variance σ^2, and let Y_n be the average of two consecutive values of X_n:

$$Y_n = \frac{X_n + X_{n-1}}{2}.$$

The mean of Y_n is zero since $E[X_i] = 0$ for all i. The covariance is

$$C_Y(i, j) = E[Y_i Y_j] = \frac{1}{4} E[(X_i + X_{i-1})(X_j + X_{j-1})]$$

$$= \frac{1}{4} \{ E[X_i X_j] + E[X_i X_{j-1}] + E[X_{i-1} X_j] + E[X_{i-1} X_{j-1}] \}$$

$$= \begin{cases} \frac{1}{2}\sigma^2 & \text{if } i = j \\ \frac{1}{4}\sigma^2 & \text{if } |i - j| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

We see that Y_n has a constant mean and a covariance function that depends only on $|i - j|$, thus Y_n is a wide-sense stationary process. Y_n is a Gaussian random variable since it is defined by a linear function of Gaussian random variables (see Section 6.4, Eq. 6.45). Thus the joint pdf of Y_n is given by Eq. (9.47) with zero-mean vector and with entries of the covariance matrix specified by $C_Y(i, j)$ above.

9.6.3   Cyclostationary Random Processes

Many random processes arise from the repetition of a given procedure every T seconds. For example, a data modulator ("modem") produces a waveform every T seconds according to some input data sequence. In another example, a "time multiplexer" interleaves n separate sequences of information symbols into a single sequence of symbols. It should not be surprising that the periodic nature of such processes is evident in their probabilistic descriptions. A discrete-time or continuous-time random process $X(t)$ is said to be **cyclostationary** if the joint cumulative distribution function of any set of samples is invariant with respect to shifts of the origin by *integer multiples of some period T*. In other words, $X(t_1), X(t_2), \ldots, X(t_k)$ and $X(t_1 + mT), X(t_2 + mT), \ldots, X(t_k + mT)$ have the same joint cdf for all k, m, and all choices of sampling times $t_1, \ldots, t_k$:

$$F_{X(t_1), X(t_2), \ldots, X(t_k)}(x_1, x_2, \ldots, x_k)$$
$$= F_{X(t_1+mT), X(t_2+mT), \ldots, X(t_k+mT)}(x_1, x_2, \ldots, x_k). \qquad (9.69)$$

We say that $X(t)$ is **wide-sense cyclostationary** if the mean and autocovariance functions are invariant with respect to shifts in the time origin by integer multiples of T, that is, for every integer m,

$$m_X(t + mT) = m_X(t) \qquad (9.70\text{a})$$

$$C_X(t_1 + mT, t_2 + mT) = C_X(t_1, t_2). \qquad (9.70\text{b})$$

Note that if $X(t)$ is cyclostationary, then it follows that $X(t)$ is also wide-sense cyclostationary.

Example 9.37

Consider a random amplitude sinusoid with period T:

$$X(t) = A \cos(2\pi t/T).$$

Is $X(t)$ cyclostationary? wide-sense cyclostationary?

Consider the joint cdf for the time samples $t_1, \ldots, t_k$:

$$P[X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_k) \leq x_k)]$$
$$= P[A \cos(2\pi t_1/T) \leq x_1, \ldots, A \cos(2\pi t_k/T) \leq x_k]$$
$$= P[A \cos(2\pi(t_1 + mT)/T) \leq x_1, \ldots, A \cos(2\pi(t_k + mT)/T) \leq x_k]$$
$$= P[X(t_1 + mT) \leq x_1, X(t_2 + mT) \leq x_2, \ldots, X(t_k + mT) \leq x_k].$$

Thus $X(t)$ is a cyclostationary random process and hence also a wide-sense cyclostationary process.

In the above example, the sample functions of the random process are always periodic. The following example shows that, in general, the sample functions of a cyclostationary random process need not be periodic.

Example 9.38   Pulse Amplitude Modulation

A modem transmits a binary iid equiprobable data sequence as follows: To transmit a binary 1, the modem transmits a rectangular pulse of duration T seconds and amplitude 1; to transmit a binary 0, it transmits a rectangular pulse of duration T seconds and amplitude -1. Let $X(t)$ be the random process that results. Is $X(t)$ wide-sense cyclostationary?

Figure 9.14(a) shows a rectangular pulse of duration T seconds, and Fig. 9.14(b) shows the waveform that results for a particular data sequence. Let A_i be the sequence of amplitudes (± 1)

(a) Individual signal pulse

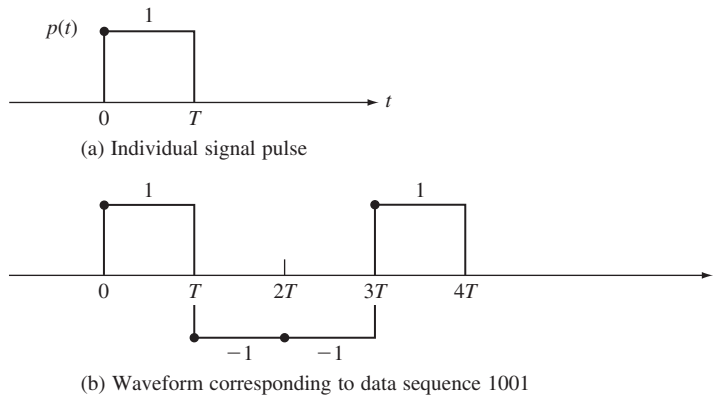(b) Waveform corresponding to data sequence 1001

FIGURE 9.14
Pulse amplitude modulation.

corresponding to the binary sequence, then $X(t)$ can be represented as the sum of amplitude-modulated time-shifted rectangular pulses:

$$X(t) = \sum_{n=-\infty}^{\infty} A_n p(t - nT). \tag{9.71}$$

The mean of $X(t)$ is

$$m_X(t) = E\left[\sum_{n=-\infty}^{\infty} A_n p(t - nT)\right] = \sum_{n=-\infty}^{\infty} E[A_n]p(t - nT) = 0$$

since $E[A_n] = 0$. The autocovariance function is

$$
\begin{aligned}
C_X(t_1, t_2) &= E[X(t_1)X(t_2)] - 0 \\
&= \begin{cases} E[X(t_1)^2] = 1 & \text{if } nT \le t_1, t_2 < (n+1)T \\ E[X(t_1)]E[X(t_2)] = 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

Figure 9.15 shows the autocovariance function in terms of t_1 and t_2. It is clear that $C_X(t_1 + mT, t_2 + mT) = C_X(t_1, t_2)$ for all integers m. Therefore the process is wide-sense cyclostationary.

We will now show how a stationary random process can be obtained from a cyclostationary process. Let $X(t)$ be a cyclostationary process with period T. We "stationarize" $X(t)$ by observing a randomly phase-shifted version of $X(t)$:

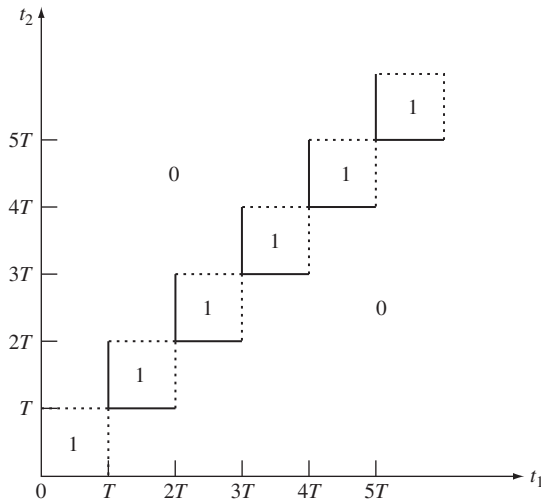$$X_s(t) = X(t + \Theta) \qquad \Theta \text{ uniform in } [0, T], \tag{9.72}$$

FIGURE 9.15
Autocovariance function of pulse amplitude-modulated
random process.

where Θ is independent of $X(t)$. $X_s(t)$ can arise when the phase of $X(t)$ is either unknown or not of interest. *If $X(t)$ is a cyclostationary random process, then $X_s(t)$ is a stationary random process.* To show this, we first use conditional expectation to find the joint cdf of $X_s(t)$:

$$P[X_s(t_1) \leq x_1, X_s(t_2) \leq x_2, \ldots, X_s(t_k) \leq x_k]$$
$$= P[X(t_1 + \Theta) \leq x_1, X(t_2 + \Theta) \leq x_2, \ldots, X(t_k + \Theta) \leq x_k]$$
$$= \int_0^T P[X(t_1 + \Theta) \leq x_1, \ldots, X(t_k + \Theta) \leq x_k | \Theta = \theta] f_\Theta(\theta) \, d\theta$$
$$= \frac{1}{T} \int_0^T P[X(t_1 + \theta) \leq x_1, \ldots, X(t_k + \theta) \leq x_k] \, d\theta. \tag{9.73}$$

Equation (9.73) shows that the joint cdf of $X_s(t)$ is obtained by integrating the joint cdf of $X(t)$ over one time period. It is easy to then show that a time-shifted version of $X_s(t)$, say $X_s(t_1 + \tau), X_s(t_2 + \tau), \ldots, X_s(t_k + \tau)$, will have the same joint cdf as $X_s(t_1)$, $X_s(t_2), \ldots, X_s(t_k)$ (see Problem 9.80). Therefore $X_s(t)$ is a stationary random process.

By using conditional expectation (see Problem 9.81), it is easy to show that *if $X(t)$ is a wide-sense cyclostationary random process, then $X_s(t)$ is a wide-sense stationary random process*, with mean and autocorrelation given by

$$E[X_s(t)] = \frac{1}{T} \int_0^T m_X(t) \, dt \tag{9.74a}$$

$$R_{X_s}(\tau) = \frac{1}{T} \int_0^T R_X(t + \tau, t) \, dt. \tag{9.74b}$$

Example 9.39   Pulse Amplitude Modulation with Random Phase Shift

Let $X_s(t)$ be the phase-shifted version of the pulse amplitude–modulated waveform $X(t)$ introduced in Example 9.38. Find the mean and autocorrelation function of $X_s(t)$.

$X_s(t)$ has zero mean since $X(t)$ is zero-mean. The autocorrelation of $X_s(t)$ is obtained from Eq. (9.74b). From Fig. 9.15, we can see that for $0 < t + \tau < T$, $R_X(t + \tau, t) = 1$ and $R_X(t + \tau, t) = 0$ otherwise. Therefore:

$$\text{for } 0 < \tau < T: \quad R_{X_s}(\tau) = \frac{1}{T} \int_0^{T-\tau} dt = \frac{T - \tau}{T};$$

$$\text{for } -T < \tau < 0: \quad R_{X_s}(\tau) = \frac{1}{T} \int_{-\tau}^T dt = \frac{T + \tau}{T}.$$

Thus $X_s(t)$ has a triangular autocorrelation function:

$$R_{X_s}(\tau) = \begin{cases} 1 - \dfrac{|\tau|}{T} & |\tau| \leq T \\ 0 & |\tau| > T. \end{cases}$$

The variance is then

$$\text{VAR}[M(t)] = E[A^2]\frac{4}{\pi^2}\sin^2\frac{2\pi t}{T} - E[A]^2\frac{4}{\pi^2}\sin^2\frac{2\pi t}{T}$$

$$= \text{VAR}[A]\frac{4}{\pi^2}\sin^2\frac{2\pi t}{T}.$$

Example 9.45   Integral of White Gaussian Noise

Let $Z(t)$ be the white Gaussian noise process introduced in Example 9.43. Find the autocorrelation function of $X(t)$, the integral of $Z(t)$ over the interval $(0, t)$.

   From Example 9.43, the white Gaussian noise process has autocorrelation function

$$R_Z(t_1, t_2) = \alpha\delta(t_1 - t_2).$$

The autocorrelation function of $X(t)$ is then given by

$$R_X(t_1, t_2) = \int_0^{t_1}\int_0^{t_2}\alpha\delta(w - v)\,dw\,dv = \alpha\int_0^{t_2}u(t_1 - v)\,dv$$

$$= \alpha\int_0^{\min(t_1, t_2)}dv = \alpha\min(t_1, t_2).$$

We thus find that $X(t)$ has the same autocorrelation as the Wiener process. In addition we have that $X(t)$ must be a Gaussian random process since $Z(t)$ is Gaussian. It then follows that $X(t)$ must be the Wiener process because it has the joint pdf given by Eq. (9.52).

9.7.4   Response of a Linear System to Random Input

We now apply the results developed in this section to develop the solution of a linear system described by a first-order differential equation. The method can be generalized to higher-order equations. In the next chapter we develop transform methods to solve the general problem.

   Consider a linear system described by the first-order differential equation:

$$X'(t) + \alpha X(t) = Z(t) \qquad t \geq 0, X(0) = 0. \tag{9.93}$$

For example, $X(t)$ may represent the voltage across the capacitor of an RC circuit with current input $Z(t)$. We now show how to obtain $m_X(t)$ and $R_X(t_1, t_2)$. If the input process $Z(t)$ is Gaussian, then the output process will also be Gaussian. Therefore, in the case of Gaussian input processes, we can then characterize the joint pdf of the output process.

We obtain a differential equation for $m_X(t)$ by taking the expected value of Eq. (9.93):

$$E[X'(t)] + E[X(t)] = m'_X(t) + m_X(t) = m_Z(t) \qquad t \geq 0 \qquad (9.94)$$

with initial condition $m_X(0) = E[X(0)] = 0$.

As an intermediate step we next find a differential equation for $R_{Z,X}(t_1, t_2)$. If we multiply Eq. (9.93) by $Z(t_1)$ and take the expected value, we obtain

$$E[Z(t_1)X'(t_2)] + \alpha E[Z(t_1)X(t_2)] = E[Z(t_1)Z(t_2)] \qquad t_2 \geq 0$$

with initial condition $E[Z(t_1)X(0)] = 0$ since $X(0) = 0$. The same derivation that led to the cross-correlation between $X(t)$ and $X'(t)$ (see Eq. 9.83) can be used to show that

$$E[Z(t_1)X'(t_2)] = \frac{\partial}{\partial t_2} R_{Z,X}(t_1, t_2).$$

Thus we obtain the following differential equation:

$$\frac{\partial}{\partial t_2} R_{Z,X}(t_1, t_2) + \alpha R_{Z,X}(t_1, t_2) = R_Z(t_1, t_2) \qquad t_2 \geq 0 \qquad (9.95)$$

with initial condition $R_{Z,X}(t_1, 0) = 0$.

Finally we obtain a differential equation for $R_Z(t_1, t_2)$. Multiply Eq. (9.93) by $X(t_2)$ and take the expected value:

$$E[X'(t_1)X(t_2)] + \alpha E[X(t_1)X(t_2)] = E[Z(t_1)X(t_2)] \qquad t_1 \geq 0$$

with initial condition $E[X(0)X(t_2)] = 0$. This leads to the differential equation

$$\frac{\partial}{\partial t_1} R_X(t_1, t_2) + \alpha R_X(t_1, t_2) = R_{Z,X}(t_1, t_2) \qquad t_1 \geq 0 \qquad (9.96)$$

with initial condition $R_{Z,X}(0, t_2) = 0$. Note that the solution to Eq. (9.95) appears as the forcing function in Eq. (9.96). Thus we conclude that by solving the differential equations in Eqs. (9.94), (9.95), and (9.96) we obtain the mean and autocorrelation function for $X(t)$.

Example 9.46   Ornstein-Uhlenbeck Process

Equation (9.93) with the input given by a zero-mean, white Gaussian noise process is called the *Langevin equation*, after the scientist who formulated it in 1908 to describe the Brownian motion of a free particle. In this formulation $X(t)$ represents the velocity of the particle, so that Eq. (9.93) results from equating the acceleration of the particle $X'(t)$ to the force on the particle due to friction $-\alpha X(t)$ and the force due to random collisions $Z(t)$. We present the solution developed by Uhlenbeck and Ornstein in 1930.

First, we note that since the input process $Z(t)$ is Gaussian, the output process $X(t)$ will also be a Gaussian random process. Next we recall that the first-order differential equation

$$x'(t) + ax(t) = g(t) \qquad t \geq 0, x(0) = 0$$

has solution

$$x(t) = \int_0^t e^{-a(t-\tau)} g(\tau) \, d\tau \qquad t \geq 0.$$

Therefore the solution to Eq. (9.94) is

$$m_X(t) = \int_0^t e^{-\alpha(t-\tau)} m_Z(\tau) \, d\tau = 0.$$

The autocorrelation of the white Gaussian noise process is

$$R_Z(t_1, t_2) = \sigma^2 \delta(t_1 - t_2).$$

Equation (9.95) is also a first-order differential equation, and it has solution

$$R_{Z,X}(t_1, t_2) = \int_0^{t_2} e^{-\alpha(t_2-\tau)} R_Z(t_1, \tau) \, d\tau$$

$$= \int_0^{t_2} e^{-\alpha(t_2-\tau)} \sigma^2 \delta(t_1 - \tau) \, d\tau$$

$$= \begin{cases} 0 & 0 \leq t_2 < t_1 \\ \sigma^2 e^{-\alpha(t_2-t_1)} & t_2 \geq t_1 \end{cases}$$

$$= \sigma^2 e^{-\alpha(t_2-t_1)} u(t_2 - t_1),$$

where $u(x)$ is the unit step function.

The autocorrelation function of the output process $X(t)$ is the solution to the first-order differential equation Eq. (9.96). The solution is given by

$$R_X(t_1, t_2) = \int_0^{t_1} e^{-\alpha(t_1-\tau)} R_{Z,X}(\tau, t_2) \, d\tau$$

$$= \sigma^2 \int_0^{t_1} e^{-\alpha(t_1-\tau)} e^{-\alpha(t_2-\tau)} u(t_2 - \tau) \, d\tau$$

$$= \sigma^2 \int_0^{\min(t_1, t_2)} e^{-\alpha(t_1-\tau)} e^{-\alpha(t_2-\tau)} \, d\tau$$

$$= \frac{\sigma^2}{2\alpha} (e^{-\alpha|t_1-t_2|} - e^{-\alpha(t_1+t_2)}) \qquad t_1 \geq 0, t_2 \geq 0. \qquad (9.97a)$$

A Gaussian random process with this autocorrelation function is called an **Ornstein-Uhlenbeck process**. Thus we conclude that the output process $X(t)$ is an Ornstein-Uhlenbeck process.

If we let $t_1 = t$ and $t_2 = t + \tau$, then as t approaches infinity,

$$R_X(t + \tau, t) \rightarrow \frac{\sigma^2}{2\alpha} e^{-\alpha|\tau|}. \qquad (9.97b)$$

This shows that the effect of the zero initial condition dies out as time progresses, and the process becomes wide-sense stationary. Since the process is Gaussian, this also implies that the process becomes strict-sense stationary.

9.8 TIME AVERAGES OF RANDOM PROCESSES AND ERGODIC THEOREMS

At some point, the parameters of a random process must be obtained through measurement. The results from Chapter 7 and the statistical methods of Chapter 8 suggest that we repeat the random experiment that gives rise to the random process a large number of times and take the arithmetic average of the quantities of interest. For example, to estimate the mean $m_X(t)$ of a random process $X(t, \zeta)$, we repeat the random experiment and take the following average:

$$\hat{m}_X(t) = \frac{1}{N} \sum_{i=1}^{N} X(t, \zeta_i), \tag{9.98}$$

where N is the number of repetitions of the experiment, and $X(t, \zeta_i)$ is the realization observed in the ith repetition.

In some situations, we are interested in estimating the mean or autocorrelation functions from the **time average** of a single realization, that is,

$$\langle X(t) \rangle_T = \frac{1}{2T} \int_{-T}^{T} X(t, \zeta)\, dt. \tag{9.99}$$

An **ergodic theorem** states conditions under which a time average converges as the observation interval becomes large. In this section, we are interested in ergodic theorems that state when time averages converge to the ensemble average (expected value).

The strong law of large numbers, presented in Chapter 7, is one of the most important ergodic theorems. It states that if X_n is an iid discrete-time random process with finite mean $E[X_n] = m$, then the time average of the samples converges to the ensemble average with probability one:

$$P\left[\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = m \right] = 1. \tag{9.100}$$

This result allows us to estimate m by taking the time average of a single realization of the process. We are interested in obtaining results of this type for a larger class of random processes, that is, for non-iid, discrete-time random processes, and for continuous-time random processes.

The following example shows that, in general, time averages do not converge to ensemble averages.

Example 9.47

Let $X(t) = A$ for all t, where A is a zero-mean, unit-variance random variable. Find the limiting value of the time average.

The mean of the process is $m_X(t) = E[X(t)] = E[A] = 0$. However, Eq. (9.99) gives

$$\langle X(t) \rangle_T = \frac{1}{2T} \int_{-T}^{T} A\, dt = A.$$

Thus the time-average mean does not always converge to $m_X(t) = 0$. Note that this process is stationary. Thus this example shows that stationary processes need not be ergodic.

Consider the estimate given by Eq. (9.99) for $E[X(t)] = m_X(t)$. The estimate yields a single number, so obviously it only makes sense to consider processes for which $m_X(t) = m$, a constant. We now develop an ergodic theorem for the time average of wide-sense stationary processes.

Let $X(t)$ be a WSS process. The expected value of $\langle X(t) \rangle_T$ is

$$E[\langle X(t) \rangle_T] = E\left[\frac{1}{2T} \int_{-T}^{T} X(t)\, dt\right] = \frac{1}{2T} \int_{-T}^{T} E[X(t)]\, dt = m. \qquad (9.101)$$

Equation (9.101) states that $\langle X(t) \rangle_T$ is an unbiased estimator for m.

Consider the variance of $\langle X(t) \rangle_T$:

$$\begin{aligned}
\mathrm{VAR}[\langle X(t) \rangle_T] &= E[(\langle X(t) \rangle_T - m)^2] \\
&= E\left[\left\{\frac{1}{2T} \int_{-T}^{T} (X(t) - m)\, dt\right\}\left\{\frac{1}{2T} \int_{-T}^{T} (X(t') - m)\, dt'\right\}\right] \\
&= \frac{1}{4T^2} \int_{-T}^{T} \int_{-T}^{T} E[(X(t) - m)(X(t') - m)]\, dt\, dt' \\
&= \frac{1}{4T^2} \int_{-T}^{T} \int_{-T}^{T} C_X(t, t')\, dt\, dt'. \qquad (9.102)
\end{aligned}$$

Since the process $X(t)$ is WSS, Eq. (9.102) becomes

$$\mathrm{VAR}[\langle X(t) \rangle_T] = \frac{1}{4T^2} \int_{-T}^{T} \int_{-T}^{T} C_X(t - t')\, dt\, dt'. \qquad (9.103)$$

Figure 9.17 shows the region of integration for this integral. The integrand is constant along the line $u = t - t'$ for $-2T < u < 2T$, so we can evaluate the integral as the
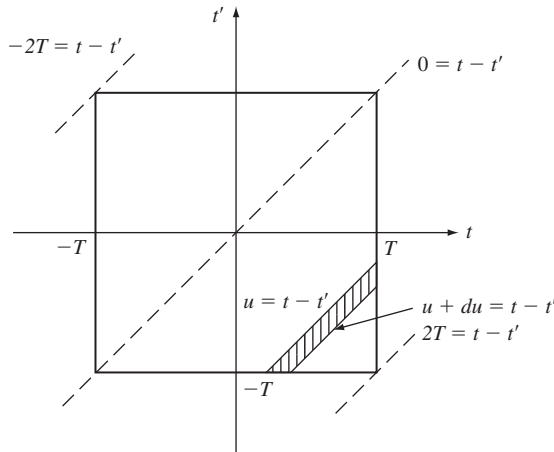

FIGURE 9.17
Region of integration for integral in Eq. (9.102).

sums of infinitesimal strips as shown in the figure. It can be shown that each strip has area $(2T - |u|)\,du$, so the contribution of each strip to the integral is $(2T - |u|)C_X(u)\,du$. Thus

$$\text{VAR}[\langle X(t)\rangle_T] = \frac{1}{4T^2}\int_{-2T}^{2T}(2T - |u|)C_X(u)\,du$$

$$= \frac{1}{2T}\int_{-2T}^{2T}\left(1 - \frac{|u|}{2T}\right)C_X(u)\,du. \qquad (9.104)$$

Therefore, $\langle X(t)\rangle_T$ will approach m in the mean square sense, that is, $E[(\langle X(t)\rangle_T - m)^2] \to 0$, if the expression in Eq. (9.104) approaches zero with increasing T. We have just proved the following ergodic theorem.

Theorem

Let $X(t)$ be a WSS process with $m_X(t) = m$, then

$$\lim_{T\to\infty}\langle X(t)\rangle_T = m$$

in the mean square sense, if and only if

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-2T}^{2T}\left(1 - \frac{|u|}{2T}\right)C_X(u)\,du = 0.$$

In keeping with engineering usage, we say that a WSS process is **mean ergodic** if it satisfies the conditions of the above theorem.

The above theorem can be used to obtain ergodic theorems for the time average of other quantities. For example, if we replace $X(t)$ with $Y(t + \tau)Y(t)$ in Eq. (9.99), we obtain a time-average estimate for the autocorrelation function of the process $Y(t)$:

$$\langle Y(t + \tau)Y(t)\rangle_T = \frac{1}{2T}\int_{-T}^{T}Y(t + \tau)Y(t)\,dt. \qquad (9.105)$$

It is easily shown that $E[\langle Y(t + \tau)Y(t)\rangle_T] = R_Y(\tau)$ if $Y(t)$ is WSS. The above ergodic theorem then implies that the time-average autocorrelation converges to $R_Y(\tau)$ in the mean square sense if the term in Eq. (9.104) with $X(t)$ replaced by $Y(t)Y(t + \tau)$ converges to zero.

Example 9.48

Is the random telegraph process mean ergodic?

The covariance function for the random telegraph process is $C_X(\tau) = e^{-2\alpha|\tau|}$, so the variance of $\langle X(t)\rangle_T$ is

$$\text{VAR}[\langle X(t)\rangle_T] = \frac{2}{2T}\int_{0}^{2T}\left(1 - \frac{u}{2T}\right)e^{-2\alpha u}\,du$$

$$< \frac{1}{T}\int_{0}^{2T}e^{-2\alpha u}\,du = \frac{1 - e^{-4\alpha T}}{2\alpha T}.$$

The bound approaches zero as $T \to \infty$, so $\text{VAR}[\langle X(t)\rangle_T] \to 0$. Therefore the process is mean ergodic.

If the random process under consideration is discrete-time, then the time-average estimate for the mean and the autocorrelation functions of X_n are given by

$$\langle X_n \rangle_T = \frac{1}{2T+1} \sum_{n=-T}^{T} X_n \tag{9.106}$$

$$\langle X_{n+k} X_n \rangle_T = \frac{1}{2T+1} \sum_{n=-T}^{T} X_{n+k} X_n. \tag{9.107}$$

If X_n is a WSS random process, then $E[\langle X_n \rangle_T] = m$, and so $\langle X_n \rangle_T$ is an unbiased estimate for m. It is easy to show that the variance of $\langle X_n \rangle_T$ is

$$\text{VAR}[\langle X_n \rangle_T] = \frac{1}{2T+1} \sum_{k=-2T}^{2T} \left(1 - \frac{|k|}{2T+1}\right) C_X(k). \tag{9.108}$$

Therefore, $\langle X_n \rangle_T$ approaches m in the mean square sense and is mean ergodic if the expression in Eq. (9.108) approaches zero with increasing T.

Example 9.49   Ergodicity and Exponential Correlation

Let X_n be a wide-sense stationary discrete-time process with mean m and covariance function $C_X(k) = \sigma^2 \rho^{-|k|}$, for $|\rho| < 1$ and $k = 0, \pm 1, +2, \dots$. Show that X_n is mean ergodic.

The variance of the sample mean (Eq. 9.106) is:

$$\text{VAR}[\langle X_n \rangle_T = \frac{1}{2T+1} \sum_{k=-2T}^{2T} \left(1 - \frac{|k|}{2T+1}\right) \sigma^2 \rho^{|k|}$$

$$< \frac{2}{2T+1} \sum_{k=0}^{\infty} \sigma^2 \rho^k = \frac{2\sigma^2}{2T+1} \frac{1}{1-\rho}.$$

The bound on the right-hand side approaches zero as T increases and so X_n is mean ergodic.

Example 9.50   Ergodicity of Self-Similar Process and Long-Range Dependence

Let X_n be a wide-sense stationary discrete-time process with mean m and covariance function

$$C_X(k) = \frac{\sigma^2}{2} \{|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\} \tag{9.109}$$

for $1/2 < H < 1$ and $k = 0, \pm 1, +2, \dots$ X_n is said to be **second-order self-similar**. We will investigate the ergodicity of X_n.

We rewrite the variance of the sample mean in (Eq. 9.106) as follows:

$$\text{VAR}[\langle X_n \rangle_T] = \frac{1}{(2T+1)^2} \sum_{k=-2T}^{2T} (2T+1-|k|) C_X(k)$$

$$= \frac{1}{(2T+1)^2} \{(2T+1)C_X(0) + 2(2TC_X(1)) + \dots + 2C_X(2T)\}.$$

It is easy to show (See Problem 9.132) that the sum inside the braces is $\sigma^2(2T + 1)^{2H}$. Therefore the variance becomes:

$$\text{VAR}[\langle X_n \rangle_T] = \frac{1}{(2T + 1)^2} \sigma^2(2T + 1)^{2H} = \sigma^2(2T + 1)^{2H-2}. \tag{9.110}$$

The value of H, which is called the **Hurst parameter**, affects the convergence behavior of the sample mean. Note that if $H = 1/2$, the covariance function becomes $C_X(k) = 1/2\sigma^2\delta_k$ which corresponds to an iid sequence. In this case, the variance becomes $\sigma^2/(2T + 1)$ which is the convergence rate of the sample mean for iid samples. However, for $H > 1/2$, the variance becomes:

$$\text{VAR}[\langle X_n \rangle_T] = \frac{\sigma^2}{2T + 1}(2T + 1)^{2H-1}, \tag{9.111}$$

so the convergence of the sample mean is slower by a factor of $(2T + 1)^{2H-1}$ than for iid samples.

The slower convergence of the sample mean when $H > 1/2$ results from the long-range dependence of X_n. It can be shown that for large k, the covariance function is approximately given by:

$$C_X(k) = \sigma^2 H(2H - 1)k^{2H-2}. \tag{9.112}$$

For $1/2 < H < 1$, $C(k)$ decays as $1/k^\alpha$ where $0 < \alpha < 1$, which is a very slow decay rate. Thus the dependence between values of X_n decreases slowly and the process is said to have a long memory or long-range dependence.

*9.9   FOURIER SERIES AND KARHUNEN-LOEVE EXPANSION

Let $X(t)$ be a wide-sense stationary, mean square periodic random process with period T, that is, $E[(X(t + T) - X(t))^2] = 0$. In order to simplify the development, we assume that $X(t)$ is zero mean. We show that $X(t)$ can be represented in a mean square sense by a **Fourier series**:

$$X(t) = \sum_{k=-\infty}^{\infty} X_k e^{j2\pi kt/T}, \tag{9.113}$$

where the coefficients are random variables defined by

$$X_k = \frac{1}{T}\int_0^T X(t')e^{-j2\pi kt'/T}\, dt'. \tag{9.114}$$

Equation (9.114) implies that, in general, the coefficients are complex-valued random variables. *For complex-valued random variables, the correlation between two random variables X and Y is defined by* $E[XY^*]$. We also show that *the coefficients are orthogonal random variables*, that is, $E[X_k X_m^*] = 0$ for $k \neq m$.

Recall that if $X(t)$ is mean square periodic, then $R_X(\tau)$ is a periodic function in τ with period T. Therefore, it can be expanded in a Fourier series:

$$R_X(\tau) = \sum_{k=-\infty}^{\infty} a_k e^{j2\pi k\tau/T}, \tag{9.115}$$

where the coefficients a_k are given by

$$a_k = \frac{1}{T}\int_0^T R_X(t')e^{-j2\pi kt'/T}\, dt'. \tag{9.116}$$

The coefficients a_k appear in the following derivation.

First, we show that the coefficients in Eq. (9.113) are orthogonal random variables, that is, $E[X_k X_m^*] = 0$:

$$E[X_k X_m^*] = E\left[X_k \frac{1}{T} \int_0^T X^*(t') e^{j2\pi mt'/T} \, dt' \right]$$

$$= \frac{1}{T} \int_0^T E[X_k X^*(t')] e^{j2\pi mt'/T} \, dt'.$$

The integrand of the above equation has

$$E[X_k X^*(t)] = E\left[\frac{1}{T} \int_0^T X(u) e^{-j2\pi ku/T} \, du \, X^*(t) \right]$$

$$= \frac{1}{T} \int_0^T R_X(u - t) e^{-j2\pi ku/T} \, du$$

$$= \left\{ \frac{1}{T} \int_{-t}^{T-t} R_X(v) e^{-j2\pi kv/T} \, dv \right\} e^{-j2\pi kt/T}$$

$$= a_k e^{-j2\pi kt/T},$$

where we have used the fact that the Fourier coefficients can be calculated over any full period. Therefore

$$E[X_k X_m^*] = \frac{1}{T} \int_0^T a_k e^{-j2\pi kt'/T} e^{j2\pi mt'/T} \, dt' = a_k \delta_{k,m}, \tag{9.117}$$

where $\delta_{k,m}$ is the Kronecker delta function. Thus X_k and X_m are orthogonal random variables. Note that the above equation implies that $a_k = E\left[|X_k|^2\right]$, that is, the a_k are real-valued.

To show that the Fourier series equals $X(t)$ in the mean square sense, we take

$$E\left[\left| X(t) - \sum_{k=-\infty}^{\infty} X_k e^{j2\pi kt/T} \right|^2 \right]$$

$$= E[|X(t)|^2] - E\left[X(t) \sum_{k=-\infty}^{\infty} X_k^* e^{-j2\pi kt/T} \right]$$

$$- E\left[X^*(t) \sum_{k=-\infty}^{\infty} X_k e^{j2\pi kt/T} \right] + E\left[\sum_{k=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} X_k X_m^* e^{j2\pi(k-m)t/T} \right]$$

$$= R_X(0) - \sum_{k=-\infty}^{\infty} a_k - \sum_{k=-\infty}^{\infty} a_k^* + \sum_{k=-\infty}^{\infty} a_k.$$

The above equation equals zero, since the a_k are real and since $R_X(0) = \Sigma a_k$ from Eq. (9.115).

If $X(t)$ is a wide-sense stationary random process that is *not* mean square periodic, we can still expand $X(t)$ in the Fourier series in an arbitrary interval $[0, T]$. Mean square equality will hold only inside the interval. Outside the interval, the expansion repeats

itself with period T. The Fourier coefficients will no longer be orthogonal; instead they are given by

$$E[X_k X_m^*] = \frac{1}{T^2} \int_0^T \int_0^T R_X(t - u) e^{-j2\pi kt/T} e^{j2\pi mu/T} \, dt \, du. \tag{9.118}$$

It is easy to show that if $X(t)$ is mean square periodic, then this equation reduces to Eq. (9.117).

9.9.1   Karhunen-Loeve Expansion

In this section we present the **Karhunen-Loeve expansion**, which allows us to expand a (possibly nonstationary) random process $X(t)$ in a series:

$$X(t) = \sum_{k=1}^{\infty} X_k \phi_k(t) \qquad 0 \leq t \leq T, \tag{9.119a}$$

where

$$X_k = \int_0^T X(t) \phi_k^*(t) \, dt, \tag{9.119b}$$

where the equality in Eq. (9.119a) is in the mean square sense, where the coefficients $\{X_k\}$ are orthogonal random variables, and where the functions $\{\phi_k(t)\}$ are orthonormal:

$$\int_0^T \phi_i(t) \phi_j(t) \, dt = \delta_{i,j} \qquad \text{for all } i, j.$$

In other words, *the Karhunen-Loeve expansion provides us with many of the nice properties of the Fourier series for the case where $X(t)$ is not mean square periodic*. For simplicity, we again assume that $X(t)$ is zero mean.

    In order to motivate the Karhunen-Loeve expansion, we review the Karhunen-Loeve transform for vector random variables as introduced in Section 6.3. Let $\mathbf{X}$ be a zero-mean, vector random variable with covariance matrix K_X. The eigenvalues and eigenvectors of K_X are obtained from

$$K_X \mathbf{e}_i = \lambda_i \mathbf{e}_i, \tag{9.120}$$

where the $\mathbf{e}_i$ are column vectors. The set of normalized eigenvectors are orthonormal, that is, $\mathbf{e}_i^{\mathrm{T}} \mathbf{e}_j = \delta_{i, j}$. Define the matrix P of eigenvectors and Λ of eigenvalues as

$$P = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] \qquad \Lambda = \mathrm{diag}[\lambda_i],$$

then

$$K_X = P\Lambda P^{\mathrm{T}} = [\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{e}_1^{\mathrm{T}} \\ \mathbf{e}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{e}_n^{\mathrm{T}} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1\mathbf{e}_1, \lambda_2\mathbf{e}_2, \ldots, \lambda_n\mathbf{e}_n \end{bmatrix} \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix}$$

$$= \sum_{k=1}^{n} \lambda_i \mathbf{e}_i \mathbf{e}_i^T. \tag{9.121a}$$

Therefore we find that the covariance matrix can be expanded as a weighted sum of matrices, $\mathbf{e}_i\mathbf{e}_i^T$. In addition, if we let $\mathbf{Y} = P^T\mathbf{X}$, then the random variables in $\mathbf{Y}$ are orthogonal. Furthermore, since $PP^T = I$, then

$$\mathbf{X} = P\mathbf{Y} = \begin{bmatrix} \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{k=1}^{n} Y_k \mathbf{e}_k. \tag{9.121b}$$

Thus we see that the arbitrary vector random variable $\mathbf{X}$ can be expanded as a weighted sum of the eigenvectors of K_X, where the coefficients are orthogonal random variables. Furthermore the eigenvectors form an orthonormal set. These are exactly the properties we seek in the Karhunen-Loeve expansion for $X(t)$. If the vector random variable $\mathbf{X}$ is jointly Gaussian, then the components of $\mathbf{Y}$ are independent random variables. This results in tremendous simplification in a wide variety of problems.

In analogy to Eq. (9.120), we begin by considering the following eigenvalue equation:

$$\int_0^T K_X(t_1, t_2)\phi_k(t_2)\, dt_2 = \lambda_k\phi_k(t_1) \qquad 0 \le t_1 \le T. \tag{9.122}$$

The values λ_k and the corresponding functions $\phi_k(t)$ for which the above equation holds are called the eigenvalues and eigenfunctions of the covariance function $K_X(t_1, t_2)$. Note that it is possible for the eigenfunctions to be complex-valued, e.g., complex exponentials. It can be shown that if $K_X(t_1, t_2)$ is continuous, then the normalized eigenfunctions form an orthonormal set and satisfy Mercer's theorem:

$$K_X(t_1, t_2) = \sum_{k=1}^{\infty} \lambda_k\phi_k(t_1)\phi_k^*(t_2). \tag{9.123}$$

Note the correspondence between Eq. (9.121) and Eq. (9.123). Equation (9.123) in turn implies that

$$K_X(t, t) = \sum_{k=1}^{\infty} \lambda_k|\phi_k(t)|^2. \tag{9.124}$$

We are now ready to show that the equality in Eq. (9.119a) holds in the mean square sense and that the coefficients X_k are orthogonal random variables. First consider $E[X_kX_m^*]$:

$$E[X_kX_m^*] = E\left[X_m^* \int_0^T X(t')\phi_k^*(t)\, dt' \right] = \int_0^T E[X(t')X_m^*]\phi_k^*(t')\, dt'.$$

The integrand of the above equation has

$$E[X(t)X_m^*] = E\left[X(t)\int_0^T X^*(u)\phi_m(u)\,du\right] = \int_0^T K_X(t,u)\phi_m(u)\,du$$
$$= \lambda_m\phi_m(t).$$

Therefore

$$E[X_kX_m^*] = \int_0^T \lambda_m\phi_k^*(t')\phi_m(t')\,dt' = \lambda_k\delta_{k,m},$$

where $\delta_{k,m}$ is the Kronecker delta function. Thus X_k and X_m are orthogonal random variables. Note that the above equation implies that $\lambda_k = E\left[|X_k|^2\right]$, that is, the eigenvalues are real-valued.

To show that the Karhunen-Loeve expansion equals $X(t)$ in the mean square sense, we take

$$E\left[\left|X(t) - \sum_{k=-\infty}^{\infty} X_k\phi_k(t)\right|^2\right]$$
$$= E[|X(t)|^2] - E\left[X(t)\sum_{k=-\infty}^{\infty} X_k^*\phi_k^*(t)\right]$$
$$\quad - E\left[X^*(t)\sum_{k=-\infty}^{\infty} X_k\phi_k(t)\right]$$
$$\quad + E\left[\sum_{k=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} X_kX_m^*\phi_k(t)\phi_m^*(t)\right]$$
$$= R_X(t,t) - \sum_{k=-\infty}^{\infty} \lambda_k|\phi_k(t)|^2$$
$$\quad - \sum_{k=-\infty}^{\infty} \lambda_k^*|\phi_k(t)|^2 + \sum_{k=-\infty}^{\infty} \lambda_k|\phi_k(t)|^2.$$

The above equation equals zero from Eq. (9.124) and from the fact that the λ_k are real. Thus we have shown that Eq. (9.119a) holds in the mean square sense.

Finally, we note that in the important case where $X(t)$ is a Gaussian random process, then the components X_k will be independent Gaussian random variables. This result is extremely useful in solving certain signal detection and estimation problems. [Van Trees.]

Example 9.51   Wiener Process

Find the Karhunen-Loeve expansion for the Wiener process.

Equation (9.122) for the Wiener process gives, for $0 \le t_1 \le T$,

$$\lambda\phi(t_1) = \int_0^T \sigma^2\min(t_1,t_2)\phi(t_2)\,dt_2$$
$$= \sigma^2\int_0^{t_1} t_2\phi(t_2)\,dt_2 + \sigma^2\int_{t_1}^T t_1\phi(t_2)\,dt_2.$$

We differentiate the above integral equation once with respect to t_1 to obtain an integral equation and again to obtain a differential equation:

$$\sigma^2 \int_{t_1}^{T} \phi(t_2) \, dt_2 = \lambda \frac{d}{dt_1} \phi(t_1)$$

$$-\phi(t_1) = \frac{\lambda}{\sigma^2} \frac{d^2}{dt_1^2} \phi(t_1).$$

This second-order differential equation has a sinusoidal solution:

$$\phi(t_1) = a \sin \frac{\sigma t_1}{\sqrt{\lambda}} + b \cos \frac{\sigma t_1}{\sqrt{\lambda}}.$$

In order to solve the above equation for a, b, and λ, we need boundary conditions for the differential equation. We obtain these by substituting the general solution for $\phi(t)$ into the integral equation:

$$\frac{\lambda}{\sigma^2} \left(a \sin \frac{\sigma t_1}{\sqrt{\lambda}} + b \cos \frac{\sigma t_1}{\sqrt{\lambda}} \right) = \int_{0}^{t_1} t_2 \phi(t_2) \, dt_2 + \int_{t_1}^{T} t_1 \phi(t_2) \, dt_2.$$

As t_1 approaches zero, the right-hand side approaches zero. This implies that $b = 0$ in the left-hand side of the equation. A second boundary condition is obtained by letting t_1 approach T in the equation obtained after the first differentiation of the integral equation:

$$0 = \lambda \frac{d}{dt_1} \phi(T) = \frac{\sigma a}{\sqrt{\lambda}} \cos \frac{\sigma T}{\sqrt{\lambda}}.$$

This implies that

$$\frac{\sigma T}{\sqrt{\lambda}} = \left(n - \frac{1}{2} \right) \pi \qquad n = 1, 2, \ldots.$$

Therefore the eigenvalues are given by

$$\lambda_n = \frac{\sigma^2 T^2}{\left(n - \frac{1}{2} \right)^2 \pi^2} \qquad n = 1, 2, \ldots.$$

The normalization requirement implies that

$$1 = \int_{0}^{T} \left(a \sin \frac{\sigma t}{\sqrt{\lambda}} \right)^2 dt = a^2 \frac{T}{2},$$

which implies that $a = (2/T)^{1/2}$. Thus the eigenfunctions are given by

$$\phi_n(t) = \sqrt{\frac{2}{T}} \sin \left(n - \frac{1}{2} \right) \frac{\pi}{T} t \qquad 0 \le t \le T,$$

and the Karhunen-Loeve expansion for the Wiener process is

$$X(t) = \sum_{n-1}^{\infty} X_n \sqrt{\frac{2}{T}} \sin \left(n - \frac{1}{2} \right) \frac{\pi}{T} t \qquad 0 \le t < T,$$

where the X_n are zero-mean, independent Gaussian random variables with variance given by λ_n.

Example 9.52   White Gaussian Noise Process

Find the Karhunen-Loeve expansion of the white Gaussian noise process.

The white Gaussian noise process is the derivative of the Wiener process. If we take the derivative of the Karhunen-Loeve expansion of the Wiener process, we obtain

$$X'(t) = \sum_{n=1}^{\infty} \frac{\sigma}{\sqrt{\lambda}} X_n \sqrt{\frac{2}{T}} \cos\left(n - \frac{1}{2}\right)\frac{\pi}{T}t$$

$$= \sum_{n=1}^{\infty} W_n \sqrt{\frac{2}{T}} \cos\left(n - \frac{1}{2}\right)\frac{\pi}{T}t \qquad 0 \leq t < T,$$

where the W_n are independent Gaussian random variables with the same variance σ^2. This implies that the process has infinite power, a fact we had already found about the white Gaussian noise process. In the Problems we will see that any orthonormal set of eigenfunctions can be used in the Karhunen-Loeve expansion for white Gaussian noise.

9.10   GENERATING RANDOM PROCESSES

Many engineering systems involve random processes that interact in complex ways. It is not always possible to model these systems precisely using analytical methods. In such situations computer simulation methods are used to investigate the system dynamics and to measure the performance parameters of interest. In this section we consider two basic methods to generating random processes. The first approach involves generating the sum process of iid sequences of random variables. We saw that this approach can be used to generate the binomial and random walk processes, and, through limiting procedures, the Wiener and Poisson processes. The second approach involves taking the linear combination of deterministic functions of time where the coefficients are given by random variables. The Fourier series and Karhunen-Loeve expansion use this approach. Real systems, e.g., digital modulation systems, also generate random processes in this manner.

9.10.1   Generating Sum Random Processes

The generation of sample functions of the sum random process involves two steps:

1. Generate a sequence of iid random variables that drive the sum process.
2. Generate the cumulative sum of the iid sequence.

Let D be an array of samples of the desired iid random variables. The function cumsum(D) in Octave and MATLAB then provides the cumulative sum, that is, the sum process, that results from the sequence in D.

The code below generates m realizations of an n-step random walk process.

```
>p=1/2
>n=1000
>m=4
```

```
> V=-1:2:1;
> P=[1-p,p];
> D=discrete_rnd(V, P, m, n);
> X=cumsum (D);
> plot (X)
```

Figures 9.7(a) and 9.7(b) in Section 9.3 show four sample functions of the symmetric random walk process for $p = 1/2$. The sample functions vary over a wide range of positive and negative values. Figure 9.7(c) shows four sample functions for $p = 3/4$. The sample functions now have a strong linear trend consistent with the mean $n(2p - 1)$. The variability about this trend is somewhat less than in the symmetric case since the variance function is now $n4p(1 - p) = 3n/4$.

We can generate an approximation to a Poisson process by summing iid Bernoulli random variables. Figure 9.18(a) shows ten realizations of Poisson processes with $\lambda = 0.4$ arrivals per second. The sample functions for $T = 50$ seconds were generated using a 1000-step binomial process with $p = \lambda T/n = 0.02$. The linear increasing trend of the Poisson process is evident in the figure. Figure 9.18(b) shows the estimate of the mean and variance functions obtained by averaging across the 10 realizations. The linear trend in the sample mean function is very clear; the sample variance function is also linear but is much more variable. The mean and variance functions of the realizations are obtained using the commands mean(transpose(X)) and var(transpose(X)).

We can generate sample functions of the random telegraph signal by taking the Poisson process $N(t)$ and calculating $X(t) = 2(N(t) \text{ modulo } 2) - 1$. Figure 9.19(a) shows a realization of the random telegraph signal. Figure 9.19(b) shows an estimate of the covariance function of the random telegraph signal. The exponential decay in the covariance function can be seen in the figure. See Eq. (9.44).
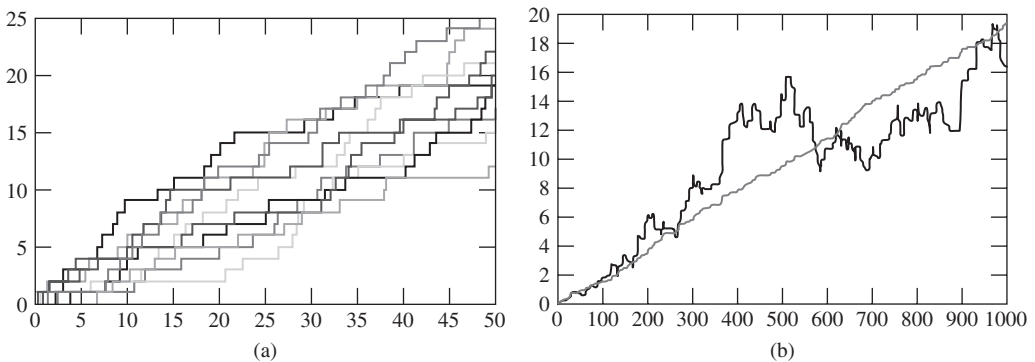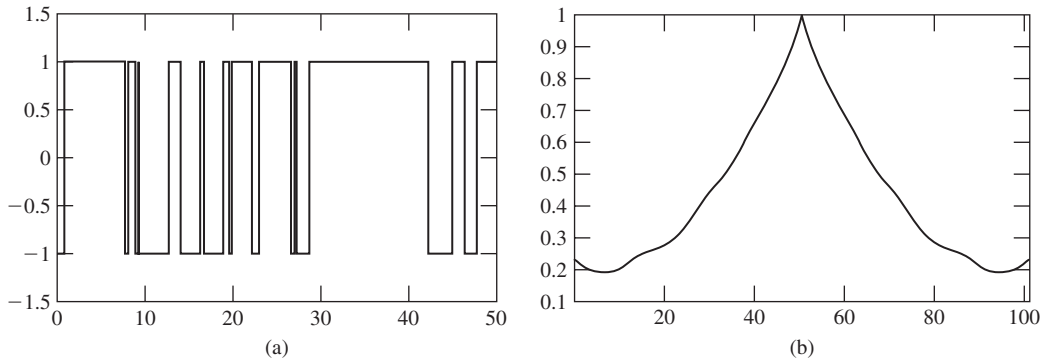

(a)   (b)

FIGURE 9.18
(a) Ten sample functions of a Poisson random process with $\lambda = 0.4$. (b) Sample mean and variance of ten sample functions of a Poisson random process with $\lambda = 0.4$.

FIGURE 9.19
(a) Sample function of a random telegraph process with $\lambda = 0.4$. (b) Estimate of covariance function of a random telegraph process.

The covariance function is computed using the function CX_est below.

```
function [CXall]=CX_est (X, L, M_est)
N=length(X);                                    % N is number of samples
CX=zeros (1,L+1);                               % L is maximum lag
M_est=mean(X)                                   % Sample mean
for m=1:L+1,                                     % Add product terms
   for n=1:N-m+1,
      CX(m)=CX(m) + (X(n) - M_est) * (X(n+m-1)- M_est);
   end;
    CX (m)=CX(m) / (N-m+1);                     % Normalize by number of terms
end;
for i=1:L,
    CXall(i)=CX(L+2-i);                         % Lags 1 to L
end
CXall(L+1:2*L+1)=CX(1:L+1);                     % Lags L + 1 to 2L + 1
```

The Wiener random process can also be generated as a sum process. One approach is to generate a properly scaled random walk process, as in Eq. (9.50). A better approach is to note that the Wiener process has independent Gaussian increments, as in Eq. (9.52), and therefore, to generate the sequence D of increments for the time subintervals, and to then find the corresponding sum process. The code below generates a sample of the Wiener process:

```
> a=2
> delta=0.001
> n=1000
> D=normal_rnd(0,a*delta,1,n)
> X=cumsum(D);
> plot(X)
```
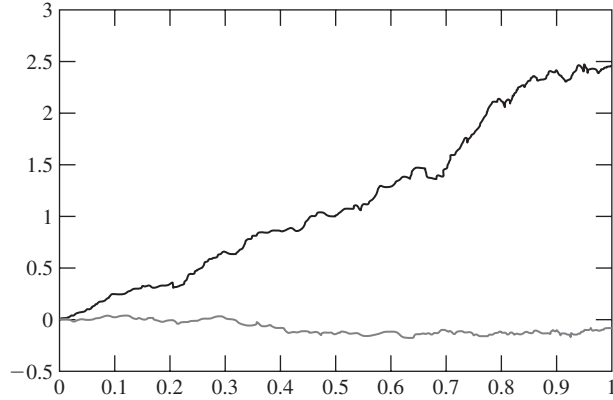
FIGURE 9.20
Sample mean and variance functions from 50 realizations of
Wiener process.

Figure 9.12 in Section 9.5 shows four sample functions of a Brownian motion process with $\alpha = 2$. Figure 9.20 shows the sample mean and sample variance of 50 sample functions of the Wiener process with $\alpha = 2$. It can be seen that the mean across the 50 realizations is close to zero which is the actual mean function for the process. The sample variance across the 50 realizations increases steadily and is close to the actual variance function which is $\alpha t = 2t$.

9.10.2   Generating Linear Combinations of Deterministic Functions

In some situations a random process can be represented as a linear combination of deterministic functions where the coefficients are random variables. The Fourier series and the Karhunen-Loeve expansions are examples of this type of representation.
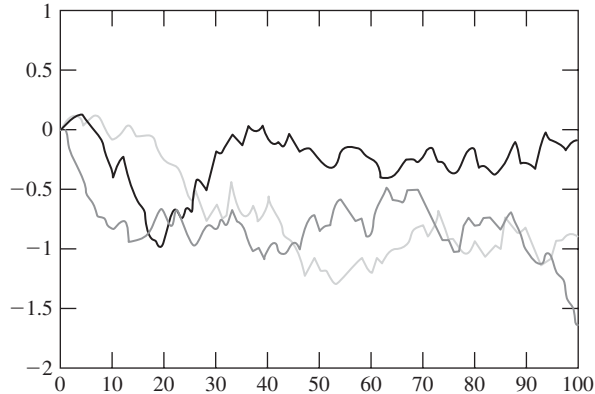
In Example 9.51 let the parameters in the Karhunen-Loeve expansion for a Wiener process in the interval $0 \le t \le T$ be $T = 1, \sigma^2 = 1$:

$$X(t) = \sum_{n=1}^{\infty} X_n \sqrt{\frac{2}{T}} \sin\left(n - \frac{1}{2}\right)\frac{\pi t}{T} = \sum_{n=1}^{\infty} X_n \sqrt{2} \sin\left(n - \frac{1}{2}\right)\pi t$$

where the X_n are zero-mean, independent Gaussian random variables with variance

$$\lambda_n = \frac{\sigma^2 T^2}{(n - 1/2)^2 \pi^2} = \frac{1}{(n - 1/2)^2 \pi^2}.$$

The following code generates the 100 Gaussian coefficients for the Karhunen-Loeve expansion for the Wiener process.

FIGURE 9.21
Sample functions for Wiener process using 100 terms in Karhunen-Loeve expansion.

```
> M=zeros(100,1);
> n=1:1:100;                      % Number of coefficients
> N=transpose(n);
> v=1./((N−0.5).^2 *pi ^2);       % Variances of coefficients
> t=0.01:0.01:1;
> p=(N−0.5)*t;                    % Argument of sinusoid
> x=normal_rnd(M,v,100,1);        % Gaussian coefficients
> y=sqrt(2)*sin(pi *p);           % sin terms
> z=transpose(x)*y
> plot(z)
```

Figure 9.21 shows the Karhunen-Loeve expansion for the Wiener process using 100 terms. The sample functions generally exhibit the same type behavior as in the previous figures. The sample functions, however, do not exhibit the jaggedness of the other examples, which are based on the generation of many more random variables.

SUMMARY

- A random process or stochastic process is an indexed family of random variables that is specified by the set of joint distributions of any number and choice of random variables in the family. The mean, autocovariance, and autocorrelation functions summarize some of the information contained in the joint distributions of pairs of time samples.

- The sum process of an iid sequence has the property of stationary and independent increments, which facilitates the evaluation of the joint pdf/pmf of the

process at any set of time instants. The binomial and random processes are sum processes. The Poisson and Wiener processes are obtained as limiting forms of these sum processes.

- The Poisson process has independent, stationary increments that are Poisson distributed. The interarrival times in a Poisson process are iid exponential random variables.
- The mean and covariance functions completely specify all joint distributions of a Gaussian random process.
- The Wiener process has independent, stationary increments that are Gaussian distributed. The Wiener process is a Gaussian random process.
- A random process is stationary if its joint distributions are independent of the choice of time origin. If a random process is stationary, then $m_X(t)$ is constant, and $R_X(t_1, t_2)$ depends only on $t_1 - t_2$.
- A random process is wide-sense stationary (WSS) if its mean is constant and if its autocorrelation and autocovariance depend only on $t_1 - t_2$. A WSS process need not be stationary.
- A wide-sense stationary Gaussian random process is also stationary.
- A random process is cyclostationary if its joint distributions are invariant with respect to shifts of the time origin by integer multiples of some period T.
- The white Gaussian noise process results from taking the derivative of the Wiener process.
- The derivative and integral of a random process are defined as limits of random variables. We investigated the existence of these limits in the mean square sense.
- The mean and autocorrelation functions of the output of systems described by a linear differential equation and subject to random process inputs can be obtained by solving a set of differential equations. If the input process is a Gaussian random process, then the output process is also Gaussian.
- Ergodic theorems state when time-average estimates of a parameter of a random process converge to the expected value of the parameter. The decay rate of the covariance function determines the convergence rate of the sample mean.

CHECKLIST OF IMPORTANT TERMS

Autocorrelation function
Autocovariance function
Average power
Bernoulli random process
Binomial counting process
Continuous-time process
Cross-correlation function
Cross-covariance function
Cyclostationary random process
Discrete-time process

Ergodic theorem
Fourier series
Gaussian random process
Hurst parameter
iid random process
Independent increments
Independent random processes
Karhunen-Loeve expansion
Markov random process
Mean ergodic random process

Mean function
Mean square continuity
Mean square derivative
Mean square integral
Mean square periodic process
Ornstein-Uhlenbeck process
Orthogonal random processes
Poisson process
Random process
Random telegraph signal
Random walk process
Realization, sample path, or sample
   function

Shot noise
Stationary increments
Stationary random process
Stochastic process
Sum random process
Time average
Uncorrelated random processes
Variance of $X(t)$
White Gaussian noise
Wide-sense cyclostationary process
Wiener process
WSS random process

ANNOTATED REFERENCES

References [1] through [6] can be consulted for further reading on random processes. Larson and Shubert [ref 7] and Yaglom [ref 8] contain excellent discussions on white Gaussian noise and Brownian motion. Van Trees [ref 9] gives detailed examples on the application of the Karhunen-Loeve expansion. Beran [ref 10] discusses long memory processes.

1. A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 2002.
2. W. B. Davenport, *Probability and Random Processes: An Introduction for Applied Scientists and Engineers*, McGraw-Hill, New York, 1970.
3. H. Stark and J. W. Woods, *Probability and Random Processes with Applications to Signal Processing*, 3d ed., Prentice Hall, Upper Saddle River, N.J., 2002.
4. R. M. Gray and L. D. Davisson, *Random Processes: A Mathematical Approach for Engineers*, Prentice Hall, Englewood Cliffs, N.J., 1986.
5. J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineering*, Cambridge University Press, Cambridge, 2006.
6. G. Grimett and D. Stirzaker, *Probability and Random Processes*, Oxford University Press, Oxford, 2006.
7. H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences*, vol. 1, Wiley, New York, 1979.
8. A. M. Yaglom, *Correlation Theory of Stationary and Related Random Functions*, vol. 1: *Basic Results*, Springer-Verlag, New York, 1987.
9. H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Wiley, New York, 1987.
10. J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall/CRC, New York, 1994.

PROBLEMS

Sections 9.1 and 9.2: Definition and Specification of a Stochastic Process

9.1. In Example 9.1, find the joint pmf for X_1 and X_2. Why are X_1 and X_2 independent?

9.2. A discrete-time random process X_n is defined as follows. A fair die is tossed and the outcome k is observed. The process is then given by $X_n = k$ for all n.

   (a) Sketch some sample paths of the process.

   (b) Find the pmf for X_n.

   (c) Find the joint pmf for X_n and X_{n+k}.

   (d) Find the mean and autocovariance functions of X_n.

9.3. A discrete-time random process X_n is defined as follows. A fair coin is tossed. If the outcome is heads, $X_n = (-1)^n$ for all n; if the outcome is tails, $X_n = (-1)^{n+1}$ for all n.

   (a) Sketch some sample paths of the process.

   (b) Find the pmf for X_n.

   (c) Find the joint pmf for X_n and X_{n+k}.

   (d) Find the mean and autocovariance functions of X_n.

9.4. A discrete-time random process is defined by $X_n = s^n$, for $n \geq 0$, where s is selected at random from the interval $(0, 1)$.

   (a) Sketch some sample paths of the process.

   (b) Find the cdf of X_n.

   (c) Find the joint cdf for X_n and X_{n+1}.

   (d) Find the mean and autocovariance functions of X_n.

   (e) Repeat parts a, b, c, and d if s is uniform in $(1, 2)$.

9.5. Let $g(t)$ be the rectangular pulse shown in Fig. P9.1. The random process $X(t)$ is defined as

$$X(t) = Ag(t),$$

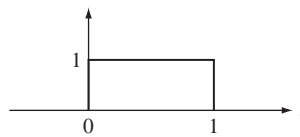where A assumes the values ± 1 with equal probability.

FIGURE P9.1

   (a) Find the pmf of $X(t)$.

   (b) Find $m_X(t)$.

   (c) Find the joint pmf of $X(t)$ and $X(t + d)$.

   (d) Find $C_X(t, t + d), d > 0$.

9.6. A random process is defined by

$$Y(t) = g(t - T),$$

where $g(t)$ is the rectangular pulse of Fig. P9.1, and T is a uniformly distributed random variable in the interval $(0, 1)$.

  (a) Find the pmf of $Y(t)$.

  (b) Find $m_Y(t)$ and $C_Y(t_1, t_2)$.

9.7. A random process is defined by

$$X(t) = g(t - T),$$

where T is a uniform random variable in the interval $(0, 1)$ and $g(t)$ is the periodic triangular waveform shown in Fig. P9.2.
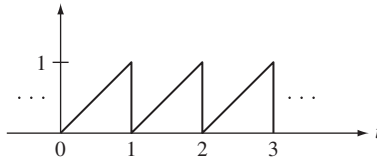

FIGURE P9.2

  (a) Find the cdf of $X(t)$ for $0 < t < 1$.

  (b) Find $m_X(t)$ and $C_X(t_1, t_2)$.

9.8. Let $Y(t) = g(t - T)$ as in Problem 9.6, but let T be an exponentially distributed random variable with parameter α.

  (a) Find the pmf of $Y(t)$.

  (b) Find the joint pmf of $Y(t)$ and $Y(t + d)$. Consider two cases: $d > 1$, and $0 < d < 1$.

  (c) Find $m_Y(t)$ and $C_Y(t, t + d)$ for $d > 1$ and $0 < d < 1$.

9.9. Let $Z(t) = At^3 + B$, where A and B are independent random variables.

  (a) Find the pdf of $Z(t)$.

  (b) Find $m_Z(t)$ and $C_Z(t_1, t_2)$.

9.10. Find an expression for $E\left[|X_{t_2} - X_{t_1}|^2\right]$ in terms of autocorrelation function.

9.11. The random process $H(t)$ is defined as the "hard-limited" version of $X(t)$:

$$H(t) = \begin{cases} +1 & \text{if} \quad X(t) \geq 0 \\ -1 & \text{if} \quad X(t) < 0. \end{cases}$$

  (a) Find the pdf, mean, and autocovariance of $H(t)$ if $X(t)$ is the sinusoid with a random amplitude presented in Example 9.2.

  (b) Find the pdf, mean, and autocovariance of $H(t)$ if $X(t)$ is the sinusoid with random phase presented in Example 9.9.

  (c) Find a general expression for the mean of $H(t)$ in terms of the cdf of $X(t)$.

9.12. (a) Are independent random processes orthogonal? Explain.

  (b) Are orthogonal random processes uncorrelated? Explain.

  (c) Are uncorrelated processes independent?

  (d) Are uncorrelated processes orthogonal?

9.13. The random process $Z(t)$ is defined by

$$Z(t) = 2Xt - Y,$$

where X and Y are a pair of random variables with means m_X, m_Y, variances σ_X^2, σ_Y^2, and correlation coefficient $\rho_{X,Y}$. Find the mean and autocovariance of $Z(t)$.

9.14. Let $H(t)$ be the output of the hard limiter in Problem 9.11.

    (a) Find the cross-correlation and cross-covariance between $H(t)$ and $X(t)$ when the input is a sinusoid with random amplitude as in Problem 9.11a.

    (b) Repeat if the input is a sinusoid with random phase as in Problem 9.11b.

    (c) Are the input and output processes uncorrelated? Orthogonal?

9.15. Let $Y_n = X_n + g(n)$ where X_n is a zero-mean discrete-time random process and $g(n)$ is a deterministic function of n.

    (a) Find the mean and variance of Y_n.

    (b) Find the joint cdf of Y_n and Y_{n+1}.

    (c) Find the autocovariance function of Y_n.

    (d) Plot typical sample functions for X_n and Y_n if: $g(n) = n$; $g(n) = 1/n^2$; $g(n) = 1/n$.

9.16. Let $Y_n = c(n)X_n$ where X_n is a zero-mean, unit-variance, discrete-time random process and $c(n)$ is a deterministic function of n.

    (a) Find the mean and variance of Y_n.

    (b) Find the joint cdf of Y_n and Y_{n+1}.

    (c) Find the autocovariance function of Y_n.

    (d) Plot typical sample functions for X_n and Y_n if: $c(n) = n$; $c(n) = 1/n^2$; $c(n) = 1/n$.

9.17. (a) Find the cross-correlation and cross-covariance for X_n and Y_n in Problem 9.15.

    (b) Find the joint pdf of X_n and Y_{n+1}.

    (c) Determine whether X_n and Y_n are uncorrelated, independent, or orthogonal random processes.

9.18. (a) Find the cross-correlation and cross-covariance for X_n and Y_n in Problem 9.16.

    (b) Find the joint pdf of X_n and Y_{n+1}.

    (c) Determine whether X_n and Y_n are uncorrelated, independent, or orthogonal random processes.

9.19. Suppose that $X(t)$ and $Y(t)$ are independent random processes and let

$$U(t) = X(t) - Y(t)$$
$$V(t) = X(t) + Y(t).$$

    (a) Find $C_{UX}(t_1, t_2)$, $C_{UY}(t_1, t_2)$, and $C_{UV}(t_1, t_2)$.

    (b) Find the $f_{U(t_1)X(t_2)}(u, x)$, and $f_{U(t_1)V(t_2)}(u, v)$. *Hint:* Use auxiliary variables.

9.20. Repeat Problem 9.19 if $X(t)$ and $Y(t)$ are independent discrete-time processes and $X(t)$ and $Y(t)$ have different iid random processes.

Section 9.3: Sum Process, Binomial Counting Process, and Random Walk

9.21. (a) Let Y_n be the process that results when individual 1's in a Bernoulli process are erased with probability α. Find the pmf of S_n', the counting process for Y_n. Does Y_n have independent and stationary increments?

    (b) Repeat part a if in addition to the erasures, individual 0's in the Bernoulli process are changed to 1's with probability β.

9.22. Let S_n denote a binomial counting process.

    (a) Show that $P[S_n = j, S_{n'} = i] \neq P[S_n = j]P[S_{n'} = i]$.

    (b) Find $P[S_{n_2} = j \,|\, S_{n_1} = i]$, where $n_2 > n_1$.

    (c) Show that $P[S_{n_2} = j \,|\, S_{n_1} = i, S_{n_0} = k] = P[S_{n_2} = j \,|\, S_{n_1} = i]$, where $n_2 > n_1 > n_0$.

9.23. (a) Find $P[S_n = 0]$ for the random walk process.

    (b) What is the answer in part a if $p = 1/2$?

9.24. Consider the following *moving average* processes:

$$Y_n = 1/2(X_n + X_{n-1}) \qquad X_0 = 0$$

$$Z_n = 2/3\, X_n + 1/3\, X_{n-1} \qquad X_0 = 0$$

    (a) Find the mean, variance, and covariance of Y_n and Z_n if X_n is a Bernoulli random process.

    (b) Repeat part a if X_n is the random step process.

    (c) Generate 100 outcomes of a Bernoulli random process X_n, and find the resulting Y_n and Z_n. Are the sample means of Y_n and Z_n in part a close to their respective means?

    (d) Repeat part c with X_n given by the random step process.

9.25. Consider the following autoregressive processes:

$$W_n = 2W_{n-1} + X_n \qquad W_0 = 0$$

$$Z_n = 3/4\, Z_{n-1} + X_n \quad Z_0 = 0.$$

    (a) Suppose that X_n is a Bernoulli process. What trends do the processes exhibit?

    (b) Express W_n and Z_n in terms of $X_n, X_{n-1}, \ldots, X_1$ and then find $E[W_n]$ and $E[Z_n]$. Do these results agree with the trends you expect?

    (c) Do W_n or Z_n have independent increments? stationary increments?

    (d) Generate 100 outcomes of a Bernoulli process. Find the resulting realizations of W_n and Z_n. Is the sample mean meaningful for either of these processes?

    (e) Repeat part d if X_n is the random step process.

9.26. Let M_n be the discrete-time process defined as the sequence of sample means of an iid sequence:

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

    (a) Find the mean, variance, and covariance of M_n.

    (b) Does M_n have independent increments? stationary increments?

9.27. Find the pdf of the processes defined in Problem 9.24 if the X_n are an iid sequence of zero-mean, unit-variance Gaussian random variables.

9.28. Let X_n consist of an iid sequence of Cauchy random variables.

    (a) Find the pdf of the sum process S_n. *Hint:* Use the characteristic function method.

    (b) Find the joint pdf of S_n and S_{n+k}.

9.29. Let X_n consist of an iid sequence of Poisson random variables with mean α.

    (a) Find the pmf of the sum process S_n.

    (b) Find the joint pmf of S_n and S_{n+k}.

9.30. Let X_n be an iid sequence of zero-mean, unit-variance Gaussian random variables.
  (a) Find the pdf of M_n defined in Problem 9.26.
  (b) Find the joint pdf of M_n and M_{n+k}. *Hint:* Use the independent increments property of S_n.

9.31. Repeat Problem 9.26 with $X_n = 1/2(Y_n + Y_{n-1})$, where Y_n is an iid random process. What happens to the variance of M_n as n increases?

9.32. Repeat Problem 9.26 with $X_n = 3/4X_{n-1} + Y_n$ where Y_n is an iid random process. What happens to the variance of M_n as n increases?

9.33. Suppose that an experiment has three possible outcomes, say $0, 1$, and 2, and suppose that these occur with probabilities p_0, p_1, and p_2, respectively. Consider a sequence of independent repetitions of the experiment, and let $X_j(n)$ be the indicator function for outcome j. The vector

$$\mathbf{X}(n) = (X_0(n), X_1(n), X_2(n))$$

then constitutes a vector-valued Bernoulli random process. Consider the counting process for $\mathbf{X}(n)$:

$$\mathbf{S}(n) = \mathbf{X}(n) + \mathbf{X}(n - 1) + \cdots + \mathbf{X}(1) \qquad \mathbf{S}(0) = 0.$$

  (a) Show that $\mathbf{S}(n)$ has a multinomial distribution.
  (b) Show that $\mathbf{S}(n)$ has independent increments, then find the joint pmf of $\mathbf{S}(n)$ and $\mathbf{S}(n + k)$.
  (c) Show that components $S_j(n)$ of the vector process are binomial counting processes.

Section 9.4: Poisson and Associated Random Processes

9.34. A server handles queries that arrive according to a Poisson process with a rate of 10 queries per minute. What is the probability that no queries go unanswered if the server is unavailable for 20 seconds?

9.35. Customers deposit \$1 in a vending machine according to a Poisson process with rate λ. The machine issues an item with probability p. Find the pmf for the number of items dispensed in time t.

9.36. Noise impulses occur in a radio transmission according to a Poisson process of rate λ.
  (a) Find the probability that no impulses occur during the transmission of a message that is t seconds long.
  (b) Suppose that the message is encoded so that the errors caused by up to 2 impulses can be corrected. What is the probability that a t-second message cannot be corrected?

9.37. Packets arrive at a multiplexer at two ports according to independent Poisson processes of rates $\lambda_1 = 1$ and $\lambda_2 = 2$ packets/second, respectively.
  (a) Find the probability that a message arrives first on line 2.
  (b) Find the pdf for the time until a message arrives on either line.
  (c) Find the pmf for $N(t)$, the total number of messages that arrive in an interval of length t.
  (d) Generalize the result of part c for the "merging" of k independent Poisson processes of rates $\lambda_1, \ldots, \lambda_k$, respectively:

$$N(t) = N_1(t) + \cdots + N_k(t).$$

9.38. (a) Find $P[N(t - d) = j \mid N(t) = k]$ with $d > 0$, where $N(t)$ is a Poisson process with rate λ.

(b) Compare your answer to $P[N(t + d) = j \mid N(t) = k]$. Explain the difference, if any.

9.39. Let $N_1(t)$ be a Poisson process with arrival rate λ_1 that is started at $t = 0$. Let $N_2(t)$ be another Poisson process that is independent of $N_1(t)$, that has arrival rate λ_2, and that is started at $t = 1$.

(a) Show that the pmf of the process $N(t) = N_1(t) + N_2(t)$ is given by:

$$P[N(t + \tau) - N(t) = k] = \frac{(m(t + \tau) - m(t))^k}{k!} e^{-(m(t+\tau)-m(t))} \qquad \text{for } k = 0, 1, \ldots$$

where $m(t) = E[N(t)]$.

(b) Now consider a Poisson process in which the arrival rate $\lambda(t)$ is a piecewise constant function of time. Explain why the pmf of the process is given by the above pmf where

$$m(t) = \int_0^t \lambda(t') \, dt'.$$

(c) For what other arrival functions $\lambda(t)$ does the pmf in part a hold?

9.40. (a) Suppose that the time required to service a customer in a queueing system is a random variable T. If customers arrive at the system according to a Poisson process with parameter λ, find the pmf for the number of customers that arrive during one customer's service time. *Hint:* Condition on the service time.

(b) Evaluate the pmf in part a if T is an exponential random variable with parameter β.

9.41. (a) Is the difference of two independent Poisson random processes also a Poisson process?

(b) Let $N_p(t)$ be the number of complete pairs generated by a Poisson process up to time t. Explain why $N_p(t)$ is or is not a Poisson process.

9.42. Let $N(t)$ be a Poisson random process with parameter λ. Suppose that each time an event occurs, a coin is flipped and the outcome (heads or tails) is recorded. Let $N_1(t)$ and $N_2(t)$ denote the number of heads and tails recorded up to time t, respectively. Assume that p is the probability of heads.

(a) Find $P[N_1(t) = j, N_2(t) = k \mid N(t) = k + j]$.

(b) Use part a to show that $N_1(t)$ and $N_2(t)$ are independent Poisson random variables of rates $p\lambda t$ and $(1 - p)\lambda t$, respectively:

$$P[N_1(t) = j, N_2(t) = k] = \frac{(p\lambda t)^j}{j!} e^{-p\lambda t} \frac{((1 - p)\lambda t)^k}{k!} e^{-(1-p)\lambda t}.$$

9.43. Customers play a $1 game machine according to a Poisson process with rate λ. Suppose the machine dispenses a random reward X each time it is played. Let $X(t)$ be the total reward issued up to time t.

(a) Find expressions for $P[X(t) = j]$ if X_n is Bernoulli.

(b) Repeat part a if X assumes the values $\{0, 5\}$ with probabilities $(5/6, 1/6)$.

- **(c)** Repeat part a if X is Poisson with mean 1.
- **(d)** Repeat part a if with probability p the machine returns all the coins.

9.44. Let $X(t)$ denote the random telegraph signal, and let $Y(t)$ be a process derived from $X(t)$ as follows: Each time $X(t)$ changes polarity, $Y(t)$ changes polarity with probability p.
- **(a)** Find the $P[Y(t) = \pm 1]$.
- **(b)** Find the autocovariance function of $Y(t)$. Compare it to that of $X(t)$.

9.45. Let $Y(t)$ be the random signal obtained by switching between the values 0 and 1 according to the events in a Poisson process of rate λ. Compare the pmf and autocovariance of $Y(t)$ with that of the random telegraph signal.

9.46. Let $Z(t)$ be the random signal obtained by switching between the values 0 and 1 according to the events in a counting process $N(t)$. Let

$$P[N(t) = k] = \frac{1}{1 + \lambda t}\left(\frac{\lambda t}{1 + \lambda t}\right)^k \qquad k = 0, 1, 2, \ldots.$$

- **(a)** Find the pmf of $Z(t)$.
- **(b)** Find $m_Z(t)$.

9.47. In the filtered Poisson process (Eq. (9.45)), let $h(t)$ be a pulse of unit amplitude and duration T seconds.
- **(a)** Show that $X(t)$ is then the increment in the Poisson process in the interval $(t - T, t)$.
- **(b)** Find the mean and autocorrelation functions of $X(t)$.

9.48. (a) Find the second moment and variance of the shot noise process discussed in Example 9.25.
- **(b)** Find the variance of the shot noise process if $h(t) = e^{-\beta t}$ for $t \geq 0$.

9.49. Messages arrive at a message center according to a Poisson process of rate λ. Every hour the messages that have arrived during the previous hour are forwarded to their destination. Find the mean of the total time waited by all the messages that arrive during the hour. *Hint:* Condition on the number of arrivals and consider the arrival instants.

Section 9.5: Gaussian Random Process, Wiener Process and Brownian Motion

9.50. Let $X(t)$ and $Y(t)$ be jointly Gaussian random processes. Explain the relation between the conditions of independence, uncorrelatedness, and orthogonality of $X(t)$ and $Y(t)$.

9.51. Let $X(t)$ be a zero-mean Gaussian random process with autocovariance function given by

$$C_X(t_1, t_2) = 4e^{-2|t_1 - t_2|}.$$

Find the joint pdf of $X(t)$ and $X(t + s)$.

9.52. Find the pdf of $Z(t)$ in Problem 9.13 if X and Y are jointly Gaussian random variables.

9.53. Let $Y(t) = X(t + d) - X(t)$, where $X(t)$ is a Gaussian random process.
- **(a)** Find the mean and autocovariance of $Y(t)$.
- **(b)** Find the pdf of $Y(t)$.
- **(c)** Find the joint pdf of $Y(t)$ and $Y(t + s)$.
- **(d)** Show that $Y(t)$ is a Gaussian random process.

9.54. Let $X(t) = A \cos \omega t + B \sin \omega t$, where A and B are iid Gaussian random variables with zero mean and variance σ^2.

    (a) Find the mean and autocovariance of $X(t)$.

    (b) Find the joint pdf of $X(t)$ and $X(t + s)$.

9.55. Let $X(t)$ and $Y(t)$ be independent Gaussian random processes with zero means and the same covariance function $C(t_1, t_2)$. Define the "amplitude-modulated signal" by

$$Z(t) = X(t) \cos \omega t + Y(t) \sin \omega t.$$

    (a) Find the mean and autocovariance of $Z(t)$.

    (b) Find the pdf of $Z(t)$.

9.56. Let $X(t)$ be a zero-mean Gaussian random process with autocovariance function given by $C_X(t_1, t_2)$. If $X(t)$ is the input to a "square law detector," then the output is

$$Y(t) = X(t)^2.$$

Find the mean and autocovariance of the output $Y(t)$.

9.57. Let $Y(t) = X(t) + \mu t$, where $X(t)$ is the Wiener process.

    (a) Find the pdf of $Y(t)$.

    (b) Find the joint pdf of $Y(t)$ and $Y(t + s)$.

9.58. Let $Y(t) = X^2(t)$, where $X(t)$ is the Wiener process.

    (a) Find the pdf of $Y(t)$.

    (b) Find the conditional pdf of $Y(t_2)$ given $Y(t_1)$.

9.59. Let $Z(t) = X(t) - aX(t - s)$, where $X(t)$ is the Wiener process.

    (a) Find the pdf of $Z(t)$.

    (b) Find $m_Z(t)$ and $C_Z(t_1, t_2)$.

9.60. (a) For $X(t)$ the Wiener process with $\alpha = 1$ and $0 < t < 1$, show that the joint pdf of $X(t)$ and $X(1)$ is given by:

$$f_{X(t), X(1)}(x_1, x_2) = \frac{\exp\left\{-\dfrac{1}{2}\left[\dfrac{x_1^2}{t} + \dfrac{(x_2 - x_1)^2}{(1 - t)}\right]\right\}}{2\pi\sqrt{t(1 - t)}}.$$

    (b) Use part a to show that for $0 < t < 1$, the conditional pdf of $X(t)$ given $X(0) = X(1) = 0$ is:

$$f_{X(t)}(x \mid X(0) = X(1) = 0) = \frac{\exp\left\{-\dfrac{1}{2}\left[\dfrac{x^2}{t(1 - t)}\right]\right\}}{2\pi\sqrt{t(1 - t)}}.$$

    (c) Use part b to find the conditional pdf of $X(t)$ given $X(t_1) = a$ and $X(t_2) = b$ for $t_1 < t < t_2$. *Hint:* Find the equivalent process in the interval $(0, t_2 - t_1)$.

Section 9.6: Stationary Random Processes

9.61. (a) Is the random amplitude sinusoid in Example 9.9 a stationary random process? Is it wide-sense stationary?

(b) Repeat part a for the random phase sinusoid in Example 9.10.

9.62. A discrete-time random process X_n is defined as follows. A fair coin is tossed; if the outcome is heads then $X_n = 1$ for all n, and $X_n = -1$ for all n, otherwise.

(a) Is X_n a WSS random process?

(b) Is X_n a stationary random process?

(c) Do the answers in parts a and b change if p is a biased coin?

9.63. Let X_n be the random process in Problem 9.3.

(a) Is X_n a WSS random process?

(b) Is X_n a stationary random process?

(c) Is X_n a cyclostationary random process?

9.64. Let $X(t) = g(t - T)$, where $g(t)$ is the periodic waveform introduced in Problem 9.7, and T is a uniformly distributed random variable in the interval $(0, 1)$. Is $X(t)$ a stationary random process? Is $X(t)$ wide-sense stationary?

9.65. Let $X(t)$ be defined by

$$X(t) = A \cos \omega t + B \sin \omega t,$$

where A and B are iid random variables.

(a) Under what conditions is $X(t)$ wide-sense stationary?

(b) Show that $X(t)$ is not stationary. *Hint:* Consider $E[X^3(t)]$.

9.66. Consider the following moving average process:

$$Y_n = 1/2(X_n + X_{n-1}) \qquad X_0 = 0.$$

(a) Is Y_n a stationary random process if X_n is an iid integer-valued process?

(b) Is Y_n a stationary random process if X_n is a stationary process?

(c) Are Y_n and X_n jointly stationary random processes if X_n is an iid process? a stationary process?

9.67. Let X_n be a zero-mean iid process, and let Z_n be an autoregressive random process

$$Z_n = 3/4 Z_{n-1} + X_n \qquad Z_0 = 0.$$

(a) Find the autocovariance of Z_n and determine whether Z_n is wide-sense stationary. *Hint:* Express Z_n in terms of $X_n, X_{n-1}, \ldots, X_1$.

(b) Does Z_n eventually settle down into stationary behavior?

(c) Find the pdf of Z_n if X_n is an iid sequence of zero-mean, unit-variance Gaussian random variables. What is the pdf of Z_n as $n \to \infty$?

9.68. Let $Y(t) = X(t + s) - \beta X(t)$, where $X(t)$ is a wide-sense stationary random process.

(a) Determine whether $Y(t)$ is also a wide-sense stationary random process.

(b) Find the cross-covariance function of $Y(t)$ and $X(t)$. Are the processes jointly wide-sense stationary?

(c) Find the pdf of $Y(t)$ if $X(t)$ is a Gaussian random process.

(d) Find the joint pdf of $Y(t_1)$ and $Y(t_2)$ in part c.

(e) Find the joint pdf of $Y(t_1)$ and $X(t_2)$ in part c.

9.69. Let $X(t)$ and $Y(t)$ be independent, wide-sense stationary random processes with zero means and the same covariance function $C_X(\tau)$. Let $Z(t)$ be defined by

$$Z(t) = 3X(t) - 5Y(t).$$

(a) Determine whether $Z(t)$ is also wide-sense stationary.

(b) Determine the pdf of $Z(t)$ if $X(t)$ and $Y(t)$ are also jointly Gaussian zero-mean random processes with $C_X(\tau) = 4e^{-|\tau|}$.

(c) Find the joint pdf of $Z(t_1)$ and $Z(t_2)$ in part b.

(d) Find the cross-covariance between $Z(t)$ and $X(t)$. Are $Z(t)$ and $X(t)$ jointly stationary random processes?

(e) Find the joint pdf of $Z(t_1)$ and $X(t_2)$ in part b. *Hint:* Use auxilliary variables.

9.70. Let $X(t)$ and $Y(t)$ be independent, wide-sense stationary random processes with zero means and the same covariance function $C_X(\tau)$. Let $Z(t)$ be defined by

$$Z(t) = X(t) \cos \omega t + Y(t) \sin \omega t.$$

(a) Determine whether $Z(t)$ is a wide-sense stationary random process.

(b) Determine the pdf of $Z(t)$ if $X(t)$ and $Y(t)$ are also jointly Gaussian zero-mean random processes with $C_X(\tau) = 4e^{-|\tau|}$.

(c) Find the joint pdf of $Z(t_1)$ and $Z(t_2)$ in part b.

(d) Find the cross-covariance between $Z(t)$ and $X(t)$. Are $Z(t)$ and $X(t)$ jointly stationary random processes?

(e) Find the joint pdf of $Z(t_1)$ and $X(t_2)$ in part b.

9.71. Let $X(t)$ be a zero-mean, wide-sense stationary Gaussian random process with autocorrelation function $R_X(\tau)$. The output of a "square law detector" is

$$Y(t) = X(t)^2.$$

Show that $R_Y(\tau) = R_X(0)^2 + 2R_X^2(\tau)$. *Hint:* For zero-mean, jointly Gaussian random variables $E[X^2Z^2] = E[X^2]E[Z^2] + 2E[XZ]^2$.

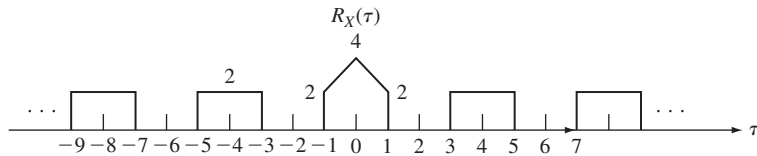**9.72.** A WSS process $X(t)$ has mean 1 and autocorrelation function given in Fig. P9.3.

FIGURE P9.3

(a) Find the mean component of $R_X(\tau)$.

(b) Find the periodic component of $R_X(\tau)$.

(c) Find the remaining component of $R_X(\tau)$.

9.73. Let X_n and Y_n be independent random processes. A multiplexer combines these two se-
quences into a combined sequence U_k, that is,

$$U_{2n} = X_n, \qquad U_{2n+1} = Y_n.$$

(a) Suppose that X_n and Y_n are independent Bernoulli random processes. Under
what conditions is U_k a stationary random process? a cyclostationary random
process?

(b) Repeat part a if X_n and Y_n are independent stationary random processes.

(c) Suppose that X_n and Y_n are wide-sense stationary random processes. Is U_k a wide-
sense stationary random process? a wide-sense cyclostationary random process?
Find the mean and autocovariance functions of U_k.

(d) If U_k is wide-sense cyclostationary, find the mean and correlation function of the
randomly phase-shifted version of U_k as defined by Eq. (9.72).

9.74. A ternary information source produces an iid, equiprobable sequence of symbols from
the alphabet $\{a, b, c\}$. Suppose that these three symbols are encoded into the respective
binary codewords 00, 01, 10. Let B_n be the sequence of binary symbols that result from
encoding the ternary symbols.

(a) Find the joint pmf of B_n and B_{n+1} for n even; n odd. Is B_n stationary? cyclostationary?

(b) Find the mean and covariance functions of B_n. Is B_n wide-sense stationary? wide-
sense cyclostationary?

(c) If B_n is cyclostationary, find the joint pmf, mean, and autocorrelation functions of the
randomly phase-shifted version of B_n as defined by Eq. (9.72).

9.75. Let $s(t)$ be a periodic square wave with period $T = 1$ which is equal to 1 for the first half
of a period and -1 for the remainder of the period. Let $X(t) = As(t)$, where A is a ran-
dom variable.

(a) Find the mean and autocovariance functions of $X(t)$.

(b) Is $X(t)$ a mean-square periodic process?

(c) Find the mean and autocovariance of $X_s(t)$ the randomly phase-shifted version of
$X(t)$ given by Eq. (9.72).

9.76. Let $X(t) = As(t)$ and $Y(t) = Bs(t)$, where A and B are independent random variables
that assume values $+1$ or -1 with equal probabilities, where $s(t)$ is the periodic square
wave in Problem 9.75.

(a) Find the joint pmf of $X(t_1)$ and $Y(t_2)$.

(b) Find the cross-covariance of $X(t_1)$ and $Y(t_2)$.

(c) Are $X(t)$ and $Y(t)$ jointly wide-sense cyclostationary? Jointly cyclostationary?

9.77. Let $X(t)$ be a mean square periodic random process. Is $X(t)$ a wide-sense cyclostationary
process?

9.78. Is the pulse amplitude modulation random process in Example 9.38 cyclostationary?

9.79. Let $X(t)$ be the random amplitude sinusoid in Example 9.37. Find the mean and autocor-
relation functions of the randomly phase-shifted version of $X(t)$ given by Eq. (9.72).

9.80. Complete the proof that if $X(t)$ is a cyclostationary random process, then $X_s(t)$, defined
by Eq. (9.72), is a stationary random process.

9.81. Show that if $X(t)$ is a wide-sense cyclostationary random process, then $X_s(t)$, defined by
Eq. (9.72), is a wide-sense stationary random process with mean and autocorrelation
functions given by Eqs. (9.74a) and (9.74b).

Section 9.7: Continuity, Derivatives, and Integrals of Random Processes

9.82. Let the random process $X(t) = u(t - S)$ be a unit step function delayed by an exponential random variable S, that is, $X(t) = 1$ for $t \geq S$, and $X(t) = 0$ for $t < S$.

    (a) Find the autocorrelation function of $X(t)$.

    (b) Is $X(t)$ mean square continuous?

    (c) Does $X(t)$ have a mean square derivative? If so, find its mean and autocorrelation functions.

    (d) Does $X(t)$ have a mean square integral? If so, find its mean and autocovariance functions.

9.83. Let $X(t)$ be the random telegraph signal introduced in Example 9.24.

    (a) Is $X(t)$ mean square continuous?

    (b) Show that $X(t)$ does not have a mean square derivative, and show that the second mixed partial derivative of its autocorrelation function has a delta function. What gives rise to this delta function?

    (c) Does $X(t)$ have a mean square integral? If so, find its mean and autocovariance functions.

9.84. Let $X(t)$ have autocorrelation function

$$R_X(\tau) = \sigma^2 e^{-\alpha \tau^2}.$$

    (a) Is $X(t)$ mean square continuous?

    (b) Does $X(t)$ have a mean square derivative? If so, find its mean and autocorrelation functions.

    (c) Does $X(t)$ have a mean square integral? If so, find its mean and autocorrelation functions.

    (d) Is $X(t)$ a Gaussian random process?

9.85. Let $N(t)$ be the Poisson process. Find $E[(N(t) - N(t_0))^2]$ and use the result to show that $N(t)$ is mean square continuous.

9.86. Does the pulse amplitude modulation random process discussed in Example 9.38 have a mean square integral? If so, find its mean and autocovariance functions.

9.87. Show that if $X(t)$ is a mean square continuous random process, then $X(t)$ has a mean square integral. *Hint:* Show that

$$R_X(t_1, t_2) - R_X(t_0, t_0) = E[(X(t_1) - X(t_0))X(t_2)] + E[X(t_0)(X(t_2) - X(t_0))],$$

and then apply the Schwarz inequality to the two terms on the right-hand side.

9.88. Let $Y(t)$ be the mean square integral of $X(t)$ in the interval $(0, t)$. Show that $Y'(t)$ is equal to $X(t)$ in the mean square sense.

9.89. Let $X(t)$ be a wide-sense stationary random process. Show that $E[X(t)X'(t)] = 0$.

9.90. A linear system with input $Z(t)$ is described by

$$X'(t) + \alpha X(t) = Z(t) \qquad t \geq 0, X(0) = 0.$$

Find the output $X(t)$ if the input is a zero-mean Gaussian random process with autocorrelation function given by

$$R_X(\tau) = \sigma^2 e^{-\beta|\tau|}.$$

Section 9.8: Time Averages of Random Processes and Ergodic Theorems

9.91. Find the variance of the time average given in Example 9.47.

9.92. Are the following processes WSS and mean ergodic?
  (a) Discrete-time dice process in Problem 9.2.
  (b) Alternating sign process in Problem 9.3.
  (c) $X_n = s^n$, for $n \geq 0$ in Problem 9.4.

9.93. Is the following WSS random process $X(t)$ mean ergodic?

$$R_X(\tau) = \begin{cases} 0 & |\tau| > 1 \\ 5(1 - |\tau|) & |\tau| \leq 1. \end{cases}$$

9.94. Let $X(t) = A \cos(2\pi f t)$, where A is a random variable with mean m and variance σ^2.
  (a) Evaluate $<X(t)>_T$, find its limit as $T \to \infty$, and compare to $m_X(t)$.
  (b) Evaluate $<X(t + \tau)X(t)>$, find its limit as $T \to \infty$, and compare to $R_X(t + \tau, t)$.

9.95. Repeat Problem 9.94 with $X(t) = A \cos(2\pi f t + \Theta)$, where A is as in Problem 9.94, Θ is a random variable uniformly distributed in $(0, 2\pi)$, and A and Θ are independent random variables.

9.96. Find an exact expression for $\text{VAR}[<X(t)>_T]$ in Example 9.48. Find the limit as $T \to \infty$.

9.97. The WSS random process X_n has mean m and autocovariance $C_X(k) = (1/2)^{|k|}$. Is X_n mean ergodic?

9.98. (a) Are the moving average processes Y_n in Problem 9.24 mean ergodic?
  (b) Are the autoregressive processes Z_n in Problem 9.25a mean ergodic?

9.99. (a) Show that a WSS random process is mean ergodic if

$$\int_{-\infty}^{\infty} |C(u)| < \infty.$$

  (b) Show that a discrete-time WSS random process is mean ergodic if

$$\sum_{k=-\infty}^{\infty} |C(k)| < \infty.$$

9.100. Let $<X^2(t)>_T$ denote a time-average estimate for the mean power of a WSS random process.
  (a) Under what conditions is this time average a valid estimate for $E[X^2(t)]$?
  (b) Apply your result in part a for the random phase sinusoid in Example 9.2.

9.101. (a) Under what conditions is the time average $<X(t + \tau)X(t)>_T$ a valid estimate for the autocorrelation $R_X(\tau)$ of a WSS random process $X(t)$?
  (b) Apply your result in part a for the random phase sinusoid in Example 9.2.

9.102. Let $Y(t)$ be the indicator function for the event $\{a < X(t) \leq b\}$, that is,

$$Y(t) = \begin{cases} 1 & \text{if } X(t) \in (a, b] \\ 0 & \text{otherwise.} \end{cases}$$

  (a) Show that $<Y(t)>_T$ is the proportion of time in the time interval $(-T, T)$ that $X(t) \in (a, b]$.

    (b) Find $E[<Y(t)>_T]$.

    (c) Under what conditions does $<Y(t)>_T \rightarrow P[a < X(t) \leq b]$?

    (d) How can $<Y(t)>_T$ be used to estimate $P[X(t) \leq x]$?

    (e) Apply the result in part d to the random telegraph signal.

9.103. **(a)** Repeat Problem 9.102 for the time average of the discrete-time Y_n, which is defined as the indicator for the event $\{a < X_n \leq b]\}$.

    (b) Apply your result in part a to an iid discrete-valued random process.

    (c) Apply your result in part a to an iid continuous-valued random process.

9.104. For $n \geq 1$, define $Z_n = u(a - X_n)$, where $u(x)$ is the unit step function, that is, $X_n = 1$ if and only if $X_n \leq a$.

    (a) Show that the time average $<Z_n>_N$ is the proportion of X_n's that are less than a in the first N samples.

    (b) Show that if the process is ergodic (in some sense), then this time average is equal to $F_X(a) = P[X \leq a]$.

9.105. In Example 9.50 show that $\mathrm{VAR}[\langle X_n \rangle_T] = (\sigma^2)(2T + 1)^{2H-2}$.

9.106. Plot the covariance function vs. k for the self-similar process in Example 9.50 with $\sigma^2 = 1$ for: $H = 0.5, H = 0.6, H = 0.75, H = 0.99$. Does the long-range dependence of the process increase or decrease with H?

9.107. **(a)** Plot the variance of the sample mean given by Eq. (9.110) vs. T with $\sigma^2 = 1$ for: $H = 0.5, H = 0.6, H = 0.75, H = 0.99$.

    (b) For the parameters in part a, plot $(2T + 1)^{2H-1}$ vs. T, which is the ratio of the variance of the sample mean of a long-range dependent process relative to the variance of the sample mean of an iid process. How does the long-range dependence manifest itself, especially for H approaching 1?

    (c) Comment on the width of confidence intervals for estimates of the mean of long-range dependent processes relative to those of iid processes.

9.108. Plot the variance of the sample mean for a long-range dependent process (Eq. 9.110) vs. the sample size T in a log-log plot.

    (a) What role does H play in the plot?

    (b) One of the remarkable indicators of long-range dependence in nature comes from a set of observations of the minimal water levels in the Nile river for the years 622–1281 [Beran, p. 22] where the log-log plot for part a gives a slope of -0.27. What value of H corresponds to this slope?

9.109. Problem 9.99b gives a sufficient condition for mean ergodicity for discrete-time random processes. Use the expression in Eq. (9.112) for a long-range dependent process to determine whether the sufficient condition is satisfied. Comment on your findings.

*Section 9.9: Fourier Series and Karhunen-Loeve Expansion

9.110. Let $X(t) = Xe^{j\omega t}$ where X is a random variable.

    (a) Find the correlation function for $X(t)$, which for complex-valued random processes is defined by $R_X(t_1, t_2) = E[X(t_1)X^*(t_2)]$, where * denotes the complex conjugate.

    (b) Under what conditions is $X(t)$ a wide-sense stationary random process?

9.111. Consider the sum of two complex exponentials with random coefficients:

$$X(t) = X_1 e^{j\omega_1 t} + X_2 e^{j\omega_2 t} \quad \text{where } \omega_1 \neq \omega_2.$$

    (a) Find the covariance function of $X(t)$.

    (b) Find conditions on the complex-valued random variables X_1, and X_2 for $X(t)$ to be a wide-sense stationary random process.

    (c) Show that if we let $\omega_1 = -\omega_2$, $X_1 = (U - jV)/2$ and $X_2 = (U + jV)/2$, where U and V are real-valued random variables, then $X(t)$ is a real-valued random process. Find an expression for $X(t)$ and for the autocorrelation function.

    (d) Restate the conditions on X_1 and X_2 from part b in terms of U and V.

    (e) Suppose that in part c, U and V are jointly Gaussian random variables. Show that $X(t)$ is a Gaussian random process.

9.112. (a) Derive Eq. (9.118) for the correlation of the Fourier coefficients for a non-mean square periodic process $X(t)$.

    (b) Show that Eq. (9.118) reduces to Eq. (9.117) when $X(t)$ is WSS and mean square periodic.

9.113. Let $X(t)$ be a WSS Gaussian random process with $R_X(\tau) = e^{-|\tau|}$.

    (a) Find the Fourier series expansion for $X(t)$ in the interval $[0, T]$.

    (b) What is the distribution of the coefficients in the Fourier series?

9.114. Show that the Karhunen-Loeve expansion of a WSS mean-square periodic process $X(t)$ yields its Fourier series. Specify the orthonormal set of eigenfunctions and the corresponding eigenvalues.

9.115. Let $X(t)$ be the white Gaussian noise process introduced in Example 9.43. Show that any set of orthonormal functions can be used as the eigenfunctions for $X(t)$ in its Karhunen-Loeve expansion. What are the eigenvalues?

9.116. Let $Y(t) = X(t) + W(t)$, where $X(t)$ and $W(t)$ are orthogonal random processes and $W(t)$ is a white Gaussian noise process. Let $\phi_n(t)$ be the eigenfunctions corresponding to $K_X(t_1, t_2)$. Show that $\phi_n(t)$ are also the eigenfunctions for $K_Y(t_1, t_2)$. What is the relation between the eigenvalues of $K_X(t_1, t_2)$ and those of $K_Y(t_1, t_2)$?

9.117. Let $X(t)$ be a zero-mean random process with autocovariance

$$R_X(\tau) = \sigma^2 e^{-\alpha|\tau|}.$$

    (a) Write the eigenvalue integral equation for the Karhunen-Loeve expansion of $X(t)$ on the interval $[-T, T]$.

    (b) Differentiate the above integral equation to obtain the differential equation

$$\frac{d^2}{dt^2}\phi(t) = \frac{\alpha^2\left(\lambda - 2\dfrac{\sigma^2}{\alpha}\right)}{\lambda}\phi(t).$$

    (c) Show that the solutions to the above differential equation are of the form $\phi(t) = A \cos bt$ and $\phi(t) = B \sin bt$. Find an expression for b.

(d) Substitute the $\phi(t)$ from part c into the integral equation of part a to show that if $\phi(t) = A \cos bt$, then b is the root of $\tan bT = \alpha/b$, and if $\phi(t) = B \sin bt$, then b is the root of $\tan bT = -b/\alpha$.

(e) Find the values of A and B that normalize the eigenfunctions.

*(f) In order to show that the frequencies of the eigenfunctions are not harmonically related, plot the following three functions versus bT: $\tan bT$, $bT/\alpha T$, $-\alpha T/bT$. The intersections of these functions yield the eigenvalues. Note that there are two roots per interval of length π.

*Section 9.10: Generating Random Processes

9.118. (a) Generate 10 realizations of the binomial counting process with $p = 1/4$, $p = 1/2$, and $p = 3/4$. For each value of p, plot the sample functions for $n = 200$ trials.

(b) Generate 50 realizations of the binomial counting process with $p = 1/2$. Find the sample mean and sample variance of the realizations for the first 200 trials.

(c) In part b, find the histogram of increments in the process for the interval $[1, 50]$, $[51, 100]$, $[101, 150]$, and $[151, 200]$. Compare these histograms to the theoretical pmf. How would you check to see if the increments in the four intervals are stationary?

(d) Plot a scattergram of the pairs consisting of the increments in the interval $[1, 50]$ and $[51, 100]$ in a given realization. Devise a test to check whether the increments in the two intervals are independent random variables.

9.119. Repeat Problem 9.118 for the random walk process with the same parameters.

9.120. Repeat Problem 9.118 for the sum process in Eq. (9.24) where the X_n are iid unit-variance Gaussian random variables with mean: $m = 0$; $m = 0.5$.

9.121. Repeat Problem 9.118 for the sum process in Eq. (9.24) where the X_n are iid Poisson random variables with $\alpha = 1$.

9.122. Repeat Problem 9.118 for the sum process in Eq. (9.24) where the X_n are iid Cauchy random variables with $\alpha = 1$.

9.123. Let $Y_n = \alpha Y_{n-1} + X_n$ where $Y_0 = 0$.

(a) Generate five realizations of the process for $\alpha = 1/4$, $1/2$, $9/10$ and with X_n given by the $p = 1/2$ and $p = 1/4$ random step process. Plot the sample functions for the first 200 steps. Find the sample mean and sample variance for the outcomes in *each* realization. Plot the histogram for outcomes in each realization.

(b) Generate 50 realizations of the process Y_n with $\alpha = 1/2$, $p = 1/4$, and $p = 1/2$. Find the sample mean and sample variance of the realizations for the first 200 trials. Find the histogram of Y_n across the realizations at times $n = 5$, $n = 50$, and $n = 200$.

(c) In part b, find the histogram of increments in the process for the interval $[1, 50]$, $[51, 100]$, $[101, 150]$, and $[151, 200]$. To what theoretical pmf should these histograms be compared? Should the increments in the process be stationary? Should the increments be independent?

9.124. Repeat Problem 9.123 for the sum process in Eq. (9.24) where the X_n are iid unit-variance Gaussian random variables with mean: $m = 0$; $m = 0.5$.

9.125. (a) Propose a method for estimating the covariance function of the sum process in Problem 9.118. Do not assume that the process is wide-sense stationary.

(b) How would you check to see if the process is wide-sense stationary?

(c) Apply the methods in parts a and b to the experiment in Problem 9.118b.

(d) Repeat part c for Problem 9.123b.

9.126. Use the binomial process to approximate a Poisson random process with arrival rate $\lambda = 1$ customer per second in the time interval (0, 100]. Try different values of n and come up with a recommendation on how n should be selected.

9.127. Generate 100 repetitions of the experiment in Example 9.21.

(a) Find the relative frequency of the event $P[N(10) = 3$ and $N(60) - N(45) = 2]$ and compare it to the theoretical probability.

(b) Find the histogram of the time that elapses until the second arrival and compare it to the theoretical pdf. Plot the empirical cdf and compare it to the theoretical cdf.

9.128. Generate 100 realizations of the Poisson random process $N(t)$ with arrival rate $\lambda = 1$ customer per second in the time interval (0, 10]. Generate the pair $(N_1(t), N_2(t))$ by assigning arrivals in $N(t)$ to $N_1(t)$ with probability $p = 0.25$ and to $N_2(t)$ with probability 0.75.

(a) Find the histograms for $N_1(10)$ and $N_2(10)$ and compare them to the theoretical pmf by performing a chi-square goodness-of-fit test at a 5% significance level.

(b) Perform a chi-square goodness-of-fit test to test whether $N_1(10)$ and $N_2(10)$ are independent random variables. How would you check whether $N_1(t)$ and $N_2(t)$ are independent random processes?

9.129. Subscribers log on to a system according to a Poisson process with arrival rate $\lambda = 1$ customer per second. The ith customer remains logged on for a random duration of T_i seconds, where the T_i are iid random variables and are also independent of the arrival times.

(a) Generate the sequence S_n of customer arrival times and the corresponding departure times given by $D_n = S_n + T_n$, where the connections times are all equal to 1.

(b) Plot: $A(t)$, the number of arrivals up to time t; $D(t)$, the number of departures up to time t; and $N(t) = A(t) - D(t)$, the number in the system at time t.

(c) Perform 100 simulations of the system operation for a duration of 200 seconds. Assume that customer connection times are an exponential random variables with mean 5 seconds. Find the customer departure time instants and the associated departure counting process $D(t)$. How would you check whether $D(t)$ is a Poisson process? Find the histograms for $D(t)$ and the number in the system $N(t)$ at $t = 50, 100, 150, 200$. Try to fit a pmf to each histogram.

(d) Repeat part c if customer connection times are exactly 5 seconds long.

9.130. Generate 100 realizations of the Wiener process with $\alpha = 1$ for the interval (0, 3.5) using the random walk limiting procedure.

(a) Find the histograms for increments in the intervals (0, 0.5], (0.5, 1.5], and (1.5, 3.5] and compare these to the theoretical pdf.

(b) Perform a test at a 5% significance level to determine whether the increments in the first two intervals are independent random variables.

9.131. Repeat Problem 9.130 using Gaussian-distributed increments to generate the Wiener process. Discuss how the increment interval in the simulation should be selected.

Problems Requiring Cumulative Knowledge

9.132. Let $X(t)$ be a random process with independent increments. Assume that the increments $X(t_2) - X(t_1)$ are gamma random variables with parameters $\lambda > 0$ and $\alpha = t_2 - t_1$.
   (a) Find the joint density function of $X(t_1)$ and $X(t_2)$.
   (b) Find the autocorrelation function of $X(t)$.
   (c) Is $X(t)$ mean square continuous?
   (d) Does $X(t)$ have a mean square derivative?

9.133. Let $X(t)$ be the pulse amplitude modulation process introduced in Example 9.38 with $T = 1$. A phase-modulated process is defined by

$$Y(t) = a\cos\left(2\pi t + \frac{\pi}{2}X(t)\right).$$

   (a) Plot the sample function of $Y(t)$ corresponding to the binary sequence 0010110.
   (b) Find the joint pdf of $Y(t_1)$ and $Y(t_2)$.
   (c) Find the mean and autocorrelation functions of $Y(t)$.
   (d) Is $Y(t)$ a stationary, wide-sense stationary, or cyclostationary random process?
   (e) Is $Y(t)$ mean square continuous?
   (f) Does $Y(t)$ have a mean square derivative? If so, find its mean and autocorrelation functions.

9.134. Let $N(t)$ be the Poisson process, and suppose we form the phase-modulated process

$$Y(t) = a\cos(2\pi f t + \pi N(t)).$$

   (a) Plot a sample function of $Y(t)$ corresponding to a typical sample function of $N(t)$.
   (b) Find the joint density function of $Y(t_1)$ and $Y(t_2)$. *Hint:* Use the independent increments property of $N(t)$.
   (c) Find the mean and autocorrelation functions of $Y(t)$.
   (d) Is $Y(t)$ a stationary, wide-sense stationary, or cyclostationary random process?
   (e) Is $Y(t)$ mean square continuous?
   (f) Does $Y(t)$ have a mean square derivative? If so, find its mean and autocorrelation functions.

9.135. Let $X(t)$ be a train of amplitude-modulated pulses with occurrences according to a Poisson process:

$$X(t) = \sum_{k=1}^{\infty} A_k h(t - S_k),$$

where the A_k are iid random variables, the S_k are the event occurrence times in a Poisson process, and $h(t)$ is a function of time. Assume the amplitudes and occurrence times are independent.
   (a) Find the mean and autocorrelation functions of $X(t)$.
   (b) Evaluate part a when $h(t) = u(t)$, a unit step function.
   (c) Evaluate part a when $h(t) = p(t)$, a rectangular pulse of duration T seconds.

9.136. Consider a linear combination of two sinusoids:

$$X(t) = A_1 \cos(\omega_0 t + \Theta_1) + A_2 \cos(\sqrt{2}\omega_0 t + \Theta_2),$$

where Θ_1 and Θ_2 are independent uniform random variables in the interval $(0, 2\pi)$, and A_1 and A_2 are jointly Gaussian random variables. Assume that the amplitudes are independent of the phase random variables.

  (a) Find the mean and autocorrelation functions of $X(t)$.

  (b) Is $X(t)$ mean square periodic? If so, what is the period?

  (c) Find the joint pdf of $X(t_1)$ and $X(t_2)$.

9.137. (a) A *Gauss-Markov random process* is a Gaussian random process that is also a Markov process. Show that the autocovariance function of such a process must satisfy

$$C_X(t_3, t_1) = \frac{C_X(t_3, t_2) C_X(t_2, t_1)}{C_X(t_2, t_2)},$$

  where $t_1 \leq t_2 \leq t_3$.

  (b) It can be shown that if the autocovariance of a Gaussian random process satisfies the above equation, then the process is Gauss-Markov. Is the Wiener process Gauss-Markov? Is the Ornstein-Uhlenbeck process Gauss-Markov?

9.138. Let A_n and B_n be two independent stationary random processes. Suppose that A_n and B_n are zero-mean, Gaussian random processes with autocorrelation functions

$$R_A(k) = \sigma_1^2 \rho_1^{|k|} \qquad R_B(k) = \sigma_2^2 \rho_2^{|k|}.$$

A block multiplexer takes blocks of two from the above processes and interleaves them to form the random process Y_m:

$$A_1 A_2 B_1 B_2 A_3 A_4 B_3 B_4 A_5 A_6 B_5 B_6 \ldots.$$

  (a) Find the autocorrelation function of Y_m.

  (b) Is Y_m cyclostationary? wide-sense stationary?

  (c) Find the joint pdf of Y_m and Y_{m+1}.

  (d) Let $Z_m = Y_{m+T}$, where T is selected uniformly from the set $\{0, 1, 2, 3\}$. Repeat parts a, b, and c for Z_m.

9.139. Let A_n be the Gaussian random process in Problem 9.138. A decimator takes every other sample to form the random process V_m:

$$A_1 A_3 A_5 A_7 A_9 A_{11}$$

  (a) Find the autocorrelation function of V_m.

  (b) Find the joint pdf of V_m and V_{m+k}.

  (c) An interpolator takes the sequence V_m and inserts zeros between samples to form the sequence W_k:

$$A_1 0 A_3 0 A_5 0 A_7 0 A_9 0 A_{11} \ldots.$$

  Find the autocorrelation function of W_k. Is W_k a Gaussian random process?

9.140. Let A_n be a sequence of zero-mean, unit-variance independent Gaussian random variables. A block coder takes pairs of A's and linearly transforms them to form the sequence Y_n:

$$\begin{bmatrix} Y_{2n} \\ Y_{2n+1} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} A_{2n} \\ A_{2n+1} \end{bmatrix}.$$

 (a) Find the autocorrelation function of Y_n.
 (b) Is Y_n stationary in any sense?
 (c) Find the joint pdf of Y_n, Y_{n+1}, and Y_{n+2}.

9.141. Suppose customer orders arrive according to a Bernoulli random process with parameter p. When an order arrives, its size is an exponential random variable with parameter λ. Let S_n be the total size of all orders up to time n.

 (a) Find the mean and autocorrelation functions of S_n.
 (b) Is S_n a stationary random process?
 (c) Is S_n a Markov process?
 (d) Find the joint pdf of S_n and S_{n+k}.

Analysis and Processing of Random Signals

In this chapter we introduce methods for analyzing and processing random signals. We cover the following topics:

- Section 10.1 introduces the notion of power spectral density, which allows us to view random processes in the frequency domain.
- Section 10.2 discusses the response of linear systems to random process inputs and introduce methods for filtering random processes.
- Section 10.3 considers two important applications of signal processing: sampling and modulation.
- Sections 10.4 and 10.5 discuss the design of optimum linear systems and introduce the Wiener and Kalman filters.
- Section 10.6 addresses the problem of estimating the power spectral density of a random process.
- Finally, Section 10.7 introduces methods for implementing and simulating the processing of random signals.

10.1 POWER SPECTRAL DENSITY

The Fourier series and the Fourier transform allow us to view deterministic time functions as the weighted sum or integral of sinusoidal functions. A time function that varies slowly has the weighting concentrated at the low-frequency sinusoidal components. A time function that varies rapidly has the weighting concentrated at higher-frequency components. Thus the rate at which a deterministic time function varies is related to the weighting function of the Fourier series or transform. This weighting function is called the "spectrum" of the time function.

The notion of a time function as being composed of sinusoidal components is also very useful for random processes. However, since a sample function of a random process can be viewed as being selected from an ensemble of allowable time functions, the weighting function or "spectrum" for a random process must refer in some way to the average rate of change of the ensemble of allowable time functions. Equation (9.66) shows that, for wide-sense stationary processes, the autocorrelation function

$R_X(\tau)$ is an appropriate measure for the average rate of change of a random process. Indeed if a random process changes slowly with time, then it remains correlated with itself for a long period of time, and $R_X(\tau)$ decreases slowly as a function of τ. On the other hand, a rapidly varying random process quickly becomes uncorrelated with itself, and $R_X(\tau)$ decreases rapidly with τ.

We now present the **Einstein-Wiener-Khinchin theorem**, which states that the power spectral density of a wide-sense stationary random process is given by the Fourier transform of the autocorrelation function.[1]

10.1.1   Continuous-Time Random Processes

Let $X(t)$ be a continuous-time WSS random process with mean m_X and autocorrelation function $R_X(\tau)$. Suppose we take the Fourier transform of a sample of $X(t)$ in the interval $0 < t < T$ as follows

$$\widetilde{x}(f) = \int_0^T X(t')e^{-j2\pi f t'}\, dt'. \tag{10.1}$$

We then approximate the power density as a function of frequency by the function:

$$\widetilde{p}_T(f) = \frac{1}{T}|\widetilde{x}(f)|^2 = \frac{1}{T}\widetilde{x}(f)\widetilde{x}^*(f) = \frac{1}{T}\left\{\int_0^T X(t')e^{-j2\pi f t'}\, dt'\right\}\left\{\int_0^T X(t')e^{j2\pi f t'}\, dt'\right\}, \tag{10.2}$$

where * denotes the complex conjugate. X(t) is a random process, so $\widetilde{p}_T(f)$ is also a random process but over a different index set. $\widetilde{p}_T(f)$ is called the **periodogram estimate** and we are interested in the **power spectral density** of $X(t)$ which is defined by:

$$S_X(f) = \lim_{T\to\infty} E[\widetilde{p}_T(f)] = \lim_{T\to\infty} \frac{1}{T}E[|\widetilde{x}(f)|^2]. \tag{10.3}$$

We show at the end of this section that *the power spectral density of X(t) is given by the Fourier transform of the autocorrelation function:*

$$S_X(f) = \mathcal{F}\{R_X(\tau)\} = \int_{-\infty}^{\infty} R_X(\tau)e^{-j2\pi f\tau}\, d\tau. \tag{10.4}$$

A table of Fourier transforms and its properties is given in Appendix B.

For real-valued random processes, the autocorrelation function is an even function of τ:

$$R_X(\tau) = R_X(-\tau). \tag{10.5}$$

---

[1]This result is usually called the Wiener-Khinchin theorem, after Norbert Wiener and A. Ya. Khinchin, who proved the result in the early 1930s. Later it was discovered that this result was stated by Albert Einstein in a 1914 paper (see Einstein).

Substitution into Eq. (10.4) implies that

$$S_X(f) = \int_{-\infty}^{\infty} R_X(\tau)\{\cos 2\pi f\tau - j\sin 2\pi f\tau\}\, d\tau$$

$$= \int_{-\infty}^{\infty} R_X(\tau)\cos 2\pi f\tau\, d\tau, \tag{10.6}$$

since the integral of the product of an even function $(R_X(\tau))$ and an odd function $(\sin 2\pi f\tau)$ is zero. Equation (10.6) implies that $S_X(f)$ *is real-valued and an even function of f.* From Eq. (10.2) we have that $S_X(f)$ *is nonnegative*:

$$S_X(f) \geq 0 \qquad \text{for all } f. \tag{10.7}$$

The autocorrelation function can be recovered from the power spectral density by applying the inverse Fourier transform formula to Eq. (10.4):

$$R_X(\tau) = \mathscr{F}^{-1}\{S_X(f)\}$$

$$= \int_{-\infty}^{\infty} S_X(f)e^{j2\pi f\tau}\, df. \tag{10.8}$$

Equation (10.8) is identical to Eq. (4.80), which relates the pdf to its corresponding characteristic function. The last section in this chapter discusses how the FFT can be used to perform numerical calculations for $S_X(f)$ and $R_X(\tau)$.

In electrical engineering it is customary to refer to the second moment of $X(t)$ as the **average power of $X(t)$**.[2] Equation (10.8) together with Eq. (9.64) gives

$$E[X^2(t)] = R_X(0) = \int_{-\infty}^{\infty} S_X(f)\, df. \tag{10.9}$$

Equation (10.9) states that the average power of $X(t)$ is obtained by integrating $S_X(f)$ over all frequencies. This is consistent with the fact that $S_X(f)$ is the "density of power" of $X(t)$ at the frequency f.

Since the autocorrelation and autocovariance functions are related by $R_X(\tau) = C_X(\tau) + m_X^2$, the power spectral density is also given by

$$S_X(f) = \mathscr{F}\{C_X(\tau) + m_X^2\}$$

$$= \mathscr{F}\{C_X(\tau)\} + m_X^2\, \delta(f), \tag{10.10}$$

where we have used the fact that the Fourier transform of a constant is a delta function. We say the m_X is the "dc" component of $X(t)$.

The notion of power spectral density can be generalized to two jointly wide-sense stationary processes. The **cross-power spectral density $S_{X,Y}(f)$** is defined by

$$S_{X,Y}(f) = \mathscr{F}\{R_{X,Y}(\tau)\}, \tag{10.11}$$

---

[2]If $X(t)$ is a voltage or current developed across a 1-ohm resistor, then $X^2(t)$ is the instantaneous power absorbed by the resistor.
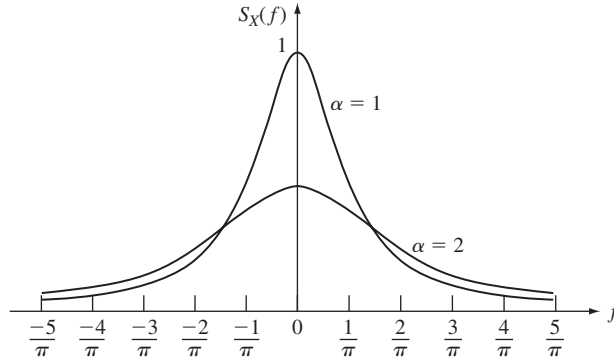
FIGURE 10.1
Power spectral density of a random telegraph signal with $\alpha = 1$ and
$\alpha = 2$ transitions per second.

where $R_{X,Y}(\tau)$ is the cross-correlation between $X(t)$ and $Y(t)$:

$$R_{X,Y}(\tau) = E[X(t + \tau)Y(t)]. \tag{10.12}$$

In general, $S_{X,Y}(f)$ is a complex function of f even if $X(t)$ and $Y(t)$ are both real-valued.

Example 10.1   Random Telegraph Signal

Find the power spectral density of the random telegraph signal.

In Example 9.24, the autocorrelation function of the random telegraph process was found to be

$$R_X(\tau) = e^{-2\alpha|\tau|},$$

where α is the average transition rate of the signal. Therefore, the power spectral density of the process is

$$S_X(f) = \int_{-\infty}^{0} e^{2\alpha\tau} e^{-j2\pi f\tau}\, d\tau + \int_{0}^{\infty} e^{-2\alpha\tau} e^{-j2\pi f\tau}\, d\tau$$

$$= \frac{1}{2\alpha - j2\pi f} + \frac{1}{2\alpha + j2\pi f}$$

$$= \frac{4\alpha}{4\alpha^2 + 4\pi^2 f^2}. \tag{10.13}$$

Figure 10.1 shows the power spectral density for $\alpha = 1$ and $\alpha = 2$ transitions per second. The process changes two times more quickly when $\alpha = 2$; it can be seen from the figure that the power spectral density for $\alpha = 2$ has greater high-frequency content.

Example 10.2   Sinusoid with Random Phase

Let $X(t) = a \cos(2\pi f_0 t + \Theta)$, where Θ is uniformly distributed in the interval $(0, 2\pi)$. Find $S_X(f)$.

From Example 9.10, the autocorrelation for $X(t)$ is

$$R_X(\tau) = \frac{a^2}{2}\cos 2\pi f_0 \tau.$$

Thus, the power spectral density is

$$S_X(f) = \frac{a^2}{2}\mathcal{F}\{\cos 2\pi f_0 \tau\}$$

$$= \frac{a^2}{4}\delta(f - f_0) + \frac{a^2}{4}\delta(f + f_0), \tag{10.14}$$

where we have used the table of Fourier transforms in Appendix B. The signal has average power $R_X(0) = a^2/2$. All of this power is concentrated at the frequencies $\pm f_0$, so the power density at these frequencies is infinite.

Example 10.3 White Noise

The power spectral density of a WSS white noise process whose frequency components are limited to the range $-W \leq f \leq W$ is shown in Fig. 10.2(a). The process is said to be "white" in analogy to white light, which contains all frequencies in equal amounts. The average power in this
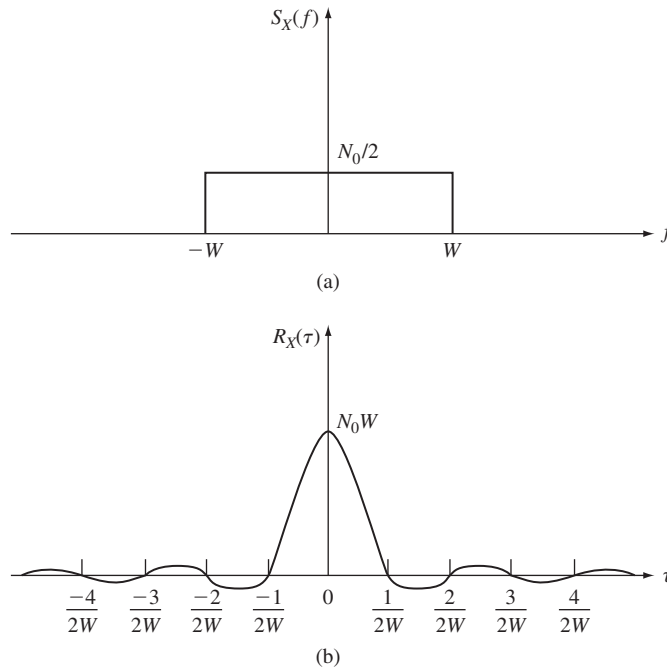
FIGURE 10.2
Bandlimited white noise: (a) power spectral density, (b) autocorrelation function.

process is obtained from Eq. (10.9):

$$E[X^2(t)] = \int_{-W}^{W} \frac{N_0}{2} df = N_0 W. \tag{10.15}$$

The autocorrelation for this process is obtained from Eq. (10.8):

$$\begin{aligned} R_X(\tau) &= \frac{1}{2} N_0 \int_{-W}^{W} e^{j2\pi f\tau} \, df \\ &= \frac{1}{2} N_0 \frac{e^{-j2\pi W\tau} - e^{j2\pi W\tau}}{-j2\pi\tau} \\ &= \frac{N_0 \sin(2\pi W_\tau)}{2\pi\tau}. \end{aligned} \tag{10.16}$$

$R_X(\tau)$ is shown in Fig. 10.2(b). Note that $X(t)$ and $X(t + \tau)$ are uncorrelated at $\tau = \pm k/2W$, $k = 1, 2, \ldots$.

The term white noise usually refers to a random process $W(t)$ whose power spectral density is $N_0/2$ for *all* frequencies:

$$S_W(f) = \frac{N_0}{2} \qquad \text{for all } f. \tag{10.17}$$

Equation (10.15) with $W = \infty$ shows that such a process must have infinite average power. By taking the limit $W \to \infty$ in Eq. (10.16), we find that the autocorrelation of such a process approaches

$$R_W(\tau) = \frac{N_0}{2} \delta(\tau). \tag{10.18}$$

If $W(t)$ is a Gaussian random process, we then see that $W(t)$ is the white Gaussian noise process introduced in Example 9.43 with $\alpha = N_0/2$.

Example 10.4   Sum of Two Processes

Find the power spectral density of $Z(t) = X(t) + Y(t)$, where $X(t)$ and $Y(t)$ are jointly WSS processes.

The autocorrelation of $Z(t)$ is

$$\begin{aligned} R_Z(\tau) &= E[Z(t + \tau)Z(t)] = E[(X(t + \tau) + Y(t + \tau))(X(t) + Y(t))] \\ &= R_X(\tau) + R_{YX}(\tau) + R_{XY}(\tau) + R_Y(\tau). \end{aligned}$$

The power spectral density is then

$$\begin{aligned} S_Z(f) &= \mathscr{F}\{R_X(\tau) + R_{YX}(\tau) + R_{XY}(\tau) + R_Y(\tau)\} \\ &= S_X(f) + S_{YX}(f) + S_{XY}(f) + S_Y(f). \end{aligned} \tag{10.19}$$

Example 10.5

Let $Y(t) = X(t - d)$, where d is a constant delay and where $X(t)$ is WSS. Find $R_{YX}(\tau)$, $S_{YX}(f)$, $R_Y(\tau)$, and $S_Y(f)$.

The definitions of $R_{YX}(\tau)$, $S_{YX}(f)$, and $R_Y(\tau)$ give

$$R_{YX}(\tau) = E[Y(t + \tau)X(t)] = E[X(t + \tau - d)X(t)] = R_X(\tau - d). \qquad (10.20)$$

The time-shifting property of the Fourier transform gives

$$S_{YX}(f) = \mathcal{F}\{R_X(\tau - d)\} = S_X(f)e^{-j2\pi fd}$$

$$= S_X(f)\cos(2\pi fd) - jS_X(f)\sin(2\pi fd). \qquad (10.21)$$

Finally,

$$R_Y(\tau) = E[Y(t + \tau)Y(t)] = E[X(t + \tau - d)X(t - d)] = R_X(\tau). \qquad (10.22)$$

Equation (10.22) implies that

$$S_Y(f) = \mathcal{F}\{R_Y(\boldsymbol{\tau})\} = \mathcal{F}\{R_X(\boldsymbol{\tau})\} = S_X(f). \qquad (10.23)$$

Note from Eq. (10.21) that the cross-power spectral density is complex. Note from Eq. (10.23) that $S_X(f) = S_Y(f)$ despite the fact that $X(t) \neq Y(t)$. Thus, $S_X(f) = S_Y(f)$ *does not imply that* $X(t) = Y(t)$.

10.1.2   Discrete-Time Random Processes

Let X_n be a discrete-time WSS random process with mean m_X and autocorrelation function $R_X(k)$. The **power spectral density of X_n** is defined as the Fourier transform of the autocorrelation sequence

$$S_X(f) = \mathcal{F}\{R_X(k)\}$$

$$= \sum_{k=-\infty}^{\infty} R_X(k)e^{-j2\pi fk}. \qquad (10.24)$$

Note that we need only consider frequencies in the range $-1/2 < f \leq 1/2$, since $S_X(f)$ is periodic in f with period 1. As in the case of continuous random processes, $S_X(f)$ can be shown to be a real-valued, nonnegative, even function of f.

The inverse Fourier transform formula applied to Eq. (10.23) implies that[3]

$$R_X(k) = \int_{-1/2}^{1/2} S_X(f)e^{j2\pi fk} \, df. \qquad (10.25)$$

Equations (10.24) and (10.25) are similar to the discrete Fourier transform. In the last section we show how to use the FFT to calculate $S_X(f)$ and $R_X(k)$.

The **cross-power spectral density $S_{X,Y}(f)$** of two jointly WSS discrete-time processes X_n and Y_n is defined by

$$S_{X,Y}(f) = \mathcal{F}\{R_{X,Y}(k)\}, \qquad (10.26)$$

where $R_{X,Y}(k)$ is the cross-correlation between X_n and Y_n:

$$R_{X,Y}(k) = E[X_{n+k}Y_n]. \qquad (10.27)$$

---

[3]You can view $R_X(k)$ as the coefficients of the Fourier series of the periodic function $S_X(f)$.

Example 10.6   White Noise

Let the process X_n be a sequence of uncorrelated random variables with zero mean and variance σ_X^2. Find $S_X(f)$.

The autocorrelation of this process is

$$R_X(k) = \begin{cases} \sigma_X^2 & k = 0 \\ 0 & k \neq 0. \end{cases}$$

The power spectral density of the process is found by substituting $R_X(k)$ into Eq. (10.24):

$$S_X(f) = \sigma_X^2 \qquad -\frac{1}{2} < f < \frac{1}{2}. \tag{10.28}$$

Thus the process X_n contains all possible frequencies in equal measure.

Example 10.7   Moving Average Process

Let the process Y_n be defined by

$$Y_n = X_n + \alpha X_{n-1}, \tag{10.29}$$

where X_n is the white noise process of Example 10.6. Find $S_Y(f)$.

It is easily shown that the mean and autocorrelation of Y_n are given by

$$E[Y_n] = 0,$$

and

$$E[Y_n Y_{n+k}] = \begin{cases} (1 + \alpha^2)\sigma_X^2 & k = 0 \\ \alpha\sigma_X^2 & k = \pm 1 \\ 0 & \text{otherwise.} \end{cases} \tag{10.30}$$

The power spectral density is then

$$\begin{aligned} S_Y(f) &= (1 + \alpha^2)\sigma_X^2 + \alpha\sigma_X^2\{e^{j2\pi f} + e^{-j2\pi f}\} \\ &= \sigma_X^2\{(1 + \alpha^2) + 2\alpha \cos 2\pi f\}. \end{aligned} \tag{10.31}$$

$S_Y(f)$ is shown in Fig. 10.3 for $\alpha = 1$.

Example 10.8   Signal Plus Noise

Let the observation Z_n be given by

$$Z_n = X_n + Y_n,$$

where X_n is the signal we wish to observe, Y_n is a white noise process with power σ_Y^2, and X_n and Y_n are independent random processes. Suppose further that $X_n = A$ for all n, where A is a random variable with zero mean and variance σ_A^2. Thus Z_n represents a sequence of noisy measurements of the random variable A. Find the power spectral density of Z_n.

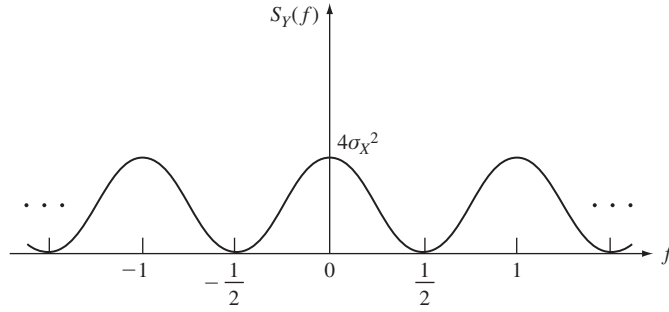The mean and autocorrelation of Z_n are

$$E[Z_n] = E[A] + E[Y_n] = 0$$

FIGURE 10.3
Power spectral density of moving average process discussed in Example 10.7.

and

$$E[Z_n Z_{n+k}] = E[(X_n + Y_n)(X_{n+k} + Y_{n+k})]$$

$$= E[X_n X_{n+k}] + E[X_n]E[Y_{n+k}]$$

$$+ E[X_{n+k}]E[Y_n] + E[Y_n Y_{n+k}]$$

$$= E[A^2] + R_Y(k).$$

Thus Z_n is also a WSS process.

The power spectral density of Z_n is then

$$S_Z(f) = E[A^2]\delta(f) + S_Y(f),$$

where we have used the fact that the Fourier transform of a constant is a delta function.

10.1.3   Power Spectral Density as a Time Average

In the above discussion, we simply stated that the power spectral density is given as the Fourier transform of the autocorrelation without supplying a proof. We now show how the power spectral density arises naturally when we take Fourier transforms of realizations of random processes.

Let $X_0, \ldots, X_{k-1}$ be k observations from the discrete-time, WSS process X_n. Let $\widetilde{x}_k(f)$ denote the discrete Fourier transform of this sequence:

$$\widetilde{x}_k(f) = \sum_{m=0}^{k-1} X_m e^{-j2\pi f m}. \tag{10.32}$$

Note that $\widetilde{x}_k(f)$ is a complex-valued random variable. The magnitude squared of $\widetilde{x}_k(f)$ is a measure of the "energy" at the frequency f. If we divide this energy by the total "time" k, we obtain an estimate for the "power" at the frequency f:

$$\widetilde{p}_k(f) = \frac{1}{k}|\widetilde{x}_k(f)|^2. \tag{10.33}$$

$\widetilde{p}_k(f)$ is called the **periodogram estimate** for the power spectral density.

Consider the expected value of the periodogram estimate:

$$E[\tilde{p}_k(f)] = \frac{1}{k}E[\tilde{x}_k(f)\tilde{x}_k^*(f)]$$

$$= \frac{1}{k}E\left[\sum_{m=0}^{k-1} X_m e^{-j2\pi fm}\sum_{i=0}^{k-1} X_i e^{j2\pi fi}\right]$$

$$= \frac{1}{k}\sum_{m=0}^{k-1}\sum_{i=0}^{k-1} E[X_m X_i]e^{-j2\pi f(m-i)}$$

$$= \frac{1}{k}\sum_{m=0}^{k-1}\sum_{i=0}^{k-1} R_X(m-i)e^{-j2\pi f(m-i)}. \qquad (10.34)$$

Figure 10.4 shows the range of the double summation in Eq. (10.34). Note that all the terms along the diagonal $m' = m - i$ are equal, that m' ranges from $-(k-1)$ to $k-1$, and that .here are $k - |m'|$ terms along the diagonal $m' = m - i$. Thus Eq. (10.34) becomes

$$E[\tilde{p}_k(f)] = \frac{1}{k}\sum_{m'=-(k-1)}^{k-1}\{k - |m'|\}R_X(m')e^{-j2\pi fm'}$$

$$= \sum_{m'=-(k-1)}^{k-1}\left\{1 - \frac{|m'|}{k}\right\}R_X(m')e^{-j2\pi fm'}. \qquad (10.35)$$

Comparison of Eq. (10.35) with Eq. (10.24) shows that the mean of the periodogram estimate is not equal to $S_X(f)$ for two reasons. First, Eq. (10.34) does not have the term in brackets in Eq. (10.25). Second, the limits of the summation in Eq. (10.35) are not $\pm\infty$. We say that $\tilde{p}_k(f)$ is a "biased" estimator for $S_X(f)$. However, as $k \to \infty$, we see
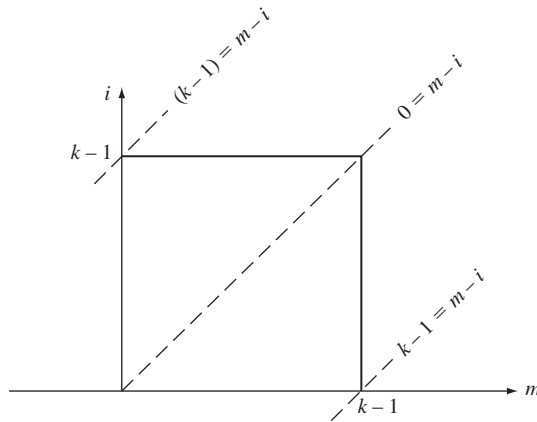
FIGURE 10.4
Range of summation in Eq. (10.34).

that the term in brackets approaches one, and that the limits of the summation approach $\pm\infty$. Thus

$$E[\widetilde{p}_k(f)] \rightarrow S_X(f) \qquad \text{as } k \rightarrow \infty, \tag{10.36}$$

that is, the mean of the periodogram estimate does indeed approach $S_X(f)$. Note that Eq. (10.36) shows that $S_X(f)$ is nonnegative for all f, since $\widetilde{p}_k(f)$ is nonnegative for all f.

In order to be useful, the variance of the periodogram estimate should also approach zero. The answer to this question involves looking more closely at the problem of power spectral density estimation. We defer this topic to Section 10.6.

All of the above results hold for a continuous-time WSS random process $X(t)$ after appropriate changes are made from summations to integrals. The periodogram estimate for $S_X(f)$, for an observation in the interval $0 < t < T$, was defined in Eq. 10.2. The same derivation that led to Eq. (10.35) can be used to show that the mean of the periodogram estimate is given by

$$E[\widetilde{p}_T(f)] = \int_{-T}^{T} \left\{ 1 - \frac{|\tau|}{T} \right\} R_X(\tau) e^{-j2\pi f\tau}\, d\tau. \tag{10.37a}$$

It then follows that

$$E[\widetilde{p}_T(f)] \rightarrow S_X(f) \qquad \text{as } T \rightarrow \infty. \tag{10.37b}$$

10.2   RESPONSE OF LINEAR SYSTEMS TO RANDOM SIGNALS

Many applications involve the processing of random signals (i.e., random processes) in order to achieve certain ends. For example, in prediction, we are interested in predicting future values of a signal in terms of past values. In filtering and smoothing, we are interested in recovering signals that have been corrupted by noise. In modulation, we are interested in converting low-frequency information signals into high-frequency transmission signals that propagate more readily through various transmission media.

Signal processing involves converting a signal from one form into another. Thus a signal processing method is simply a transformation or mapping from one time function into another function. If the input to the transformation is a random process, then the output will also be a random process. In the next two sections, we are interested in determining the statistical properties of the output process when the input is a wide-sense stationary random process.

10.2.1   Continuous-Time Systems

Consider a **system** in which an input signal $x(t)$ is mapped into the output signal $y(t)$ by the transformation

$$y(t) = T[x(t)].$$

The system is **linear** if superposition holds, that is,

$$T[\alpha x_1(t) + \beta x_2(t)] = \alpha T[x_1(t)] + \beta T[x_2(t)],$$

where $x_1(t)$ and $x_2(t)$ are arbitrary input signals, and α and β are arbitrary constants.[4] Let $y(t)$ be the response to input $x(t)$, then the system is said to be **time-invariant** if the response to $x(t - \tau)$ is $y(t - \tau)$. The **impulse response** $h(t)$ of a linear, time-invariant system is defined by

$$h(t) = T[\delta(t)]$$

where $\delta(t)$ is a unit delta function input applied at $t = 0$. The response of the system to an arbitrary input $x(t)$ is then

$$y(t) = h(t) * x(t) = \int_{-\infty}^{\infty} h(s)x(t - s)\, ds = \int_{-\infty}^{\infty} h(t - s)x(s)\, ds. \quad (10.38)$$

Therefore a linear, time-invariant system is completely specified by its impulse response. The impulse response $h(t)$ can also be specified by giving its Fourier transform, the **transfer function** of the system:

$$H(f) = \mathcal{F}\{h(t)\} = \int_{-\infty}^{\infty} h(t)e^{-j2\pi ft}\, dt. \quad (10.39)$$

A system is said to be **causal** if the response at time t depends only on past values of the input, that is, if $h(t) = 0$ for $t < 0$.

If the input to a linear, time-invariant system is a random process $X(t)$ as shown in Fig. 10.5, then the output of the system is the random process given by

$$Y(t) = \int_{-\infty}^{\infty} h(s)X(t - s)\, ds = \int_{-\infty}^{\infty} h(t - s)X(s)\, ds. \quad (10.40)$$

We assume that the integrals exist in the mean square sense as discussed in Section 9.7. We now show that if $X(t)$ is a wide-sense stationary process, then $Y(t)$ is also wide-sense stationary.[5]

The mean of $Y(t)$ is given by

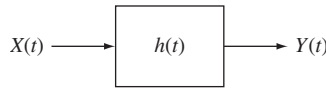$$E[Y(t)] = E\left[\int_{-\infty}^{\infty} h(s)X(t - s)\, ds\right] = \int_{-\infty}^{\infty} h(s)E[X(t - s)]\, ds.$$

$$X(t) \longrightarrow \boxed{\quad h(t) \quad} \longrightarrow Y(t)$$

FIGURE 10.5
A linear system with a random input
signal.

---

[4]For examples of nonlinear systems see Problems 9.11 and 9.56.
[5]Equation (10.40) supposes that the input was applied at an infinite time in the past. If the input is applied at $t = 0$, then $Y(t)$ is not wide-sense stationary. However, it becomes wide-sense stationary as the response reaches "steady state" (see Example 9.46 and Problem 10.29).

Now $m_X = E[X(t - \tau)]$ since $X(t)$ is wide-sense stationary, so

$$E[Y(t)] = m_X \int_{-\infty}^{\infty} h(\tau) \, d\tau = m_X H(0), \tag{10.41}$$

where $H(f)$ is the transfer function of the system. Thus the mean of the output $Y(t)$ is the constant $m_Y = H(0)m_X$.

The autocorrelation of $Y(t)$ is given by

$$E[Y(t)Y(t + \tau)] = E\left[\int_{-\infty}^{\infty} h(s)X(t - s) \, ds \int_{-\infty}^{\infty} h(r)X(t + \tau - r) \, dr\right]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h(r)E[X(t - s)X(t + \tau - r)] \, ds \, dr$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h(r)R_X(\tau + s - r) \, ds \, dr, \tag{10.42}$$

where we have used the fact that $X(t)$ is wide-sense stationary. The expression on the right-hand side of Eq. (10.42) depends only on τ. Thus the autocorrelation of $Y(t)$ depends only on τ, and since the $E[Y(t)]$ is a constant, we conclude that $Y(t)$ is a wide-sense stationary process.

We are now ready to compute the power spectral density of the output of a linear, time-invariant system. Taking the transform of $R_Y(\tau)$ as given in Eq. (10.42), we obtain

$$S_Y(f) = \int_{-\infty}^{\infty} R_Y(\tau)e^{-j2\pi f\tau} \, d\tau$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h(r)R_X(\tau + s - r)e^{-j2\pi f\tau} \, ds \, dr \, d\tau.$$

Change variables, letting $u = \tau + s - r$:

$$S_Y(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(s)h(r)R_X(u)e^{-j2\pi f(u-s+r)} \, ds \, dr \, du$$

$$= \int_{-\infty}^{\infty} h(s)e^{j2\pi fs} \, ds \int_{-\infty}^{\infty} h(r)e^{-j2\pi fr} \, dr \int_{-\infty}^{\infty} R_X(u)e^{-j2\pi fu} \, du$$

$$= H^*(f)H(f)S_X(f)$$

$$= |H(f)|^2 S_X(f), \tag{10.43}$$

where we have used the definition of the transfer function. *Equation (10.43) relates the input and output power spectral densities to the system transfer function.* Note that $R_Y(\tau)$ can also be found by computing Eq. (10.43) and then taking the inverse Fourier transform.

Equations (10.41) through (10.43) only enable us to determine the mean and autocorrelation function of the output process $Y(t)$. In general this is not enough to determine probabilities of events involving $Y(t)$. However, if the input process is a

Gaussian WSS random process, then as discussed in Section 9.7 the output process will also be a Gaussian WSS random process. Thus the mean and autocorrelation function provided by Eqs. (10.41) through (10.43) are enough to determine all joint pdf's involving the Gaussian random process $Y(t)$.

The cross-correlation between the input and output processes is also of interest:

$$R_{Y,X}(\tau) = E[Y(t + \tau)X(t)]$$

$$= E\left[X(t) \int_{-\infty}^{\infty} X(t + \tau - r)h(r)\, dr \right]$$

$$= \int_{-\infty}^{\infty} E[X(t)X(t + \tau - r)]h(r)\, dr$$

$$= \int_{-\infty}^{\infty} R_X(\tau - r)h(r)\, dr$$

$$= R_X(\tau) * h(\tau). \tag{10.44}$$

By taking the Fourier transform, we obtain the cross-power spectral density:

$$S_{Y,X}(f) = H(f)S_X(f). \tag{10.45a}$$

Since $R_{X,Y}(\tau) = R_{Y,X}(-\tau)$, we have that

$$S_{X,Y}(f) = S_{Y,X}^*(f) = H^*(f)S_X(f). \tag{10.45b}$$

Example 10.9 Filtered White Noise

Find the power spectral density of the output of a linear, time-invariant system whose input is a white noise process.

Let $X(t)$ be the input process with power spectral density

$$S_X(f) = \frac{N_0}{2} \qquad \text{for all } f.$$

The power spectral density of the output $Y(t)$ is then

$$S_Y(f) = |H(f)|^2 \frac{N_0}{2}. \tag{10.46}$$

Thus the transfer function completely determines the shape of the power spectral density of the output process.

Example 10.9 provides us with a method for generating WSS processes with arbitrary power spectral density $S_Y(f)$. We simply need to filter white noise through a filter with transfer function $H(f) = \sqrt{S_Y(f)}$. In general this filter will be noncausal. We can usually, but not always, obtain a *causal* filter with transfer function $H(f)$ such that $S_Y(f) = H(f)H^*(f)$. For example, if $S_Y(f)$ is a rational function, that is, if it consists of the ratio of two polynomials, then it is easy to factor $S_X(f)$ into the above form, as

shown in the next example. Furthermore any power spectral density can be approximated by a rational function. Thus filtered white noise can be used to synthesize WSS random processes with arbitrary power spectral densities, and hence arbitrary autocorrelation functions.

Example 10.10   Ornstein-Uhlenbeck Process

Find the impulse response of a causal filter that can be used to generate a Gaussian random process with output power spectral density and autocorrelation function

$$S_Y(f) = \frac{\sigma^2}{\alpha^2 + 4\pi^2 f^2} \quad \text{and} \quad R_Y(\tau) = \frac{\sigma^2}{2\alpha} e^{-\alpha|\tau|}$$

This power spectral density factors as follows:

$$S_Y(f) = \frac{1}{(\alpha - j2\pi f)} \frac{1}{(\alpha + j2\pi f)} \sigma^2.$$

If we let the filter transfer function be $H(f) = 1/(\alpha + j2\pi f)$, then the impulse response is

$$h(t) = e^{-\alpha t} \quad \text{for } t \geq 0,$$

which is the response of a causal system. Thus if we filter white Gaussian noise with power spectral density σ^2 using the above filter, we obtain a process with the desired power spectral density.

In Example 9.46, we found the autocorrelation function of the transient response of this filter for a white Gaussian noise input (see Eq. (9.97a)). As was already indicated, when dealing with power spectral densities we assume that the processes are in steady state. Thus as $t \to \infty$ Eq. (9.97a) approaches Eq. (9.97b).

Example 10.11   Ideal Filters

Let $Z(t) = X(t) + Y(t)$, where $X(t)$ and $Y(t)$ are independent random processes with power spectral densities shown in Fig. 10.6(a). Find the output if $Z(t)$ is input into an ideal lowpass filter with transfer function shown in Fig. 10.6(b). Find the output if $Z(t)$ is input into an ideal bandpass filter with transfer function shown in Fig. 10.6(c).

The power spectral density of the output $W(t)$ of the lowpass filter is

$$S_W(f) = |H_{LP}(f)|^2 S_X(f) + |H_{LP}(f)|^2 S_Y(f) = S_X(f),$$

since $H_{LP}(f) = 1$ for the frequencies where $S_X(f)$ is nonzero, and $H_{LP}(f) = 0$ where $S_Y(f)$ is nonzero. Thus $W(t)$ has the same power spectral density as $X(t)$. As indicated in Example 10.5, this does not imply that $W(t) = X(t)$.

To show that $W(t) = X(t)$, in the mean square sense, consider $D(t) = W(t) - X(t)$. It is easily shown that

$$R_D(\tau) = R_W(\tau) - R_{WX}(\tau) - R_{XW}(\tau) + R_X(\tau).$$

The corresponding power spectral density is

$$\begin{aligned} S_D(f) &= S_W(f) - S_{WX}(f) - S_{XW}(f) + S_X(f) \\ &= |H_{LP}(f)|^2 S_X(f) - H_{LP}(f) S_X(f) - H_{LP}^*(f) S_X(f) + S_X(f) \\ &= 0. \end{aligned}$$
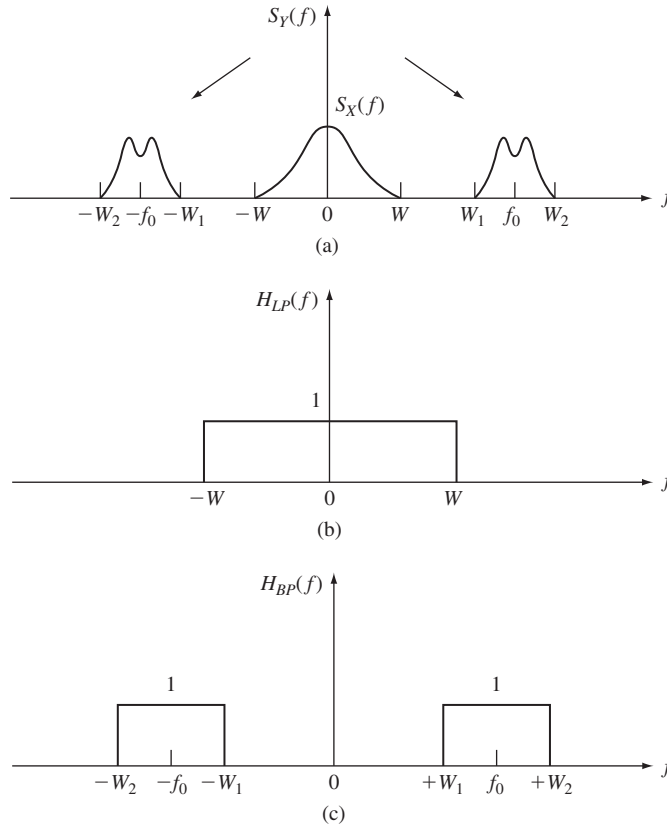
FIGURE 10.6
(a) Input signal to filters is $X(t) + Y(t)$, (b) lowpass filter, (c) bandpass filter.

Therefore $R_D(\tau) = 0$ for all τ, and $W(t) = X(t)$ in the mean square sense since

$$E[(W(t) - X(t))^2] = E[D^2(t)] = R_D(0) = 0.$$

Thus we have shown that the lowpass filter removes $Y(t)$ and passes $X(t)$. Similarly, the bandpass filter removes $X(t)$ and passes $Y(t)$.

Example 10.12

A random telegraph signal is passed through an RC lowpass filter which has transfer function

$$H(f) = \frac{\beta}{\beta + j2\pi f},$$

where $\beta = 1/RC$ is the time constant of the filter. Find the power spectral density and autocorrelation of the output.

In Example 10.1, the power spectral density of the random telegraph signal with transition rate α was found to be

$$S_X(f) = \frac{4\alpha}{4\alpha^2 + 4\pi^2 f^2}.$$

From Eq. (10.43) we have

$$S_Y(f) = \left(\frac{\beta^2}{\beta^2 + 4\pi^2 f^2} \right) \left(\frac{4\alpha}{4\alpha^2 + 4\pi^2 f^2} \right)$$

$$= \frac{4\alpha\beta^2}{\beta^2 - 4\alpha^2} \left\{ \frac{1}{4\alpha^2 + 4\pi^2 f^2} - \frac{1}{\beta^2 + 4\pi^2 f^2} \right\}.$$

$R_Y(\tau)$ is found by inverting the above expression:

$$R_Y(\tau) = \frac{1}{\beta^2 - 4\alpha^2} \{ \beta^2 e^{-2\alpha|\tau|} - 2\alpha\beta e^{-\beta|\tau|} \}.$$

10.2.2 Discrete-Time Systems

The results obtained above for continuous-time signals also hold for discrete-time signals after appropriate changes are made from integrals to summations.

Let the **unit-sample response** h_n be the response of a discrete-time, linear, time-invariant system to a unit-sample input δ_n:

$$\delta_n = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0. \end{cases} \tag{10.47}$$

The response of the system to an arbitrary input random process X_n is then given by

$$Y_n = h_n * X_n = \sum_{j=-\infty}^{\infty} h_j X_{n-j} = \sum_{j=-\infty}^{\infty} h_{n-j} X_j. \tag{10.48}$$

Thus discrete-time, linear, time-invariant systems are determined by the unit-sample response h_n. The **transfer function** of such a system is defined by

$$H(f) = \sum_{i=-\infty}^{\infty} h_i e^{-j2\pi fi}. \tag{10.49}$$

The derivation from the previous section can be used to show that if X_n is a wide-sense stationary process, then Y_n is also wide-sense stationary. The mean of Y_n is given by

$$m_Y = m_X \sum_{j=-\infty}^{\infty} h_j = m_X H(0). \tag{10.50}$$

The autocorrelation of Y_n is given by

$$R_Y(k) = \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} h_j h_i R_X(k + j - i). \tag{10.51}$$

By taking the Fourier transform of $R_Y(k)$ it is readily shown that the power spectral density of Y_n is

$$S_Y(f) = |H(f)|^2 S_X(f). \tag{10.52}$$

This is the same equation that was found for continuous-time systems.

Finally, we note that if the input process X_n is a Gaussian WSS random process, then the output process Y_n is also a Gaussian WSS random whose statistics are completely determined by the mean and autocorrelation function provided by Eqs. (10.50) through (10.52).

Example 10.13 Filtered White Noise

Let X_n be a white noise sequence with zero mean and average power σ_X^2. If X_n is the input to a linear, time-invariant system with transfer function $H(f)$, then the output process Y_n has power spectral density:

$$S_Y(f) = |H(f)|^2 \sigma_X^2. \tag{10.53}$$

Equation (10.53) provides us with a method for generating discrete-time random processes with arbitrary power spectral densities or autocorrelation functions. If the power spectral density can be written as a rational function of $z = e^{j2\pi f}$ in Eq. (10.24), then a causal filter can be found to generate a process with the power spectral density. Note that this is a generalization of the methods presented in Section 6.6 for generating vector random variables with arbitrary covariance matrix.

Example 10.14 First-Order Autoregressive Process

A first-order autoregressive (AR) process Y_n with zero mean is defined by

$$Y_n = \alpha Y_{n-1} + X_n, \tag{10.54}$$

where X_n is a zero-mean white noise input random process with average power σ_X^2. Note that Y_n can be viewed as the output of the system in Fig. 10.7(a) for an iid input X_n. Find the power spectral density and autocorrelation of Y_n.

The unit-sample response can be determined from Eq. (10.54):

$$h_n = \begin{cases} 0 & n < 0 \\ 1 & n = 0 \\ \alpha^n & n > 0. \end{cases}$$

Note that we require $|\alpha| < 1$ for the system to be stable.[6] Therefore the transfer function is

$$H(f) = \sum_{n=0}^{\infty} \alpha^n e^{-j2\pi fn} = \frac{1}{1 - \alpha e^{-j2\pi f}}.$$

---

[6]A system is said to be **stable** if $\sum_n |h_n| < \infty$. The response of a stable system to any bounded input is also bounded.
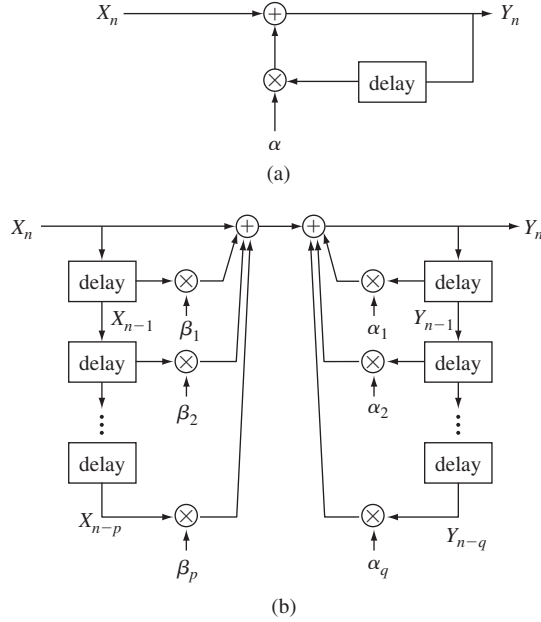
FIGURE 10.7
(a) Generation of AR process; (b) Generation of ARMA process.

Equation (10.52) then gives

$$S_Y(f) = \frac{\sigma_X^2}{(1 - \alpha e^{-j2\pi f})(1 - \alpha e^{j2\pi f})}$$

$$= \frac{\sigma_X^2}{1 + \alpha^2 - (\alpha e^{-j2\pi f} + \alpha e^{j2\pi f})}$$

$$= \frac{\sigma_X^2}{1 + \alpha^2 - 2\alpha \cos 2\pi f}.$$

Equation (10.51) gives

$$R_Y(k) = \sum_{j=0}^{\infty}\sum_{i=0}^{\infty} h_j h_i \sigma_X^2 \delta_{k+j-i} = \sigma_X^2 \sum_{j=0}^{\infty} \alpha^j \alpha^{j+k} = \frac{\sigma_X^2 \alpha^k}{1 - \alpha^2}.$$

Example 10.15   ARMA Random Process

An **autoregressive moving average (ARMA) process** is defined by

$$Y_n = -\sum_{i=1}^{q} \alpha_i Y_{n-i} + \sum_{i'=0}^{p} \beta_{i'} W_{n-i'}, \tag{10.55}$$

where W_n is a WSS, white noise input process. Y_n can be viewed as the output of the recursive system in Fig. 10.7(b) to the input X_n. It can be shown that the transfer function of the linear system

defined by the above equation is

$$H(f) = \frac{\sum_{i'=0}^{p} \beta_{i'} e^{-j2\pi f i'}}{1 + \sum_{i=1}^{q} \alpha_i e^{-j2\pi f i}}.$$

The power spectral density of the ARMA process is

$$S_Y(f) = |H(f)|^2 \sigma_W^2.$$

ARMA models are used extensively in random time series analysis and in signal processing. The general **autoregressive process** is the special case of the ARMA process with $\beta_1 = \beta_2 = \cdots = \beta_p = 0$. The general **moving average process** is the special case of the ARMA process with $\alpha_1 = \alpha_2 = \cdots = \alpha_q = 0$. Octave has a function filter(b, a, x) which takes a set of coefficients $b = (\beta_1, \beta_2, \ldots, \beta_{p+1})$ and $a = (\alpha_1, \alpha_2, \ldots, \alpha_q)$ as coefficient for a filter as in Eq. (10.55) and produces the output corresponding to the input sequence x. The choice of a and b can lead to a broad range of discrete-time filters.

For example, if we let $a = (1/N, 1/N, \ldots, 1/N)$ we obtain a moving average filter:

$$Y_n = (W_n + W_{n-1} + \cdots + W_{n-N+1})/N.$$

Figure 10.8 shows a zero-mean, unit-variance Gaussian iid sequence W_n and the outputs from an $N = 3$ and an $N = 10$ moving average filter. It can be seen that the $N = 3$ filter moderates the extreme variations but generally tracks the fluctuations in X_n. The $N = 10$ filter on the other hand severely limits the variations and only tracks slower longer-lasting trends.

Figures 10.9(a) and (b) show the result of passing an iid Gaussian sequence X_n through first-order autoregressive filters as in Eq. (10.54). The AR sequence with $\alpha = 0.1$ has low correlation between adjacent samples and so the sequence remains similar to the underlying iid random process. The AR sequence with $\alpha = 0.75$ has higher correlation between adjacent samples which tends to cause longer lasting trends as evident in Fig. 10.9(b).
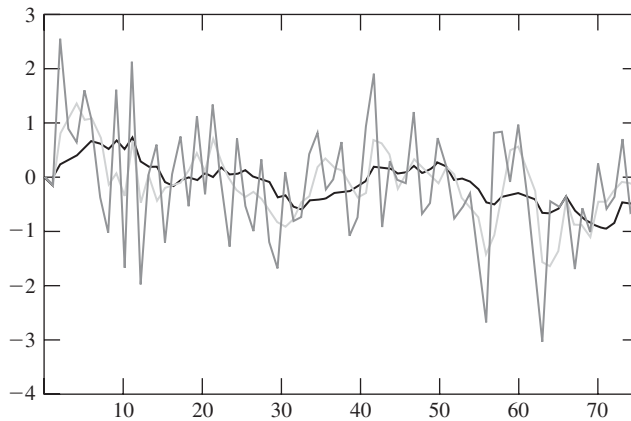

FIGURE 10.8
Moving average process showing iid Gaussian sequence and corresponding
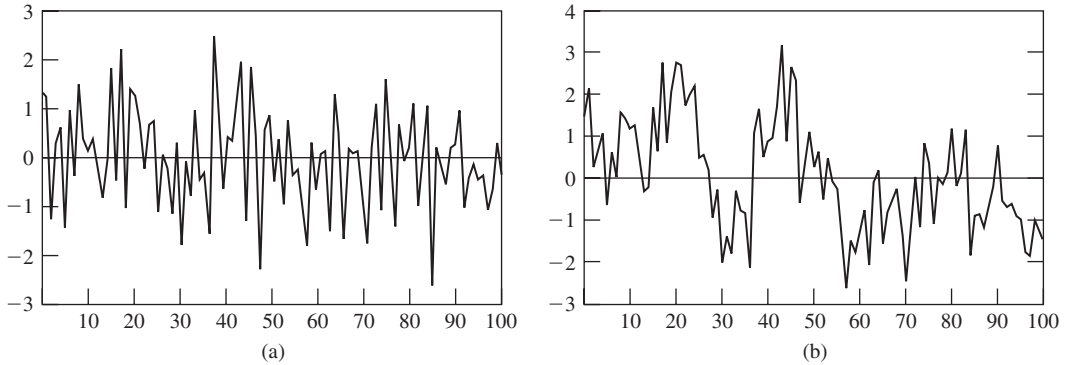$N = 3, N = 10$ moving average processes.

FIGURE 10.9
(a) First-order autoregressive process with $\alpha = 0.1$; (b) with $\alpha = 0.75$.

10.3   BANDLIMITED RANDOM PROCESSES

In this section we consider two important applications that involve random processes with power spectral densities that are nonzero over a finite range of frequencies. The first application involves the sampling theorem, which states that bandlimited random processes can be represented in terms of a sequence of their time samples. This theorem forms the basis for modern digital signal processing systems. The second application involves the modulation of sinusoidal signals by random information signals. Modulation is a key element of all modern communication systems.

10.3.1   Sampling of Bandlimited Random Processes

One of the major technology advances in the twentieth century was the development of digital signal processing technology. All modern multimedia systems depend in some way on the processing of digital signals. Many information signals, e.g., voice, music, imagery, occur naturally as analog signals that are continuous-valued and that vary continuously in time or space or both. The two key steps in making these signals amenable to digital signal processing are: (1). Convert the continuous-time signals into discrete-time signals by sampling the amplitudes; (2) Representing the samples using a fixed number of bits. In this section we introduce the sampling theorem for wide-sense stationary bandlimited random processes, which addresses the conversion of signals into discrete-time sequences.

Let $x(t)$ be a deterministic, finite-energy time signal that has Fourier transform $\widetilde{X}(f) = \mathcal{F}\{x(t)\}$ that is nonzero only in the frequency range $|f| \leq W$. Suppose we sample $x(t)$ every T seconds to obtain the sequence of sample values: $\{\dots, x(-2T), x(-T), x(0), x(T), \dots\}$. *The sampling theorem for deterministic signals states that $x(t)$ can be recovered exactly from the sequence of samples if $T \leq 1/2W$ or equivalently $1/T \geq 2W$, that is, the sampling rate is at least twice the bandwidth of the signal.* The minimum sampling rate $1/2W$ is called the **Nyquist sampling rate**. The sampling
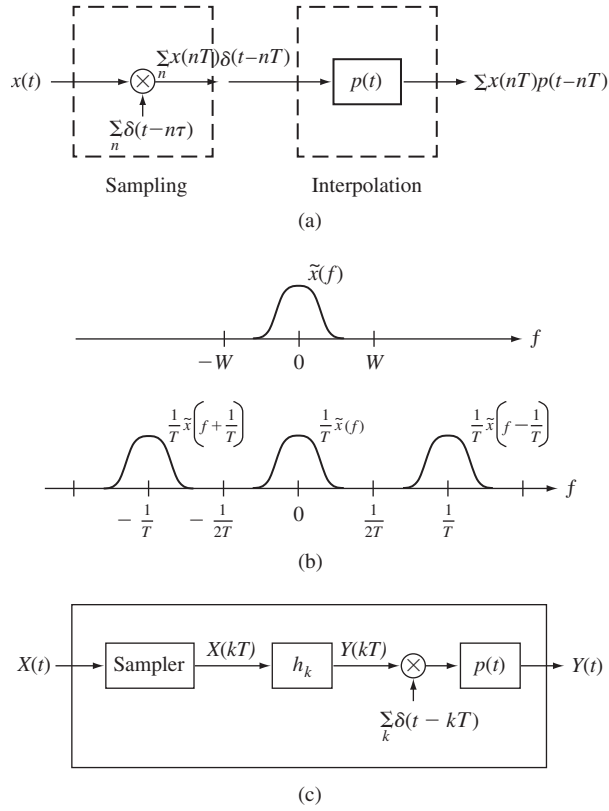
FIGURE 10.10
(a) Sampling and interpolation; (b) Fourier transform of sampled
deterministic signal; (c) Sampling, digital filtering, and interpolation.

theorem provides the following interpolation formula for recovering $x(t)$ from the samples:

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT)p(t - nT) \quad \text{where} \quad p(t) = \frac{\sin(\pi t/T)}{\pi t/T}. \qquad (10.56)$$

Eq. (10.56) provides us with the interesting interpretation depicted in Fig. 10.10(a). The process of sampling $x(t)$ can be viewed as the multiplication of $x(t)$ by a train of delta functions spaced T seconds apart. The sampled function is then represented by:

$$x_s(t) = \sum_{n=-\infty}^{\infty} x(nT)\delta(t - nT). \qquad (10.57)$$

Eq. (10.56) can be viewed as the response of a linear system with impulse response $p(t)$ to the signal $x_s(t)$. It is easy to show that the $p(t)$ in Eq. (10.56) corresponds to the ideal lowpass filter in Fig. 10.6:

$$P(f) = \mathcal{F}\{p(t)\} = \begin{cases} 1 & -W \leq f \leq W \\ 0 & |f| > W. \end{cases}$$

The proof of the sampling theorem involves the following steps. We show that

$$\mathcal{F}\left\{ \sum_{n=-\infty}^{\infty} x(nT)p(t - nT) \right\} = \frac{1}{T}P(f) \sum_{k=-\infty}^{\infty} \widetilde{X}(f - \frac{k}{T}), \qquad (10.58)$$

which consists of the sum of translated versions of $\widetilde{X}(f) = \mathcal{F}\{x(t)\}$, as shown in Fig. 10.10(b). We then observe that as long as $1/T \geq 2W$, then $P(f)$ in the above expressions selects the $k = 0$ term in the summation, which corresponds to $X(f)$. See Problem 10.45 for details.

Example 10.16   Sampling a WSS Random Process

Let $X(t)$ be a WSS process with autocorrelation function $R_X(\tau)$. Find the mean and covariance functions of the discrete-time sampled process $X_n = X(nT)$ for $n = 0, \pm 1, \pm 2, \ldots$.

Since $X(t)$ is WSS, the mean and covariance functions are:

$$m_X(n) = E[X(nT)] = m$$

$$E[X_{n_1}X_{n_2}] = E[X(n_1T)X(n_2T)] = R_X(n_1T - n_2T) = R_X((n_1 - n_2)T).$$

This shows X_n is a WSS discrete-time process.

Let $X(t)$ be a WSS process with autocorrelation function $R_X(\tau)$ and power spectral density $S_X(f)$. Suppose that $S_X(f)$ is bandlimited, that is,

$$S_X(f) = 0 \quad |f| > W.$$

We now show that the sampling theorem can be extended to $X(t)$. Let

$$\hat{X}(t) = \sum_{n=-\infty}^{\infty} X(nT)p(t - nT) \quad \text{where} \quad p(t) = \frac{\sin(\pi t/T)}{\pi t/T}, \qquad (10.59)$$

then $\hat{X}(t) = X(t)$ *in the mean square sense*. Recall that equality in the mean square sense does not imply equality for all sample functions, so this version of the sampling theorem is weaker than the version in Eq. (10.56) for finite energy signals.

To show Eq. (10.59) we first note that since $S_X(f) = \mathcal{F}\{R_X(\tau)\}$, we can apply the sampling theorem for deterministic signals to $R_X(\tau)$:

$$R_X(\tau) = \sum_{n=-\infty}^{\infty} R_X(nT)p(\tau - nT). \qquad (10.60)$$

Next we consider the mean square error associated with Eq. (10.59):

$$E[\{X(t) - \hat{X}(t)\}^2] = E[\{X(t) - \hat{X}(t)\}X(t)] - E[\{X(t) - \hat{X}(t)\}\hat{X}(t)]$$

$$= \left\{ E[X(t)X(t)] - E[\hat{X}(t)X(t)] \right\} -$$

$$\left\{ E[X(t)\hat{X}(t)] - E[\hat{X}(t)\hat{X}(t)] \right\}.$$

It is easy to show that Eq. (10.60) implies that each of the terms in braces is equal to zero. (See Problem 10.48.) We then conclude that $\hat{X}(t) = X(t)$ in the mean square sense.

Example 10.17   Digital Filtering of a Sampled WSS Random Process

Let $X(t)$ be a WSS process with power spectral density $S_X(f)$ that is nonzero only for $|f| \leq W$. Consider the sequence of operations shown in Fig. 10.10(c): (1) $X(t)$ is sampled at the Nyquist rate; (2) the samples $X(nT)$ are input into a digital filter in Fig. 10.7(b) with $\alpha_1 = \alpha_2 = \cdots = \alpha_q = 0$; and (3) the resulting output sequence Y_n is fed into the interpolation filter. Find the power spectral density of the output $Y(t)$.

The output of the digital filter is given by:

$$Y(kT) = \sum_{n=0}^{p} \beta_n X((k-n)T)$$

and the corresponding autocorrelation from Eq. (10.51) is:

$$R_Y(kT) = \sum_{n=0}^{p}\sum_{i=0}^{p} \beta_n \beta_i R_X((k+n-i)T).$$

The autocorrelation of $Y(t)$ is found from the interpolation formula (Eq. 10.60):

$$R_Y(\tau) = \sum_{k=-\infty}^{\infty} R_Y(kT)p(\tau - kT) = \sum_{k=-\infty}^{\infty}\sum_{n=0}^{p}\sum_{i=0}^{p} \beta_n\beta_i R_X((k+n-i)T)p(\tau - kT)$$

$$= \sum_{n=0}^{p}\sum_{i=0}^{p} \beta_n\beta_i \left\{ \sum_{k=-\infty}^{\infty} R_X((k+n-i)T)p(\tau - kT) \right\}$$

$$= \sum_{n=0}^{p}\sum_{i=0}^{p} \beta_n\beta_i R_X(\tau + (n-i)T).$$

The output power spectral density is then:

$$S_Y(f) = \mathcal{F}\{R_Y(\tau)\} = \sum_{n=0}^{p}\sum_{i=0}^{p} \beta_n\beta_i \mathcal{F}\{R_X(\tau + (n-i)T)\}$$

$$= \sum_{n=0}^{p}\sum_{i=0}^{p} \beta_n\beta_i S_X(f)e^{-j2\pi f(n-i)T}$$

$$= \left\{ \sum_{n=0}^{p} \beta_n e^{-j2\pi fnT} \right\}\left\{ \sum_{i=0}^{p} \beta_i e^{j2\pi fiT} \right\} S_X(f)$$

$$= |H(fT)|^2 S_X(f) \tag{10.61}$$

where $H(f)$ is the transfer function of the digital filter as per Eq. (10.49). The key finding here is the appearance of $H(f)$ *evaluated at* fT. We have obtained a very nice result that characterizes the overall system response in Fig. 10.8 to the continuous-time input $X(t)$. This result is true for more general digital filters, see [Oppenheim and Schafer].

The sampling theorem provides an important bridge between continuous-time and discrete-time signal processing. It gives us a means for implementing the real as well as the simulated processing of random signals. First, we must sample the random process above its Nyquist sampling rate. We can then perform whatever digital processing is necessary. We can finally recover the continuous-time signal by interpolation. The only difference between real signal processing and simulated signal processing is that the former usually has real-time requirements, whereas the latter allows us to perform our processing at whatever rate is possible using the available computing power.

10.3.2   Amplitude Modulation by Random Signals

Many of the transmission media used in communication systems can be modeled as linear systems and their behavior can be specified by a transfer function $H(f)$, which passes certain frequencies and rejects others. Quite often the information signal $A(t)$ (i.e., a speech or music signal) is not at the frequencies that propagate well. The purpose of a **modulator** is to map the information signal $A(t)$ into a transmission signal $X(t)$ that is in a frequency range that propagates well over the desired medium. At the receiver, we need to perform an inverse mapping to recover $A(t)$ from $X(t)$. In this section, we discuss two of the amplitude modulation methods.

Let $A(t)$ be a WSS random process that represents an information signal. In general $A(t)$ will be "lowpass" in character, that is, its power spectral density will be concentrated at low frequencies, as shown in Fig. 10.11(a). An **amplitude modulation** (AM) system produces a transmission signal by multiplying $A(t)$ by a "carrier" signal $\cos(2\pi f_c t + \Theta)$:

$$X(t) = A(t)\cos(2\pi f_c t + \Theta), \tag{10.62}$$

where we assume Θ is a random variable that is uniformly distributed in the interval $(0, 2\pi)$, and Θ and $A(t)$ are independent.

The autocorrelation of $X(t)$ is

$$E[X(t + \tau)X(t)]$$
$$= E[A(t + \tau)\cos(2\pi f_c(t + \tau) + \Theta)A(t)\cos(2\pi f_c t + \Theta)]$$
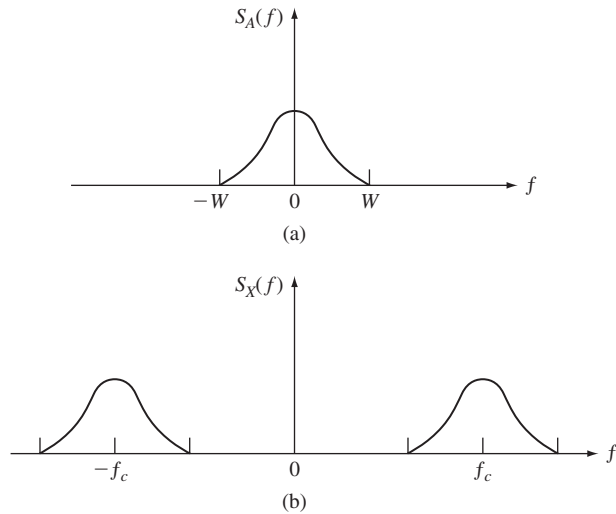$$= E[A(t + \tau)A(t)]E[\cos(2\pi f_c(t + \tau) + \Theta)\cos(2\pi f_c t + \Theta)]$$

(a)

(b)

FIGURE 10.11
(a) A lowpass information signal; (b) an amplitude-modulated signal.

$$= R_A(\tau)E\left[\frac{1}{2}\cos(2\pi f_c\tau) + \frac{1}{2}\cos(2\pi f_c(2t + \tau) + 2\Theta)\right]$$

$$= \frac{1}{2}R_A(\tau)\cos(2\pi f_c\tau), \tag{10.63}$$

where we used the fact that $E[\cos(2\pi f_c(2t + \tau) + 2\Theta)] = 0$ (see Example 9.10). Thus $X(t)$ is also a wide-sense stationary random process.

The power spectral density of $X(t)$ is

$$S_X(f) = \mathcal{F}\left\{\frac{1}{2}R_A(\tau)\cos(2\pi f_c\tau)\right\}$$

$$= \frac{1}{4}S_A(f + f_c) + \frac{1}{4}S_A(f - f_c), \tag{10.64}$$

where we used the table of Fourier transforms in Appendix B. Figure 10.11(b) shows $S_X(f)$. It can be seen that the power spectral density of the information signal has been shifted to the regions around $\pm f_c$. $X(t)$ is an example of a **bandpass signal**. Bandpass signals are characterized as having their power spectral density concentrated about some frequency much greater than zero.

The transmission signal is demodulated by multiplying it by the carrier signal and lowpass filtering, as shown in Fig. 10.12. Let

$$Y(t) = X(t)2\cos(2\pi f_c t + \Theta). \tag{10.65}$$

Proceeding as above, we find that

$$S_Y(f) = \frac{1}{2}S_X(f + f_c) + \frac{1}{2}S_X(f - f_c)$$

$$= \frac{1}{2}\{S_A(f + 2f_c) + S_A(f)\} + \frac{1}{2}\{S_A(f) + S_A(f - 2f_c)\}.$$

The ideal lowpass filter passes $S_A(f)$ and blocks $S_A(f \pm 2f_c)$, which is centered about $\pm f$, so the output of the lowpass filter has power spectral density

$$S_Y(f) = S_A(f).$$

In fact, from Example 10.11 we know the output is the original information signal, $A(t)$.
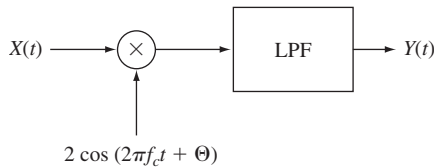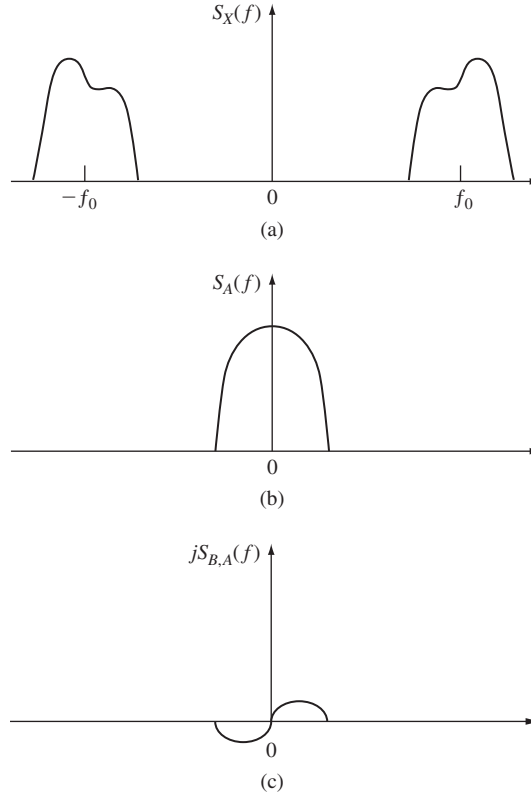

$X(t) \longrightarrow \otimes \longrightarrow \boxed{\text{LPF}} \longrightarrow Y(t)$

$2\cos(2\pi f_c t + \Theta)$

FIGURE 10.12
AM demodulator.

FIGURE 10.13
(a) A general bandpass signal. (b) a real-valued even function
of f. (c) an imaginary odd function of f.

The modulation method in Eq. (10.56) can only produce bandpass signals for which $S_X(f)$ is locally symmetric about f_c, $S_X(f_c + \delta f) = S_X(f_c - \delta f)$ for $|\delta f| < W$, as in Fig. 10.11(b). The method cannot yield real-valued transmission signals whose power spectral density lack this symmetry, such as shown in Fig. 10.13(a). The following **quadrature amplitude modulation** (QAM) method can be used to produce such signals:

$$X(t) = A(t) \cos(2\pi f_c t + \Theta) + B(t) \sin(2\pi f_c t + \Theta), \qquad (10.66)$$

where $A(t)$ and $B(t)$ are real-valued, jointly wide-sense stationary random processes, and we require that

$$R_A(\tau) = R_B(\tau) \qquad (10.67a)$$

$$R_{B,A}(\tau) = -R_{A,B}(\tau). \qquad (10.67b)$$

Note that Eq. (10.67a) implies that $S_A(f) = S_B(f)$, a real-valued, even function of f, as shown in Fig. 10.13(b). Note also that Eq. (10.67b) implies that $S_{B,A}(f)$ is a purely imaginary, odd function of f, as also shown in Fig. 10.13(c) (see Problem 10.57).

Proceeding as before, we can show that $X(t)$ is a wide-sense stationary random process with autocorrelation function

$$R_X(\tau) = R_A(\tau) \cos(2\pi f_c \tau) + R_{B,A}(\tau) \sin(2\pi f_c \tau) \tag{10.68}$$

and power spectral density

$$S_X(f) = \frac{1}{2}\{S_A(f - f_c) + S_A(f + f_c)\} + \frac{1}{2j}\{S_{BA}(f - f_c) - S_{BA}(f + f_c)\}. \tag{10.69}$$

The resulting power spectral density is as shown in Fig. 10.13(a). Thus QAM can be used to generate real-valued bandpass signals with arbitrary power spectral density.

Bandpass random signals, such as those in Fig. 10.13(a), arise in communication systems when wide-sense stationary white noise is filtered by bandpass filters. Let $N(t)$ be such a process with power spectral density $S_N(f)$. It can be shown that $N(t)$ can be represented by

$$N(t) = N_c(t) \cos(2\pi f_c t + \Theta) - N_s(t) \sin(2\pi f_c t + \Theta), \tag{10.70}$$

where $N_c(t)$ and $N_s(t)$ are jointly wide-sense stationary processes with

$$S_{N_c}(f) = S_{N_s}(f) = \{S_N(f - f_c) + S_N(f + f_c)\}_L \tag{10.71}$$

and

$$S_{N_c,N_s}(f) = j\{S_N(f - f_c) - S_N(f + f_c)\}_L, \tag{10.72}$$

where the subscript L denotes the lowpass portion of the expression in brackets. In words, every real-valued bandpass process can be treated as if it had been generated by a QAM modulator.

Example 10.18 Demodulation of Noisy Signal

The received signal in an AM system is

$$Y(t) = A(t) \cos(2\pi f_c t + \Theta) + N(t),$$

where $N(t)$ is a bandlimited white noise process with spectral density

$$S_N(f) = \begin{cases} \dfrac{N_0}{2} & |f \pm f_c| < W \\ 0 & \text{elsewhere.} \end{cases}$$

Find the signal-to-noise ratio of the recovered signal.

Equation (10.70) allows us to represent the received signal by

$$Y(t) = \{A(t) + N_c(t)\} \cos(2\pi f_c t + \Theta) - N_s(t) \sin(2\pi f_c t + \Theta).$$

The demodulator in Fig. 10.12 is used to recover $A(t)$. After multiplication by $2\cos(2\pi f_c t + \Theta)$, we have

$$2Y(t) \cos(2\pi f_c t + \Theta) = \{A(t) + N_c(t)\}2\cos^2(2\pi f_c t + \Theta)$$
$$- N_s(t)2\cos(2\pi f_c t + \Theta)\sin(2\pi f_c t + \Theta)$$
$$= \{A(t) + N_c(t)\}(1 + \cos(4\pi f_c t + 2\Theta))$$
$$- N_s(t)\sin(4\pi f_c t + 2\Theta).$$

After lowpass filtering, the recovered signal is

$$A(t) + N_c(t).$$

The power in the signal and noise components, respectively, are

$$\sigma_A^2 = \int_{-W}^{W} S_A(f) \, df$$

$$\sigma_{N_c}^2 = \int_{-W}^{W} S_{N_c}(f) \, df = \int_{-W}^{W} \left(\frac{N_0}{2} + \frac{N_0}{2} \right) df = 2WN_0.$$

The output signal-to-noise ratio is then

$$\text{SNR} = \frac{\sigma_A^2}{2WN_0}.$$

10.4   OPTIMUM LINEAR SYSTEMS

Many problems can be posed in the following way. We observe a discrete-time, zero-mean process X_α over a certain time interval $I = \{t - a, \dots, t + b\}$, and we are required to use the $a + b + 1$ resulting observations $\{X_{t-a}, \dots, X_t, \dots, X_{t+b}\}$ to obtain an estimate Y_t for some other (presumably related) zero-mean process Z_t. The estimate Y_t is required to be linear, as shown in Fig. 10.14:

$$Y_t = \sum_{\beta=t-a}^{t+b} h_{t-\beta} X_\beta = \sum_{\beta=-b}^{a} h_\beta X_{t-\beta}. \tag{10.73}$$

The figure of merit for the estimator is the mean square error

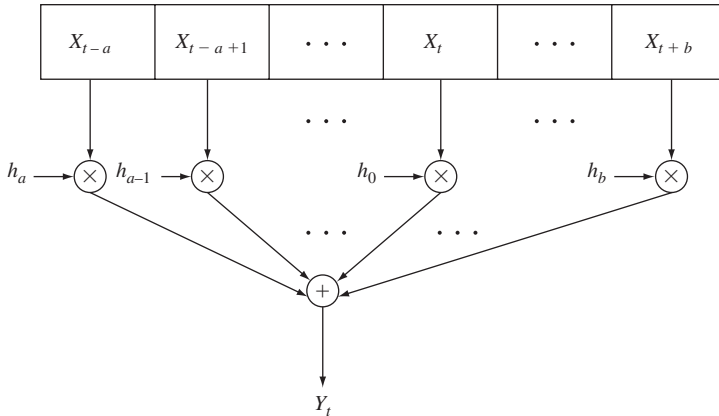$$E[e_t^2] = E[(Z_t - Y_t)^2], \tag{10.74}$$

FIGURE 10.14
A linear system for producing an estimate Y_t.

and we seek to find the **optimum filter**, which is characterized by the impulse response h_β that minimizes the mean square error.

Examples 10.19 and 10.20 show that different choices of Z_t and X_α and of observation interval correspond to different estimation problems.

Example 10.19   Filtering and Smoothing Problems

Let the observations be the sum of a "desired signal" Z_α plus unwanted "noise" N_α:

$$X_\alpha = Z_\alpha + N_\alpha \qquad \alpha \in I.$$

We are interested in estimating the desired signal at time t. The relation between t and the observation interval I gives rise to a variety of estimation problems.

If $I = (-\infty, t)$, that is, $a = \infty$ and $b = 0$, then we have a **filtering** problem where we estimate Z_t in terms of noisy observations of the past and present. If $I = (t - a, t)$, then we have a filtering problem in which we estimate Z_t in terms of the $a + 1$ most recent noisy observations.

If $I = (-\infty, \infty)$, that is, $a = b = \infty$, then we have a **smoothing** problem where we are attempting to recover the signal from its entire noisy version. There are applications where this makes sense, for example, if the entire realization X_α has been recorded and the estimate Z_t is obtained by "playing back" X_α.

Example 10.20   Prediction

Suppose we want to predict Z_t in terms of its recent past: $\{Z_{t-a}, \dots, Z_{t-1}\}$. The general estimation problem becomes this **prediction** problem if we let the observation X_α be the past a values of the signal Z_α, that is,

$$X_\alpha = Z_\alpha \qquad t - a \le \alpha \le t - 1.$$

The estimate Y_t is then a linear prediction of Z_t in terms of its most recent values.

10.4.1   The Orthogonality Condition

It is easy to show that the optimum filter must satisfy the **orthogonality condition** (see Eq. 6.56), which states that the error e_t must be orthogonal to all the observations X_α, that is,

$$0 = E[e_t X_\alpha] \qquad \text{for all } \alpha \in I$$

$$= E[(Z_t - Y_t) X_\alpha] = 0, \tag{10.75}$$

or equivalently,

$$E[Z_t X_\alpha] = E[Y_t X_\alpha] \qquad \text{for all } \alpha \in I. \tag{10.76}$$

If we substitute Eq. (10.73) into Eq. (10.76) we find

$$E[Z_t X_\alpha] = E\left[\sum_{\beta=-b}^{a} h_\beta X_{t-\beta} X_\alpha \right] \qquad \text{for all } \alpha \in I$$

$$= \sum_{\beta=-b}^{a} h_\beta E[X_{t-\beta} X_\alpha]$$

$$= \sum_{\beta=-b}^{a} h_\beta R_X(t - \alpha - \beta) \qquad \text{for all } \alpha \in I. \tag{10.77}$$

Equation (10.77) shows that $E[Z_t X_\alpha]$ depends only on $t - \alpha$, and thus X_α and Z_t are jointly wide-sense stationary processes. Therefore, we can rewrite Eq. (10.77) as follows:

$$R_{Z,X}(t - \alpha) = \sum_{\beta=-b}^{a} h_\beta R_X(t - \beta - \alpha) \qquad t - a \le \alpha \le t + b.$$

Finally, letting $m = t - \alpha$, we obtain the following key equation:

$$R_{Z,X}(m) = \sum_{\beta=-b}^{a} h_\beta R_X(m - \beta) \qquad -b \le m \le a. \qquad (10.78)$$

The optimum linear filter must satisfy the set of $a + b + 1$ linear equations given by Eq. (10.78). Note that Eq. (10.78) is identical to Eq. (6.60) for estimating a random variable by a linear combination of several random variables. The wide-sense stationarity of the processes reduces this estimation problem to the one considered in Section 6.5.

In the above derivation we deliberately used the notation Z_t instead of Z_n to suggest that the same development holds for *continuous-time estimation*. In particular, suppose we seek a linear estimate $Y(t)$ for the continuous-time random process $Z(t)$ in terms of observations of the continuous-time random process $X(\alpha)$ in the time interval $t - a \le \alpha \le t + b$:

$$Y(t) = \int_{t-a}^{t+b} h(t - \beta)X(\beta)\, d\beta = \int_{-b}^{a} h(\beta)X(t - \beta)\, d\beta.$$

It can then be shown that the filter $h(\beta)$ that minimizes the mean square error is specified by

$$R_{Z,X}(\tau) = \int_{-b}^{a} h(\beta)R_X(\tau - \beta)\, d\beta \qquad -b \le \tau \le a. \qquad (10.79)$$

Thus in the time-continuous case we obtain an integral equation instead of a set of linear equations. The analytic solution of this integral equation can be quite difficult, but the equation can be solved numerically by approximating the integral by a summation.[7]

We now determine the mean square error of the optimum filter. First we note that for the optimum filter, the error e_t and the estimate Y_t are orthogonal since

$$E[e_t Y_t] = E\left[e_t \sum h_{t-\beta} X_\beta\right] = \sum h_{t-\beta} E[e_t X_\beta] = 0,$$

where the terms inside the last summation are 0 because of Eq. (10.75). Since $e_t = Z_t - Y_t$, the mean square error is then

$$E[e_t^2] = E[e_t(Z_t - Y_t)]$$
$$= E[e_t Z_t],$$

---

[7]Equation (10.79) can also be solved by using the Karhunen-Loeve expansion.

since e_t and Y_t are orthogonal. Substituting for e_t yields

$$
\begin{aligned}
E[e_t^2] &= E[(Z_t - Y_t)Z_t] = E[Z_t Z_t] - E[Y_t Z_t] \\
&= R_Z(0) - E[Z_t Y_t] \\
&= R_Z(0) - E\left[Z_t \sum_{\beta=-b}^{a} h_\beta X_{t-\beta} \right] \\
&= R_Z(0) - \sum_{\beta=-b}^{a} h_\beta R_{Z,X}(\beta).
\end{aligned}
\tag{10.80}
$$

Similarly, it can be shown that the mean square error of the optimum filter in the continuous-time case is

$$
E[e^2(t)] = R_Z(0) = \int_{-b}^{a} h(\beta) R_{Z,X}(\beta) \, d\beta.
\tag{10.81}
$$

The following theorems summarize the above results.

Theorem

Let X_t and Z_t be discrete-time, zero-mean, jointly wide-sense stationary processes, and let Y_t be an estimate for Z_t of the form

$$
Y_t = \sum_{\beta=t-a}^{t+b} h_{t-\beta} X_\beta = \sum_{\beta=-b}^{a} h_\beta X_{t-\beta}.
$$

The filter that minimizes $E[(Z_t - Y_t)^2]$ satisfies the equation

$$
R_{Z,X}(m) = \sum_{\beta=-b}^{a} h_\beta R_X(m - \beta) \qquad -b \leq m \leq a
$$

and has mean square error given by

$$
E[(Z_t - Y_t)^2] = R_Z(0) - \sum_{\beta=-b}^{a} h_\beta R_{Z,X}(\beta).
$$

Theorem

Let $X(t)$ and $Z(t)$ be continuous-time, zero-mean, jointly wide-sense stationary processes, and let $Y(t)$ be an estimate for $Z(t)$ of the form

$$
Y(t) = \int_{t-a}^{t+b} h(t - \beta) X(\beta) \, d\beta = \int_{-b}^{a} h(\beta) X(t - \beta) \, d\beta.
$$

The filter $h(\beta)$ that minimizes $E[(Z(t) - Y(t))^2]$ satisfies the equation

$$
R_{Z,X}(\tau) = \int_{-b}^{a} h(\beta) R_X(\tau - \beta) \, d\beta \qquad -b \leq \tau \leq a
$$

and has mean square error given by

$$E[(Z(t) - Y(t))^2] = R_Z(0) - \int_{-b}^{a} h(\beta)R_{Z,X}(\beta)\, d\beta.$$

Example 10.21   Filtering of Signal Plus Noise

Suppose we are interested in estimating the signal Z_n from the $p + 1$ most recent noisy observations:

$$X_\alpha = Z_\alpha + N_\alpha \qquad \alpha \in I = \{n - p, \ldots, n - 1, n\}.$$

Find the set of linear equations for the optimum filter if Z_α and N_α are independent random processes.

For this choice of observation interval, Eq. (10.78) becomes

$$R_{Z,X}(m) = \sum_{\beta=0}^{p} h_\beta R_X(m - \beta) \qquad m \in \{0, 1, \ldots, p\}. \tag{10.82}$$

The cross-correlation terms in Eq. (10.82) are given by

$$R_{Z,X}(m) = E[Z_n X_{n-m}] = E[Z_n(Z_{n-m} + N_{n-m})] = R_Z(m).$$

The autocorrelation terms are given by

$$\begin{aligned} R_X(m - \beta) &= E[X_{n-\beta} X_{n-m}] = E[(Z_{n-\beta} + N_{n-\beta})(Z_{n-m} + N_{n-m})] \\ &= R_Z(m - \beta) + R_{Z,N}(m - \beta) \\ &\quad + R_{N,Z}(m - \beta) + R_N(m - \beta) \\ &= R_Z(m - \beta) + R_N(m - \beta), \end{aligned}$$

since Z_α and N_α are independent random processes. Thus Eq. (10.82) for the optimum filter becomes

$$R_Z(m) = \sum_{\beta=0}^{p} h_\beta \{R_Z(m - \beta) + R_N(m - \beta)\} \quad m \in \{0, 1, \ldots, p\}. \tag{10.83}$$

This set of $p + 1$ linear equations in $p + 1$ unknowns h_β is solved by matrix inversion.

Example 10.22   Filtering of AR Signal Plus Noise

Find the set of equations for the optimum filter in Example 10.21 if Z_α is a first-order autoregressive process with average power σ_Z^2 and parameter r, $|r| < 1$, and N_α is a white noise process with average power σ_N^2.

The autocorrelation for a first-order autoregressive process is given by

$$R_Z(m) = \sigma_Z^2 r^{|m|} \qquad m = 0, \pm 1, \pm 2, \ldots.$$

(See Problem 10.42.) The autocorrelation for the white noise process is

$$R_N(m) = \sigma_N^2\, \delta(m).$$

Substituting $R_Z(m)$ and $R_N(m)$ into Eq. (10.83) yields the following set of linear equations:

$$\sigma_Z^2 r^{|m|} = \sum_{\beta=0}^{p} h_\beta(\sigma_Z^2 r^{|m-\beta|} + \sigma_N^2 \delta(m - \beta)) \qquad m \in \{0, \ldots, p\}. \tag{10.84}$$

If we divide both sides of Eq. (10.84) by σ_Z^2 and let $\Gamma = \sigma_N^2/\sigma_Z^2$, we obtain the following matrix equation:

$$
\begin{bmatrix}
1 + \Gamma & r & r^2 & \cdots & r^p \\
r & 1 + \Gamma & r & \cdots & r^{p-1} \\
r^2 & r & 1 + \Gamma & \cdots & r^{p-2} \\
\cdot & \cdot & \cdot & \cdots & \cdot \\
r^p & r^{p-1} & r^{p-2} & \cdots & 1 + \Gamma
\end{bmatrix}
\begin{bmatrix}
h_0 \\ h_1 \\ \cdot \\ \cdot \\ h_p
\end{bmatrix}
=
\begin{bmatrix}
1 \\ r \\ \cdot \\ \cdot \\ r^p
\end{bmatrix}.
\tag{10.85}
$$

Note that when the noise power is zero, i.e., $\Gamma = 0$, then the solution is $h_0 = 1, h_j = 0$, $j = 1, \ldots, p$, that is, no filtering is required to obtain Z_n.

Equation (10.85) can be readily solved using Octave. The following function will compute the optimum linear coefficients and the mean square error of the optimum predictor:

```
function [mse]= Lin_Est_AR (order,rho,varsig,varnoise)
n=[0:1:order-1]
r=varsig*rho.^n;
R=varnoise*eye(order)+toeplitz(r);
H=inv(R)*transpose(r)
mse=varsig-transpose(H)*transpose(r);
endfunction
```

Table 10.1 gives the values of the optimal predictor coefficients and the mean square error as the order of the estimator is increased for the first-order autoregressive process with $\sigma_Z^2 = 4$, $r = 0.9$, and noise variance $\sigma_N^2 = 4$. It can be seen that the predictor places heavier weight on more recent samples, which is consistent with the higher correlation of such samples with the current sample. For smaller values of r, the correlation for distant samples drops off more quickly and the coefficients place even lower weighting on them. The mean square error can also be seen to decrease with increasing order $p + 1$ of the estimator. Increasing the first few orders provides significant improvements, but a point of diminishing returns is reached around $p + 1 = 3$.

10.4.2 Prediction

The linear prediction problem arises in many signal processing applications. In Example 6.31 in Chapter 6, we already discussed the linear prediction of speech signals. In general, we wish to predict Z_n in terms of $Z_{n-1}, Z_{n-2}, \ldots, Z_{n-p}$:

$$
Y_n = \sum_{\beta=1}^{p} h_\beta Z_{n-\beta}.
$$

TABLE 10.1   Effect of predictor order on MSE performance.

$p + 1$	MSE	Coefficients				
1	2.0000	0.5				
2	1.4922	0.37304	0.28213			
3	1.3193	0.32983	0.22500	0.17017		
4	1.2549	0.31374	0.20372	0.13897	0.10510	
5	1.2302	0.30754	0.19552	0.12696	0.08661	0.065501

For this problem, $X_\alpha = Z_\alpha$, so Eq. (10.79) becomes

$$R_Z(m) = \sum_{\beta=1}^{p} h_\beta R_Z(m - \beta) \qquad m \in \{1, \ldots, p\}. \tag{10.86a}$$

In matrix form this equation becomes

$$\begin{bmatrix} R_Z(1) \\ R_Z(2) \\ . \\ . \\ R_Z(p) \end{bmatrix} = \begin{bmatrix} R_Z(0) & R_Z(1) & R_Z(2) & \cdots & R_Z(p-1) \\ R_Z(1) & R_Z(0) & R_Z(1) & \cdots & R_Z(p-2) \\ . & . & . & . & . \\ . & . & . & . & R_Z(1) \\ R_Z(p-1) & . & . & R_Z(1) & R_Z(0) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ . \\ . \\ h_p \end{bmatrix}$$

$$= \mathbf{R_Z}\mathbf{h}. \tag{10.86b}$$

Equations (10.86a) and (10.86b) are called the **Yule-Walker equations**.

Equation (10.80) for the mean square error becomes

$$E[e_n^2] = R_Z(0) - \sum_{\beta=1}^{p} h_\beta R_Z(\beta). \tag{10.87}$$

By inverting the $p \times p$ matrix $\mathbf{R_Z}$, we can solve for the vector of filter coefficients h.

Example 10.23 Prediction for Long-Range and Short-Range Dependent Processes

Let $X_1(t)$ be a discrete-time first-order autoregressive process with $\sigma_X^2 = 1$ and $r = 0.7411$, and let $X_2(t)$ be a discrete-time long-range dependent process with autocovariance given by Eq. (9.109), $\sigma_X^2 = 1$, and $H = 0.9$. Both processes have $C_X(1) = 0.7411$, but the autocovariance of $X_1(t)$ decreases exponentially while that of $X_2(t)$ has long-range dependence. Compare the performance of the optimal linear predictor for these processes for short-term as well as long-term predictions.

The optimum linear coefficients and the associated mean square error for the long-range dependent process can be calculated using the following code. The function can be modified for the autoregressive case.

```
function mse= Lin_Pred_LR(order,Hurst,varsig)
n=[0:1:order-1]
H2=2*Hurst
r=varsig*((1+n).^H2−2*(n.^H2)+abs(n−1).^H2)/2
rz=varsig*((2+n).^H2−2*((n+1).^H2)+(n).^H2)/2
R=toeplitz(r);
H=transpose(inv(R)*transpose(rz))
mse=varsig-H*transpose(rz)
endfunction
```

Table 10.2 below compares the mean square errors and the coefficients of the two processes in the case of short-term prediction. The predictor for $X_1(t)$ attains all of the benefit of prediction with a $p = 1$ system. The optimum predictors for higher-order systems set the other coefficients to zero, and the mean square error remains at 0.4577. The predictor for $X_2(t)$

TABLE 10.2(a)   Short-term prediction: autoregressive, $r = 0.7411, \sigma_X^2 = 1, C_X(1) = 0.7411$.

p	MSE	Coefficients	
1	0.45077	0.74110	
2	0.45077	0.74110	0

TABLE 10.2(b)   Short-term prediction: long-range dependent process, Hurst $= 0.9, \sigma_X^2 = 1, C_X(1) = 0.7411$.

p	MSE	Coefficients				
1	0.45077	0.74110				
2	0.43625	0.60809	0.17948			
3	0.42712	0.582127	0.091520	0.144649		
4	0.42253	0.567138	0.082037	0.084329	0.103620	
5	0.41964	0.558567	0.075061	0.077543	0.056707	0.082719

achieves most of the possible performance with a $p = 1$ system, but small reductions in mean square error do accrue by adding more coefficients. This is due to the persistent correlation among the values in $X_2(t)$.

Table 10.3 shows the dramatic impact of long-range dependence on prediction performance. We modified Eq. (10.86) to provide the optimum linear predictor for X_t based on two observations X_{t-10} and X_{t-20} that are in the relatively remote past. $X_1(t)$ and its previous values are almost uncorrelated, so the best predictor has a mean square error of almost 1, which is the variance of $X_1(t)$. On the other hand, $X_2(t)$ retains significant correlation with its previous values and so the mean square error provides a significant reduction from the unit variance. Note that the second-order predictor places significant weight on the observation 20 samples in the past.

TABLE 10.3(a)   Long-term prediction: autoregressive, $r = 0.7411, \sigma_X^2 = 1, C_X(1) = 0.7411$.

p	MSE	Coefficients	
1	0.99750	0.04977	
2	0.99750	0.04977	0

TABLE 10.3(b)   Long-term prediction: long-range dependent process, Hurst $= 0.9, \sigma_X^2 = 1, C_X(1) = 0.7411$.

p	MSE	Coefficients	
10	0.79354	0.45438	
10;20	0.74850	0.34614	0.23822

10.4.3   Estimation Using the Entire Realization of the Observed Process

Suppose that Z_t is to be estimated by a linear function Y_t of the entire realization of X_t, that is, $a = b = \infty$ and Eq. (10.73) becomes

$$Y_t = \sum_{\beta=-\infty}^{\infty} h_\beta X_{t-\beta}.$$

In the case of continuous-time random processes, we have

$$Y(t) = \int_{-\infty}^{\infty} h(\beta)X(t - \beta)\, d\beta.$$

The optimum filters must satisfy Eqs. (10.78) and (10.79), which in this case become

$$R_{Z,X}(m) = \sum_{\beta=-\infty}^{\infty} h_\beta R_X(m - \beta) \qquad \text{for all } m \tag{10.88a}$$

$$R_{Z,X}(\tau) = \int_{-\infty}^{\infty} h(\beta)R_X(\tau - \beta)\, d\beta \qquad \text{for all } \tau. \tag{10.88b}$$

The Fourier transform of the first equation and the Fourier transform of the second equation both yield the same expression:

$$S_{Z,X}(f) = H(f)S_X(f),$$

which is readily solved for the transfer function of the optimum filter:

$$H(f) = \frac{S_{Z,X}(f)}{S_X(f)}. \tag{10.89}$$

The impulse response of the optimum filter is then obtained by taking the appropriate inverse transform. In general the filter obtained from Eq. (10.89) will be noncausal, that is, its impulse response is nonzero for $t < 0$. We already indicated that there are applications where this makes sense, namely, in situations where the entire realization X_α is recorded and the estimate Z_t is obtained in "nonreal time" by "playing back" X_α.

Example 10.24   Infinite Smoothing

Find the transfer function for the optimum filter for estimating $Z(t)$ from $X(\alpha) = Z(\alpha) + N(\alpha)$, $\alpha \in (-\infty, \infty)$, where $Z(\alpha)$ and $N(\alpha)$ are independent, zero-mean random processes.

The cross-correlation between the observation and the desired signal is

$$\begin{aligned}
R_{Z,X}(\tau) = E[Z(t + \tau)X(t)] &= E[Z(t + \tau)(Z(t) + N(t))] \\
&= E[Z(t + \tau)Z(t)] + E[Z(t + \tau)N(t)] \\
&= R_Z(\tau),
\end{aligned}$$

since $Z(t)$ and $N(t)$ are zero-mean, independent random processes. The cross-power spectral density is then

$$S_{Z,X}(t) = S_Z(f). \tag{10.90}$$

The autocorrelation of the observation process is

$$R_X(\tau) = E[(Z(t + \tau) + N(t + \tau))(Z(t) + N(t))]$$

$$= R_Z(\tau) + R_N(\tau).$$

The corresponding power spectral density is

$$S_X(f) = S_Z(f) + S_N(f). \tag{10.91}$$

Substituting Eqs. (10.90) and (10.91) into Eq. (10.89) gives

$$H(f) = \frac{S_Z(f)}{S_Z(f) + S_N(f)}. \tag{10.92}$$

Note that the optimum filter $H(f)$ is nonzero only at the frequencies where $S_Z(f)$ is nonzero, that is, where the signal has power content. By dividing the numerator and denominator of Eq. (10.92) by $S_Z(f)$, we see that $H(f)$ emphasizes the frequencies where the ratio of signal to noise power density is large.

---

*10.4.4 Estimation Using Causal Filters

Now, suppose that Z_t is to be estimated using only the past and present of X_α, that is, $I = (-\infty, t)$. Equations (10.78) and (10.79) become

$$R_{Z,X}(m) = \sum_{\beta=0}^{\infty} h_\beta R_X(m - \beta) \qquad \text{for all } m \tag{10.93a}$$

$$R_{Z,X}(\tau) = \int_0^\infty h(\beta) R_X(\tau - \beta) \, d\beta \qquad \text{for all } \tau. \tag{10.93b}$$

Equations (10.93a) and (10.93b) are called the **Wiener-Hopf equations** and, though similar in appearance to Eqs. (10.88a) and (10.88b), are considerably more difficult to solve.

First, let us consider the special case where the observation process is white, that is, for the discrete-time case $R_X(m) = \delta_m$. Equation (10.93a) is then

$$R_{Z,X}(m) = \sum_{\beta=0}^{\infty} h_\beta \, \delta_{m-\beta} = h_m \qquad m \geq 0. \tag{10.94}$$

Thus in this special case, the optimum causal filter has coefficients given by

$$h_m = \begin{cases} 0 & m < 0 \\ R_{Z,X}(m) & m \geq 0. \end{cases}$$

The corresponding transfer function is

$$H(f) = \sum_{m=0}^{\infty} R_{Z,X}(m) e^{-j2\pi fm}. \tag{10.95}$$

Note Eq. (10.95) is *not* $S_{Z,X}(f)$, since the limits of the Fourier transform in Eq. (10.95) do not extend from $-\infty$ to $+\infty$. However, $H(f)$ can be obtained from $S_{Z,X}(f)$ by finding $h_m = \mathscr{F}^{-1}[S_{Z,X}(f)]$, keeping the causal part (i.e., h_m for $m \geq 0$) and setting the noncausal part to 0.

We now show how the solution of the above special case can be used to solve the general case. It can be shown that under very general conditions, the power spectral density of a random process can be factored into the form

$$S_X(f) = |G(f)|^2 = G(f)G^*(f), \tag{10.96}$$

where $G(f)$ and $1/G(f)$ are *causal* filters.[8] This suggests that we can find the optimum filter in two steps, as shown in Fig. 10.15. First, we pass the observation process through a "whitening" filter with transfer function $W(f) = 1/G(f)$ to produce a white noise process X'_n, since

$$S_{X'}(f) = |W(f)|^2 S_X(f) = \frac{|G(f)|^2}{|G(f)|^2} = 1 \qquad \text{for all } f.$$

Second, we find the best estimator for Z_n using the whitened observation process X'_n as given by Eq. (10.95). The filter that results from the tandem combination of the whitening filter and the estimation filter is the solution to the Wiener-Hopf equations.

The transfer function of the second filter in Fig. 10.15 is

$$H_2(f) = \sum_{m=0}^{\infty} R_{Z,X'}(m)e^{-j2\pi fm} \tag{10.97}$$

by Eq. (10.95). To evaluate Eq. (10.97) we need to find

$$\begin{aligned} R_{Z,X'}(k) &= E[Z_{n+k}X'_n] \\ &= \sum_{i=0}^{\infty} w_i E[Z_{n+k}X_{n-i}] \\ &= \sum_{i=0}^{\infty} w_i R_{Z,X}(k+i), \end{aligned} \tag{10.98}$$

where w_i is the impulse response of the whitening filter. The Fourier transform of Eq. (10.98) gives an expression that is easier to work with:

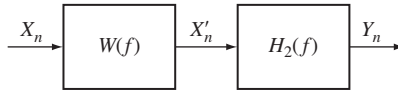$$S_{Z,X'}(f) = W^*(f)S_{Z,X}(f) = \frac{S_{Z,X}(f)}{G^*(f)}. \tag{10.99}$$

FIGURE 10.15
Whitening filter approach for solving Wiener-Hopf equations.

[8]The method for factoring $S_X(f)$ as specified by Eq. (10.96) is called **spectral factorization**. See Example 10.10 and the references at the end of the chapter.

The inverse Fourier transform of Eq. (10.99) yields the desired $R_{Z,X'}(k)$, which can then be substituted into Eq. (10.97) to obtain $H_2(f)$.

In summary, the optimum filter is found using the following procedure:

1. Factor $S_X(f)$ as in Eq. (10.96) and obtain a causal whitening filter $W(f) = 1/G(f)$.
2. Find $R_{Z,X'}(k)$ from Eq. (10.98) or from Eq. (10.99).
3. $H_2(f)$ is then given by Eq. (10.97).
4. The optimum filter is then

$$H(f) = W(f)H_2(f). \qquad (10.100)$$

This procedure is valid for the continuous-time version of the optimum causal filter problem, after appropriate changes are made from summations to integrals. The following example considers a continuous-time problem.

Example 10.25   Wiener Filter

Find the optimum causal filter for estimating a signal $Z(t)$ from the observation $X(t) = Z(t) + N(t)$, where $Z(t)$ and $N(t)$ are independent random processes, $N(t)$ is zero-mean white noise density 1, and $Z(t)$ has power spectral density

$$S_Z(f) = \frac{2}{1 + 4\pi^2 f^2}.$$

The optimum filter in this problem is called the **Wiener filter**.

The cross-power spectral density between $Z(t)$ and $X(t)$ is

$$S_{Z,X}(f) = S_Z(f),$$

since the signal and noise are independent random processes. The power spectral density for the observation process is

$$S_X(f) = S_Z(f) + S_N(f)$$

$$= \frac{3 + 4\pi^2 f^2}{1 + 4\pi^2 f^2}$$

$$= \left(\frac{j2\pi f + \sqrt{3}}{j2\pi f + 1}\right)\left(\frac{-j2\pi f + \sqrt{3}}{-j2\pi f + 1}\right).$$

If we let

$$G(f) = \frac{j2\pi f + \sqrt{3}}{j2\pi f + 1},$$

then it is easy to verify that $W(f) = 1/G(f)$ is the whitening causal filter.

Next we evaluate Eq. (10.99):

$$S_{Z,X'}(f) = \frac{S_{Z,X}(f)}{G^*(f)} = \frac{2}{1 + 4\pi^2 f^2}\frac{1 - j2\pi f}{\sqrt{3} - j2\pi f}$$

$$= \frac{2}{(1 + j2\pi f)(\sqrt{3} - j2\pi f)}$$

$$= \frac{c}{1 + j2\pi f} + \frac{c}{\sqrt{3} - j2\pi f}, \qquad (10.101)$$

where $c = 2/(1 + \sqrt{3})$. If we take the inverse Fourier transform of $S_{Z,X'}(f)$, we obtain

$$R_{Z,X'}(\tau) = \begin{cases} ce^{-\tau} & \tau > 0 \\ ce^{\sqrt{3}\tau} & \tau < 0. \end{cases}$$

Equation (10.97) states that $H_2(f)$ is given by the Fourier transform of the $\tau > 0$ portion of $R_{Z,X'}(\tau)$:

$$H_2(f) = \mathscr{F}\{ce^{-\tau}u(\tau)\} = \frac{c}{1 + j2\pi f}.$$

Note that we could have gotten this result directly from Eq. (10.101) by noting that only the first term gives rise to the positive-time (i.e., causal) component.

The optimum filter is then

$$H(f) = \frac{1}{G(f)}H_2(f) = \frac{c}{\sqrt{3} + j2\pi f}.$$

The impulse response of this filter is

$$h(t) = ce_t^{-\sqrt{3}} \qquad t > 0.$$

10.5   THE KALMAN FILTER

The optimum linear systems considered in the previous section have two limitations: (1) They assume wide-sense stationary signals; and (2) The number of equations grows with the size of the observation set. In this section, we consider an estimation approach that assumes signals have a certain structure. This assumption keeps the dimensionality of the problem fixed even as the observation set grows. It also allows us to consider certain nonstationary signals.

We will consider the class of signals that can be represented as shown in Fig. 10.16(a):

$$Z_n = a_{n-1}Z_{n-1} + W_{n-1} \qquad n = 1, 2, \ldots, \tag{10.102}$$

where Z_0 is the random variable at time 0, a_n is a known sequence of constants, and W_n is a sequence of zero-mean uncorrelated random variables with possibly time-varying variances $\{E[W_n^2]\}$. The resulting process Z_n is nonstationary in general. We assume that the process Z_n is not available to us, and that instead, as shown in Fig. 10.16(a), we observe

$$X_n = Z_n + N_n \qquad n = 0, 1, 2, \ldots, \tag{10.103}$$

where the observation noise N_n is a zero-mean, uncorrelated sequence of random variables with possibly time-varying variances $\{E[N_n^2]\}$. We assume that W_n and N_n are uncorrelated at all times n_1 and n_2. In the special case where W_n and N_n are Gaussian random processes, then Z_n and X_n will also be Gaussian random processes. We will develop the Kalman filter, which has the structure in Fig. 10.16(b).

Our objective is to find *for each time n* the minimum mean square estimate (actually prediction) of Z_n based on the observations $X_0, X_1, \ldots, X_{n-1}$ using a linear estimator that possibly varies with time:

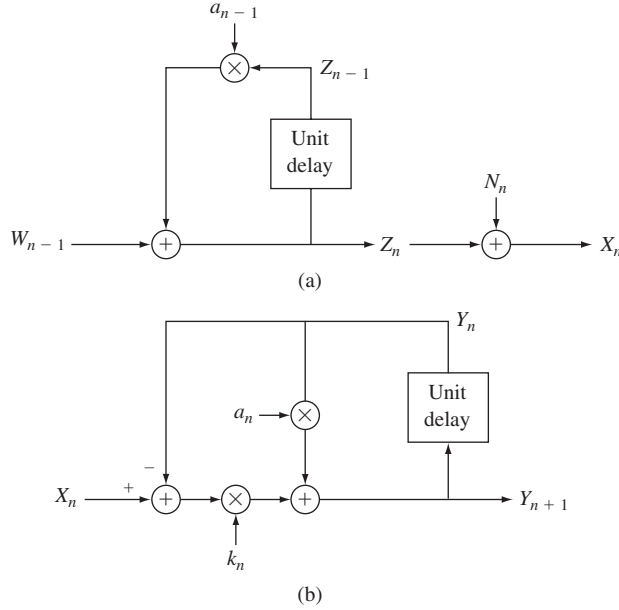$$Y_n = \sum_{j=i}^{n} h_j^{(n-1)} X_{n-j}. \tag{10.104}$$

FIGURE 10.16
(a) Signal structure. (b) Kalman filter.

The orthogonality principle implies that the optimum filter $\{h_j^{(n-1)}\}$ satisfies

$$E\left[\left(Z_n - \sum_{j=1}^{n} h_j^{(n-1)} X_{n-j}\right) X_l\right] = 0 \qquad \text{for } l = 0, 1, \ldots, n-1,$$

which leads to a set of n equations in n unknowns:

$$R_{Z,X}(n, l) = \sum_{j=1}^{n} h_j^{(n-1)} R_X(n - j, l) \qquad \text{for } l = 0, 1, \ldots, n-1. \quad (10.105)$$

At the next time instant, we need to find

$$Y_{n+1} = \sum_{j=1}^{n+1} h_j^{(n)} X_{n+1-j} \qquad (10.106)$$

by solving a system of $(n + 1) \times (n + 1)$ equations:

$$R_{Z,X}(n + 1, l) = \sum_{j=1}^{n+1} h_j^{(n)} R_X(n + 1 - j, l) \qquad \text{for } l = 0, 1, \ldots, n. \quad (10.107)$$

Up to this point we have followed the procedure of the previous section and we find that the dimensionality of the problem grows with the number of observations. We now use the signal structure to develop a recursive method for solving Eq. (10.106).

We first need the following two results: For $l < n$, we have

$$R_{Z,X}(n + 1, l) = E[Z_{n+1}X_l] = E[(a_nZ_n + W_n)X_l]$$

$$= a_nR_{Z,X}(n, l) + E[W_nX_l] = a_nR_{Z,X}(n, l), \qquad (10.108)$$

since $E[W_nX_l] = E[W_n]E[X_l] = 0$, that is, W_n is uncorrelated with the past of the process and the observations prior to time n, as can be seen from Fig. 10.16(a). Also for $l < n$, we have

$$R_{Z,X}(n, l) = E[Z_nX_l] = E[(X_n - N_n)X_l]$$

$$= R_X(n, l) - E[N_nX_l] = R_X(n, l), \qquad (10.109)$$

since $E[N_nX_l] = E[N_n]E[X_l] = 0$, that is, the observation noise at time n is uncorrelated with prior observations.

We now show that the set of equations in Eq. (10.107) can be related to the set in Eq. (10.105). For $l < n$, we can equate the right-hand sides of Eqs. (10.108) and (10.107):

$$a_nR_{Z,X}(n, l) = \sum_{j=1}^{n+1} h_j^{(n)} R_X(n + 1 - j, l)$$

$$= h_1^{(n)}R_X(n, l) + \sum_{j=2}^{n+1} h_j^{(n)} R_X(n + 1 - j, l)$$

$$\text{for } l = 0, 1, \ldots, n - 1. \ (10.110)$$

From Eq. (10.109) we have $R_X(n, l) = R_{Z,X}(n, l)$, so we can replace the first term on the right-hand of Eq. (10.110) and then move the resulting term to the left-hand side:

$$(a_n - h_1^{(n)})R_{Z,X}(n, l) = \sum_{j=2}^{n+1} h_j^{(n)} R_X(n + 1 - j, l)$$

$$= \sum_{j'=1}^{n} h_{j'+1}^{(n)} R_X(n - j', l). \qquad (10.111)$$

By dividing both sides by $a_n - h_1^{(n)}$ we finally obtain

$$R_{Z,X}(n, l) = \sum_{j'=1}^{n} \frac{h_{j'+1}^{(n)}}{a_n - h_1^{(n)}} R_X(n - j', l)$$

$$\text{for } l = 0, 1, \ldots, n - 1. \ (10.112)$$

This set of equations is identical to Eq. (10.105) if we set

$$h_j^{(n-1)} = \frac{h_{j+1}^{(n)}}{a_n - h_1^{(n)}} \qquad \text{for } j = 1, \ldots, n. \qquad (10.113a)$$

Therefore, if at step n we have found $h_1^{(n-1)}, \ldots, h_n^{(n-1)}$, and if somehow we have found $h_1^{(n)}$, then we can find the remaining coefficients from

$$h_{j+1}^{(n)} = (a_n - h_1^{(n)})h_j^{(n-1)} \qquad j = 1, \ldots, n. \qquad (10.113b)$$

Thus the key question is how to find $h_1^{(n)}$.

Suppose we substitute the coefficients in Eq. (10.113b) into Eq. (10.106):

$$Y_{n+1} = h_1^{(n)} X_n + \sum_{j'=1}^{n} (a_n - h_1^{(n)}) h_{j'}^{(n-1)} X_{n-j'}$$

$$= h_1^{(n)} X_n + (a_n - h_1^{(n)}) Y_n$$

$$= a_n Y_n + h_1^{(n)} (X_n - Y_n), \tag{10.114}$$

where the second equality follows from Eq. (10.104). The above equation has a very pleasing interpretation, as shown in Fig. 10.16(b). Since Y_n is the prediction for time n, $a_n Y_n$ is the prediction for the next time instant, $n + 1$, based on the "old" information (see Eq. (10.102)). The term $(X_n - Y_n)$ is called the "innovations," and it gives the discrepancy between the old prediction and the observation. Finally, the term $h_1^{(n)}$ is called the *gain*, henceforth denoted by k_n, and it indicates the extent to which the innovations should be used to correct $a_n Y_n$ to obtain the "new" prediction Y_{n+1}. If we denote the **innovations** by

$$I_n = X_n - Y_n \tag{10.115}$$

then Eq. (10.114) becomes

$$Y_{n+1} = a_n Y_n + k_n I_n. \tag{10.116}$$

We still need to determine a means for computing the gain k_n.

From Eq. (10.115), we have that the innovations satisfy

$$I_n = X_n - Y_n = Z_n + N_n - Y_n = Z_n - Y_n + N_n = \varepsilon_n + N_n,$$

where $\varepsilon_n = Z_n - Y_n$ is the prediction error. A recursive equation can be obtained for the prediction error:

$$\varepsilon_{n+1} = Z_{n+1} - Y_{n+1} = a_n Z_n + W_n - a_n Y_n - k_n I_n$$

$$= a_n(Z_n - Y_n) + W_n - k_n(\varepsilon_n + N_n)$$

$$= (a_n - k_n)\varepsilon_n + W_n - k_n N_n, \tag{10.117}$$

with initial condition $\varepsilon_0 = Z_0$. Since X_0, W_n, and N_n are zero-mean, it then follows that $E[\varepsilon_n] = 0$ for all n. A recursive equation for the mean square prediction error is obtained from Eq. (10.117):

$$E[\varepsilon_{n+1}^2] = (a_n - k_n)^2 E[\varepsilon_n^2] + E[W_n^2] + k_n^2 E[N_n^2], \tag{10.118}$$

with initial condition $E[\varepsilon_0^2] = E[Z_0^2]$. We are finally ready to obtain an expression for the gain k_n.

The gain k_n must minimize the mean square error $E[\varepsilon_{n+1}^2]$. Therefore we can differentiate Eq. (10.118) with respect to k_n and set it equal to zero:

$$0 = -2(a_n - k_n)E[\varepsilon_n^2] + 2k_n E[N_n^2].$$

Then we can solve for k_n:

$$k_n = \frac{a_n E[\varepsilon_n^2]}{E[\varepsilon_n^2] + E[N_n^2]}. \tag{10.119}$$

The expression for the mean square prediction error in Eq. (10.118) can be simplified by using Eq. (10.119) (see Problem 10.72):

$$E[\varepsilon_{n+1}^2] = a_n(a_n - k_n)E[\varepsilon_n^2] + E[W_n^2]. \tag{10.120}$$

Equations (10.119), (10.116), and (10.120) when combined yield the recursive procedure that constitutes the Kalman filtering algorithm:

Kalman filter algorithm:[9]

Initialization: $Y_0 = 0$ $E[\varepsilon_0^2] = E[Z_0^2]$
For $n = 0, 1, 2, \ldots$

$$k_n = \frac{a_n E[\varepsilon_n^2]}{E[\varepsilon_n^2] + E[N_n^2]}$$

$$Y_{n+1} = a_n Y_n + k_n(X_n - Y_n)$$

$$E[\varepsilon_{n+1}^2] = a_n(a_n - k_n)E[\varepsilon_n^2] + E[W_n^2].$$

Note that the algorithm requires knowledge of the signal structure, i.e., the a_n, and the variances $E[N_n^2]$ and $E[W_n^2]$. The algorithm can be implemented easily and has consequently found application in a broad range of detection, estimation, and signal processing problems. The algorithm can be extended in matrix form to accommodate a broader range of processes.

Example 10.26   First-Order Autoregressive Process

Consider a signal defined by

$$Z_n = aZ_{n-1} + W_n \qquad n = 1, 2, \ldots \qquad Z_0 = 0,$$

where $E[W_n^2] = \sigma_W^2 = 0.36$, and $a = 0.8$, and suppose the observations are made in additive white noise

$$X_n = Z_n + N_n \qquad n = 0, 1, 2, \ldots,$$

where $E[N_n^2] = 1$. Find the form of the predictor and its mean square error as $n \to \infty$.

The gain at step n is given by

$$k_n = \frac{a E[\varepsilon_n^2]}{E[\varepsilon_n^2] + 1}.$$

The mean square error sequence is therefore given by

$$E[\varepsilon_0^2] = E[Z_0^2] = 0$$

---

[9]We caution the student that there are two common ways of defining the gain. The statement of the Kalman filter algorithm will differ accordingly in various textbooks.

$$E[\varepsilon_{n+1}^2] = a(a - k_n)E[\varepsilon_n^2] + \sigma_W^2$$

$$= a\left(\frac{a}{1 + E[\varepsilon_n^2]}\right)E[\varepsilon_n^2] + \sigma_W^2 \qquad \text{for } n = 1, 2, \ldots.$$

The steady state mean square error e_∞ must satisfy

$$e_\infty = \frac{a^2}{1 + e_\infty}e_\infty + \sigma_W^2.$$

For $a = 0.8$ and $\sigma_W^2 = 0.36$, the resulting quadratic equation yields $k_\infty = 0.3$ and $e_\infty = 0.6$. Thus at steady state the predictor is

$$Y_{n+1} = 0.8Y_n + 0.3(X_n - Y_n).$$

*10.6 ESTIMATING THE POWER SPECTRAL DENSITY

Let $X_0, \ldots, X_{k-1}$ be k observations of the discrete-time, zero-mean, wide-sense stationary process X_n. The periodogram estimate for $S_X(f)$ is defined as

$$\tilde{p}_k(f) = \frac{1}{k}|\tilde{x}_k(f)|^2, \tag{10.121}$$

where $\tilde{x}_k(f)$ is obtained as a Fourier transform of the observation sequence:

$$\tilde{x}_k(f) = \sum_{m=0}^{k-1} X_m e^{-j2\pi fm}. \tag{10.122}$$

In Section 10.1 we showed that the expected value of the periodogram estimate is

$$E[\tilde{p}_k(f)] = \sum_{m'=-(k-1)}^{k-1}\left\{1 - \frac{|m'|}{k}\right\}R_X(m')e^{-j2\pi fm'}, \tag{10.123}$$

so $\tilde{p}_k(f)$ is a biased estimator for $S_X(f)$. However, as $k \to \infty$,

$$E[\tilde{p}_k(f)] \to S_X(f), \tag{10.124}$$

so the mean of the periodogram estimate approaches $S_X(f)$.

Before proceeding to find the variance of the periodogram estimate, we note that the periodogram estimate is equivalent to taking the Fourier transform of an estimate for the autocorrelation sequence; that is,

$$\tilde{p}_k(f) = \sum_{m=-(k-1)}^{k-1} \hat{r}_k(m)e^{-j2\pi fm}, \tag{10.125}$$

where the estimate for the autocorrelation is

$$\hat{r}_k(m) = \frac{1}{k}\sum_{n=0}^{k-|m|-1} X_n X_{n+m}. \tag{10.126}$$

(See Problem 10.77.)
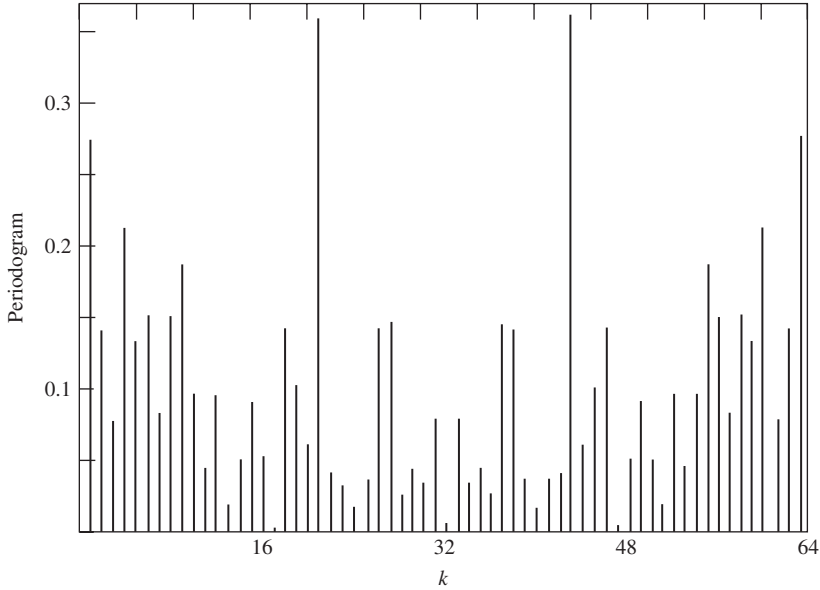
FIGURE 10.17
Periodogram for 64 samples of white noise sequence X_n iid uniform in (0, 1), $S_X(f) = \sigma_X^2 = 1/12 = 0.083$.

We might expect that as we increase the number of samples k, the periodogram estimate converges to $S_X(f)$. This does not happen. Instead we find that $\widetilde{p}_k(f)$ fluctuates wildly about the true spectral density, and that this random variation does not decrease with increased k (see Fig. 10.17). To see why this happens, in the next section we compute the statistics of the periodogram estimate for a white noise Gaussian random process. We find that the estimates given by the periodogram have a variance that does *not* approach zero as the number of samples is increased. This explains the lack of improvement in the estimate as k is increased. Furthermore, we show that the periodogram estimates are uncorrelated at uniformly spaced frequencies in the interval $-1/2 \le f < 1/2$. This explains the erratic appearance of the periodogram estimate as a function of f. In the final section, we obtain another estimate for $S_X(f)$ whose variance does approach zero as k increases.

10.6.1 Variance of Periodogram Estimate

Following the approach of [Jenkins and Watts, pp. 230–233], we consider the periodogram of samples of a white noise process with $S_X(f) = \sigma_X^2$ at the frequencies $f = n/k, -k/2 \le n < k/2$, which will cover the frequency range $-1/2 \le f < 1/2$. (In practice these are the frequencies we would evaluate if we were using the FFT algorithm to compute $\widetilde{x}_k(f)$.) First we rewrite Eq. (10.122) at $f = n/k$ as follows:

$$\widetilde{x}_k\left(\frac{n}{k}\right) = \sum_{m=0}^{k-1} X_m\left(\cos\left(\frac{2\pi mn}{k}\right) - j\sin\left(\frac{2\pi mn}{k}\right)\right)$$

$$= A_k(n) - jB_k(n) \qquad -k/2 \le n < k/2, \tag{10.127}$$

where

$$A_k(n) = \sum_{m=0}^{k-1} X_m \cos\left(\frac{2\pi mn}{k}\right) \tag{10.128}$$

and

$$B_k(n) = \sum_{m=0}^{k-1} X_m \sin\left(\frac{2\pi mn}{k}\right). \tag{10.129}$$

Then it follows that the periodogram estimate is

$$\widetilde{p}_k\left(\frac{n}{k}\right) = \frac{1}{k}\left|\hat{x}_k\left(\frac{n}{k}\right)\right|^2 = \frac{1}{k}\{A_k^2(n) + B_k^2(n)\}. \tag{10.130}$$

We find the variance of $\widetilde{p}_k(n/k)$ from the statistics of $A_k(n)$ and $B_k(n)$.

The random variables $A_k(n)$ and $B_k(n)$ are defined as linear functions of the jointly Gaussian random variables $X_0, \ldots, X_{k-1}$. Therefore $A_k(n)$ and $B_k(n)$ are also jointly Gaussian random variables. If we take the expected value of Eqs. (10.128) and (10.129) we find

$$E[A_k(n)] = 0 = E[B_k(n)] \qquad \text{for all } n. \tag{10.131}$$

Note also that the $n = -k/2$ and $n = 0$ terms are different in that

$$B_k(-k/2) = 0 = B_k(0) \tag{10.132a}$$

$$A_k(-k/2) = \sum_{i=0}^{k-1}(-1)^i X_i \qquad A_k(0) = \sum_{i=0}^{k-1} X_i. \tag{10.132b}$$

The correlation between $A_k(n)$ and $A_k(m)$ (for n, m not equal to $-k/2$ or 0) is

$$E[A_k(n)A_k(m)] = \sum_{i=0}^{k-1}\sum_{l=0}^{k-1} E[X_i X_l] \cos\left(\frac{2\pi ni}{k}\right) \cos\left(\frac{2\pi ml}{k}\right)$$

$$= \sigma_X^2 \sum_{i=0}^{k-1} \cos\left(\frac{2\pi ni}{k}\right) \cos\left(\frac{2\pi mi}{k}\right)$$

$$= \sigma_X^2 \sum_{i=0}^{k-1} \frac{1}{2}\cos\left(\frac{2\pi(n-m)i}{k}\right) + \sigma_X^2 \sum_{i=0}^{k-1}\frac{1}{2}\cos\left(\frac{2\pi(n+m)i}{k}\right),$$

where we used the fact that $E[X_i X_1] = \sigma_X^2 \delta_{il}$ since the noise is white. The second summation is equal to zero, and the first summation is zero except when $n = m$. Thus

$$E[A_k(n)A_k(m)] = \frac{1}{2}k\sigma_X^2\,\delta_{nm} \qquad \text{for all } n, m \neq -k/2, 0. \tag{10.133a}$$

It can similarly be shown that

$$E[B_k(n)B_k(m)] = \frac{1}{2}k\sigma_X^2\,\delta_{nm} \qquad n, m \neq 0 - k/2, 0 \tag{10.133b}$$

$$E[A_k(n)B_k(m)] = 0 \qquad \text{for all } n, m. \tag{10.133c}$$

When $n = -k/2$ or 0, we have

$$E[A_k(n)A_k(m)] = k\sigma_X^2 \, \delta_{nm} \qquad \text{for all } m. \tag{10.133d}$$

Equations (10.133a) through (10.133d) imply that $A_k(n)$ and $B_k(m)$ are uncorrelated random variables. Since $A_k(n)$ and $B_k(n)$ are jointly Gaussian random variables, this implies that they are zero-mean, *independent* Gaussian random variables.

   We are now ready to find the statistics of the periodogram estimates at the frequencies $f = n/k$. Equation (10.130) gives

$$\widetilde{p}_k\left(\frac{n}{k}\right) = \frac{1}{k}\{A_k^2(n) + B_k^2(n)\} \qquad n \neq -k/2, 0$$

$$= \frac{1}{2}\sigma_X^2 \left\{\frac{A_k^2(n)}{(1/2)k\sigma_X^2} + \frac{B_k^2(n)}{(1/2)k\sigma_X^2}\right\}. \tag{10.134}$$

The quantity in brackets is the sum of the squares of two zero-mean, unit-variance, independent Gaussian random variables. This is a chi-square random variable with two degrees of freedom (see Problem 7.6). From Table 4.1, we see that a chi-square random variable with v degrees of freedom has variance $2v$. Thus the expression in the brackets has variance 4, and the periodogram estimate $\hat{p}_k(n/k)$ has variance

$$\text{VAR}\left[\widetilde{p}_k\left(\frac{n}{k}\right)\right] = \left(\frac{1}{2}\sigma_X^2\right)^2 4 = \sigma_X^4 = S_X(f)^2. \tag{10.135a}$$

For $n = -k/2$ and $n = 0$,

$$\widetilde{p}_k\left(\frac{n}{k}\right) = \sigma_X^2 \left\{\frac{A_k^2(n)}{k\sigma_X^2}\right\}.$$

The quantity in brackets is a chi-square random variable with one degree of freedom and variance 2, so the variance of the periodogram estimate is

$$\text{VAR}\left[\widetilde{p}_k\left(\frac{n}{k}\right)\right] = 2\sigma_X^4 \qquad n = -k/2, 0. \tag{10.135b}$$

   Thus we conclude from Eqs. (10.135a) and (10.135b) that *the variance of the periodogram estimate is proportional to the square of the power spectral density and does not approach zero as k increases*. In addition, Eqs. (10.133a) through (10.133d) imply that *the periodogram estimates at the frequencies $f = -n/k$ are uncorrelated random variables*. A more detailed analysis [Jenkins and Watts, p. 238] shows that for arbitrary f,

$$\text{VAR}[\widetilde{p}_k(f)] = S_X(f)^2\left\{1 + \left(\frac{\sin(2\pi f k)}{k \sin(2\pi f)}\right)^2\right\}. \tag{10.136}$$

Thus variance of the periodogram estimate does not approach zero as the number of samples is increased.

   The above discussion has only considered the spectrum estimation for a white noise, Gaussian random process, but the general conclusions are also valid for non-white, non-Gaussian processes. If the X_i are not Gaussian, we note from Eqs. (10.128)

and (10.129) that A_k and B_k are approximately Gaussian by the central limit theorem if k is large. Thus the periodogram estimate is then approximately a chi-square random variable.

If the process X_i is not white, then it can be viewed as filtered white noise:

$$X_n = h_n * W_n,$$

where $S_W(f) = \sigma_W^2$ and $|H(f)|^2 S_W(f) = S_X(f)$. The periodograms of X_n and W_n are related by

$$\frac{1}{k}\left|\widetilde{x}_k\left(\frac{n}{k}\right)\right|^2 = \frac{1}{k}\left|H\left(\frac{n}{k}\right)\right|^2 \left|\widetilde{w}_k\left(\frac{n}{k}\right)\right|^2. \tag{10.137}$$

Thus

$$\left|\widetilde{w}_k\left(\frac{n}{k}\right)\right|^2 = \frac{|\widetilde{x}_k(n/k)|^2}{|H(n/k)|^2}. \tag{10.138}$$

From our previous results, we know that $|\widetilde{w}_k(n/k)|^2/k$ is a chi-square random variable with variance σ_W^4. This implies that

$$\mathrm{VAR}\left[\frac{|\widetilde{x}_k(n/k)|^2}{k}\right] = \left|H\left(\frac{n}{k}\right)\right|^4 \sigma_W^4 = S_X(f)^2. \tag{10.139}$$

Thus we conclude that the variance of the periodogram estimate for nonwhite noise is also proportional to $S_X(f)^2$.

10.6.2 Smoothing of Periodogram Estimate

A fundamental result in probability theory is that the sample mean of a sequence of *independent* realizations of a random variable approaches the true mean with probability one. We obtain an estimate for $S_X(f)$ that goes to zero with the number of observations k by taking the average of N *independent* periodograms on samples of size k:

$$\langle \widetilde{p}_k(f) \rangle_N = \frac{1}{N}\sum_{i=1}^{N} \widetilde{p}_{k,i}(f), \tag{10.140}$$

where $\{\widetilde{p}_{k,i}(f)\}$ are N independent periodograms computed using separate sets of k samples each. Figures 10.18 and 10.19 show the $N = 10$ and $N = 50$ smoothed periodograms corresponding to the unsmoothed periodogram of Fig. 10.17. It is evident that the variance of the power spectrum estimates is decreasing with N.
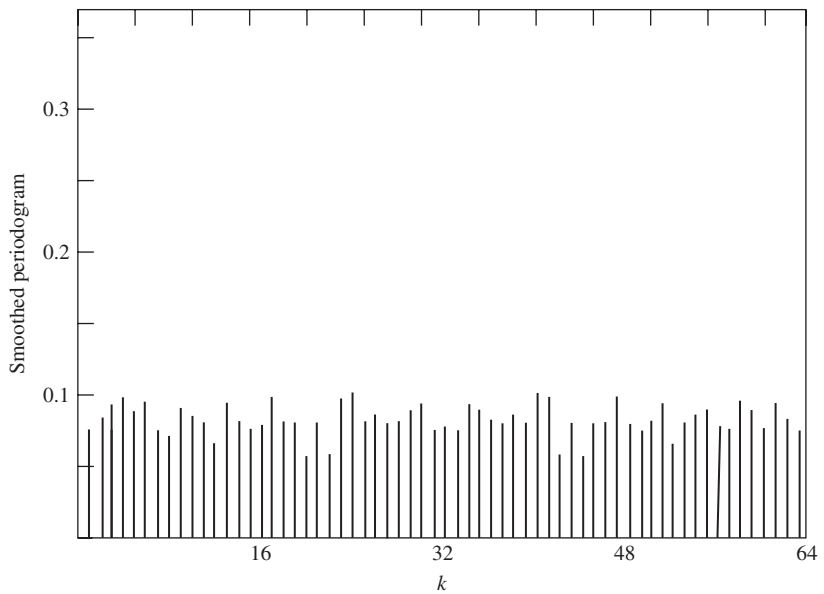
The mean of the smoothed estimator is

$$E\langle \widetilde{p}_k(f) \rangle_N = \frac{1}{N}\sum_{i=1}^{N} E[\widetilde{p}_{k,i}(f)] = E[\widetilde{p}_k(f)]$$

$$= \sum_{m'=-(k-1)}^{k-1}\left\{1 - \frac{|m'|}{k}\right\} R_X(m') e^{-j2\pi f m'}, \tag{10.141}$$

where we have used Eq. (10.35). Thus the smoothed estimator has the same mean as the periodogram estimate on a sample of size k.
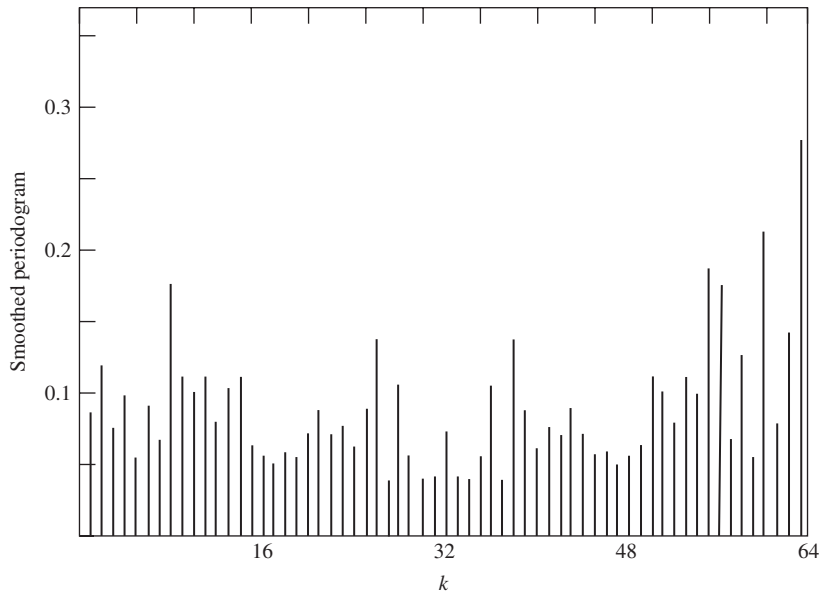
FIGURE 10.18
Sixty-four-point smoothed periodogram with $N = 10$, X_n iid uniform in $(0, 1)$,
$S_X(f) = 1/12 = 0.083$.

FIGURE 10.19
Sixty-four-point smoothed periodogram with $N = 50$, X_n iid uniform in $(0, 1)$,
$S_X(f) = 1/12 = 0.083$.

The variance of the smoothed estimator is

$$\mathrm{VAR}[\langle \widetilde{p}_k(f) \rangle_N] = \frac{1}{N^2} \sum_{i=1}^{N} \mathrm{VAR}[\widetilde{p}_{k,i}(f)]$$

$$= \frac{1}{N} \mathrm{VAR}[\widetilde{p}_k(f)]$$

$$\simeq \frac{1}{N} S_X(f)^2.$$

Thus the variance of the smoothed estimator can be reduced by increasing N, the number of periodograms used in Eq. (10.140).

In practice, a sample set of size Nk, $X_0, \ldots, X_{Nk-1}$ is divided into N blocks and a separate periodogram is computed for each block. The smoothed estimate is then the average over the N periodograms. This method is called **Bartlett's smoothing procedure**. Note that, in general, the resulting periodograms are not independent because the underlying blocks are not independent. Thus this smoothing procedure must be viewed as an approximation to the computation and averaging of independent periodograms.

The choice of k and N is determined by the desired frequency resolution and variance of the estimate. The blocksize k determines the number of frequencies for which the spectral density is computed (i.e., the frequency resolution). The variance of the estimate is controlled by the number of periodograms N. The actual choice of k and N depends on the nature of the signal being investigated.

10.7 NUMERICAL TECHNIQUES FOR PROCESSING RANDOM SIGNALS

In this chapter our discussion has combined notions from random processes with basic concepts from signal processing. The processing of signals is a very important area in modern technology and a rich set of techniques and methodologies have been developed to address the needs of specific application areas such as communication systems, speech compression, speech recognition, video compression, face recognition, network and service traffic engineering, etc. In this section we briefly present a number of general tools available for the processing of random signals. We focus on the tools provided in Octave since these are quite useful as well as readily available.

10.7.1 FFT Techniques

The Fourier transform relationship between $R_X(\tau)$ and $S_X(f)$ is fundamental in the study of wide-sense stationary processes and plays a key role in random signal analysis. The fast fourier transform (FFT) methods we developed in Section 7.6 can be applied to the numerical transformation from autocorrelation functions to power spectral densities and back.

Consider the computation of $R_X(\tau)$ and $S_X(f)$ for continuous-time processes:

$$R_X(\tau) = \int_{-\infty}^{\infty} S_X(f) e^{-j2\pi f \tau} \, df \approx \int_{-W}^{W} S_X(f) e^{-j2\pi f \tau} \, df.$$

First we limit the integral to the region where $S_X(f)$ has significant power. Next we restrict our attention to a discrete set of $N = 2M$ frequency values at kf_0 so that $-W = -Mf_0 < (-M + 1)f_0 < \cdots < (M - 1)f_0 < W$, and then approximate the integral by a sum:

$$R_X(\tau) \approx \sum_{m=-M}^{M-1} S_X(mf_0)e^{-j2\pi mf_0\tau}f_0.$$

Finally, we also focus on a set of discrete lag values: kt_0 so that $-T = -Mt_0 < (-M + 1)t_0 < \cdots < (M - 1)t_0 < T$. We obtain the DFT as follows:

$$R_X(kt_0) \approx f_0 \sum_{m=-M}^{M-1} S_X(mf_0)e^{-j2\pi mkt_0f_0} = f_0 \sum_{m=-M}^{M-1} S_X(mf_0)e^{-j2\pi mk/N}. \quad (10.142)$$

In order to have a discrete Fourier transform, *we must have* $t_0f_0 = 1/N$, which is equivalent to: $t_0 = 1/Nf_0$ and $T = Mt_0 = 1/2f_0$ and $W = Mf_0 = 1/2t_0$. We can use the FFT function introduced in Section 7.6 to perform the transformation in Eq. (10.142) to obtain the set of values $\{R_X(kt_0), k \in [-M, M - 1]\}$ from $\{S_X(mt_0), k \in [-M, M - 1]\}$. The transformation in the reverse direction is done in the same way. Since $R_X(\tau)$ and $S_X(f)$ are even functions various simplifications are possible. We discuss some of these in the problems.

   Consider the computation of $S_X(f)$ and $R_X(k)$ for discrete-time processes. $S_X(f)$ spans the range of frequencies $|f| < 1/2$, so we restrict attention to N points $1/N$ apart:

$$S_X\left(\frac{m}{N}\right) = \sum_{k=-\infty}^{\infty} R_X(k)e^{-j2\pi kf}\bigg|_{f=m/N} \approx \sum_{k=-M}^{M-1} R_X(k)e^{-j2\pi km/N}. \quad (10.143)$$

The approximation here involves neglecting autocorrelation terms outside $[-M, M - 1]$. Since $df \approx 1/N$, the transformation in the reverse direction is scaled differently:

$$R_X(k) = \int_{-1/2}^{1/2} S_X(f)e^{-j2\pi kf}\,df \approx \frac{1}{N}\sum_{k=-M}^{M-1} S_X\left(\frac{m}{N}\right)e^{-j2\pi km/N}. \quad (10.144)$$

We assume that the student has already tried the FFT exercises in Section 7.6, so we leave examples in the use of the FFT to the Problems.

   The various frequency domain results for linear systems that relate input, output, and cross-spectral densities can be evaluated numerically using the FFT.

Example 10.27   Output Autocorrelation and Cross-Correlation

Consider Example 10.12, where a random telegraph signal $X(t)$ with $\alpha = 1$ is passed through a lowpass filter with $\beta = 1$ and $\beta = 10$. Find $R_Y(\tau)$.

   The random telegraph has $S_X(f) = \alpha/(\alpha^2 + \pi^2f^2)$ and the filter has transfer function $H(f) = \beta/(\beta + j2\pi f)$, so $R_Y(\tau)$ is given by:

$$R_Y(\tau) = \mathscr{F}^{-1}\big\{|H(f)|^2 S_X(f)\big\} = \int_{-\infty}^{\infty} \frac{\beta^2}{\beta^2 + 4\pi^2 f^2}\frac{\alpha^2}{\alpha^2 + 4\pi^2 f^2}\,df.$$
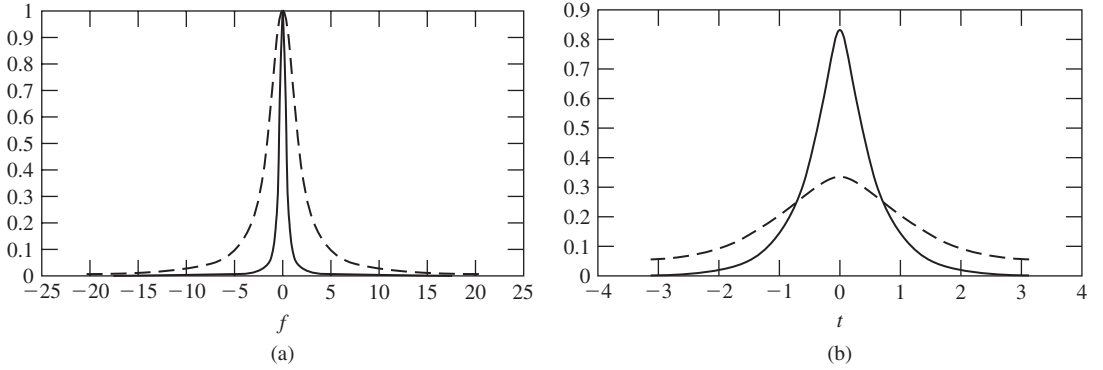
FIGURE 10.20
(a) Transfer function and input power spectral density; (b) Autocorrelation of filtered random telegraph with filter $\beta = 10$.

We used an $N = 256$ FFT to evaluate autocorrelation functions numerically for $\alpha = 1$ and $\beta = 1$ and $\beta = 10$. Figure 10.20(a) shows $|H(f)|^2$ and $S_X(f)$ for $\beta = 10$. It can be seen that the transfer function (the dashed line) is close to 1 in the region of f where $S_X(f)$ has most of its power. Consequently we expect the output for $\beta = 10$ to have an autocorrelation similar to that of the input. For $\beta = 1$, on the other hand, the filter will attenuate more of the significant frequencies of $X(t)$ and we expect more change in the output autocorrelation. Figure 10.20(b) shows the output autocorrelation and we see that indeed for $\beta = 10$ (the solid line), $R_Y(\tau)$ is close to the double-sided exponential of $R_X(\tau)$. For $\beta = 1$ the output autocorrelation differs significantly from $R_X(\tau)$.

10.7.2 Filtering Techniques

The autocorrelation and power spectral density functions provide us with information about the average behavior of the processes. We are also interested in obtaining sample functions of the inputs and outputs of systems. For linear systems the principal tools for signal processing are the convolution and Fourier transform.

Convolution in discrete-time (Eq. (10.48)) is quite simple and so convolution is the workhorse in linear signal processing. Octave provides several functions for performing convolutions with discrete-time signals. In Example 10.15 we encountered the function `filter(b,a,x)` which implements filtering of the sequence x with an ARMA filter with coefficients specified by vectors b and a in the following equation.

$$Y_n = -\sum_{i=1}^{q} \alpha_i Y_{n-i} + \sum_{j=0}^{p} \beta_j X_{n-j}.$$

Other functions use `filter(b,a,x)` to provide special cases of filtering. For example, `conv(a,b)` convolves the elements in the vectors a and b. We can obtain the output of a linear system by letting a be the impulse response and b the input random sequence. The moving average example in Fig. 10.7(b) is easily obtained using this `conv`. Octave provides other functions implementing specific digital filters.

We can also obtain the output of a linear system in the frequency domain. We take the FFT of the input sequence X_n and we then multiply it by the FFT of the transfer function. The inverse FFT will then provide Y_n of the linear system. The Octave function `fftconv(a,b,n)` implements this approach. The size of the FFT must be equal to the total number of samples in the input sequence, so this approach is not advisable for long input sequences.

10.7.3   Generation of Random Processes

Finally, we are interested in obtaining discrete-time and continuous-time sample functions of the inputs and outputs of systems. Previous chapters provide us with several tools for the generation of random signals that can act as inputs to the systems of interest.

Section 5.10 provides the method for generating independent pairs of Gaussian random variables. This method forms the basis for the generation of iid Gaussian sequences and is implemented in `normal_rnd=(M,V,Sz)`. The generation of sequences of WSS but correlated sequences of Gaussian random variables requires more work. One approach is to use the matrix approaches developed in Section 6.6 to generate individual vectors with a specified covariance matrix. To generate a vector $\mathbf{Y}$ of n outcomes with covariance $\mathbf{K_Y}$, we perform the following factorization:

$$\mathbf{K_Y} = \mathbf{A}^\mathrm{T}\,\mathbf{A}\,\mathbf{P}\,\boldsymbol{\lambda}\,\mathbf{P}\mathbf{T},$$

and we generate the vector

$$\mathbf{Y} = \mathbf{A}^\mathrm{T}\,\mathbf{X}$$

where $\mathbf{X}$ is vector of iid zero-mean, unit-variance Gaussian random variables. The Octave function `svd(B)` performs a singular value decomposition of the matrix B, see [Long]. When $B = \mathbf{K_Y}$ is a covariance matrix, `svd` returns the diagonal matrix $\mathbf{D}$ of eigenvalues of $\mathbf{K_Y}$ as well as the matrices $\mathbf{U} = \mathbf{P}$ and $\mathbf{V} = \mathbf{P}^\mathrm{T}$.

Example 10.28   Generation of Correlated Gaussian Random Variables

Generate 256 samples of the autoregressive process in Example 10.14 with $\alpha = -0.5, \sigma_X = 1$.

The autocorrelation of the process is given by $R_X(k) = (-1/2)^{|k|}$. We generate a vector r of the first 256 lags of $R_X(k)$ and use the function `toeplitz(r)` to generate the covariance matrix. We then call the `svd` to obtain A. Finally we produce the output vector $\mathbf{Y} = \mathbf{A}^\mathrm{T}\,\mathbf{X}$.

```
> n=[0:255]
> r=(−0.5).^n;
> K=toeplitz(r);
> [U,D,V]=svd(K);
> X=normal_rnd(0,1,1,256);
> y=V*(D^0.5)*transpose(X);
> plot(y)
```

Figure 10.21(a) shows a plot of $\mathbf{Y}$. To check that the sequence has the desired autocovariance we use the function `autocov(X,H)` which estimates the autocovariance function of the sequence X for the first H lag values. Figure 10.21(b) shows that the sample correlation coefficient that is obtained by dividing the autocovariance by the sample variance. The plot shows the alternating covariance values and the expected peak values of -0.5 and 0.25 to the first two lags.
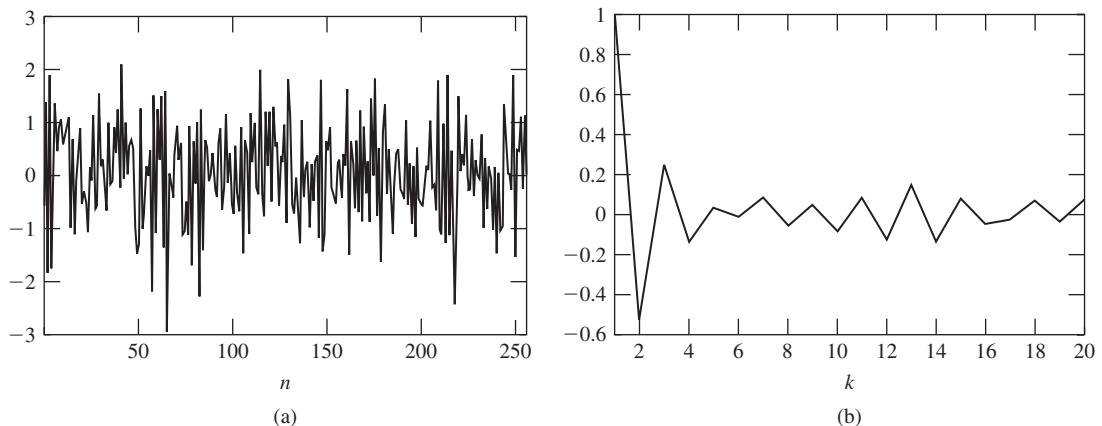
FIGURE 10.21
(a) Correlated Gaussian noise (b) Sample autocovariance.

An alternative approach to generating a correlated sequence of random variables with a specified covariance function is to input an uncorrelated sequence into a linear filter with a specific $H(f)$. Equation (10.46) allows us to determine the power spectral density of the output sequence. This approach can be implemented using convolution and is applicable to extremely long signal sequences. A large choice of possible filter functions is available for both continuous-time and discrete-time systems. For example, the ARMA model in Example 10.15 is capable of implementing a broad range of transfer functions. Indeed the entire discussion in Section 10.4 was focused on obtaining the transfer function of optimal linear systems in various scenarios.

Example 10.29   Generation of White Gaussian Noise

Find a method for generating white Gaussian noise for a simulation of a continuous-time communications system.

The generation of discrete-time white Gaussian noise is trivial and involves the generation of a sequence of iid Gaussian random variables. The generation of continuous-time white Gaussian noise is not so simple. Recall from Example 10.3 that true white noise has infinite bandwidth and hence infinite power and so is impossible to realize. Real systems however are bandlimited, and hence we always end up dealing with bandlimited white noise. If the system of interest is bandlimited to W Hertz, then we need to model white noise limited to W Hz. In Example 10.3 we found this type of noise has autocorrelation:

$$R_X(\tau) = \frac{N_0 \sin(2\pi W \tau)}{2\pi \tau}.$$

The sampling theorem discussed in Section 10.3 allows us to represent bandlimited white Gaussian noise as follows:

$$\hat{X}(t) = \sum_{n=-\infty}^{\infty} X(nT) p(t - nT) \quad \text{where} \quad p(t) = \frac{\sin(\pi t/T)}{\pi t/T},$$

where $1/T = 2W$. The coefficients $X(nT)$ have autocorrelation $R_X(nT)$ which is given by:

$$R_X(nT) = \frac{N_0 \sin(2\pi WnT)}{2\pi nT} = \frac{N_0 \sin(2\pi Wn/2W)}{2\pi n/2W}$$

$$= \frac{N_0 W \sin(\pi n)}{\pi n} = \begin{cases} N_0 W & \text{for} \quad n = 0 \\ 0 & \text{for} \quad n \neq 0. \end{cases}$$

We thus conclude that $X(nT)$ is an iid sequence of Gaussian random variables with variance $N_0 W$. Therefore we can simulate sampled bandlimited white Gaussian noise by generating a sequence $X(nT)$. We can perform any processing required in the discrete-time domain, and we can then apply the result to an interpolator to recover the continuous-time output.

SUMMARY

- The power spectral density of a WSS process is the Fourier transform of its autocorrelation function. The power spectral density of a real-valued random process is a real-valued, nonnegative, even function of frequency.
- The output of a linear, time-invariant system is a WSS random process if its input is a WSS random process that is applied an infinite time in the past.
- The output of a linear, time-invariant system is a Gaussian WSS random process if its input is a Gaussian WSS random process.
- Wide-sense stationary random processes with arbitrary rational power spectral density can be generated by filtering white noise.
- The sampling theorem allows the representation of bandlimited continuous-time processes by the sequence of periodic samples of the process.
- The orthogonality condition can be used to obtain equations for linear systems that minimize mean square error. These systems arise in filtering, smoothing, and prediction problems. Matrix numerical methods are used to find the optimum linear systems.
- The Kalman filter can be used to estimate signals with a structure that keeps the dimensionality of the algorithm fixed even as the size of the observation set increases.
- The variance of the periodogram estimate for the power spectral density does not approach zero as the number of samples is increased. An average of several independent periodograms is required to obtain an estimate whose variance does approach zero as the number of samples is increased.
- The FFT, convolution, and matrix techniques are basic tools for analyzing, simulating, and implementing processing of random signals.

CHECKLIST OF IMPORTANT TERMS

Amplitude modulation	Cross-power spectral density
ARMA process	Einstein-Wiener-Khinchin theorem
Autoregressive process	Filtering
Bandpass signal	Impulse response
Causal system	Innovations

Kalman filter
Linear system
Long-range dependence
Moving average process
Nyquist sampling rate
Optimum filter
Orthogonality condition
Periodogram
Power spectral density
Prediction
Quadrature amplitude modulation

Sampling theorem
Smoothed periodogram
Smoothing
System
Time-invariant system
Transfer function
Unit-sample response
White noise
Wiener filter
Wiener-Hopf equations
Yule-Walker equations

ANNOTATED REFERENCES

References [1] through [6] contain good discussions of the notion of power spectral density and of the response of linear systems to random inputs. References [6] and [7] give accessible introductions to the spectral factorization problem. References [7] through [9] discuss linear filtering and power spectrum estimation in the context of digital signal processing. Reference [10] discusses the basic theory underlying power spectrum estimation.

1. A. Papoulis and S. Pillai, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 2002.
2. H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, 3d ed., Prentice Hall, Upper Saddle River, N.J., 2002.
3. R. M. Gray and L. D. Davisson, *Random Processes: A Mathematical Approach for Engineers*, Prentice Hall, Englewood Cliffs, N.J., 1986.
4. R. D. Yates and D. J. Goodman, *Probability and Stochastic Processes*, Wiley, New York, 2005.
5. J. A. Gubner, *Probability and Random Processes for Electrical and Computer Engineering*, Cambridge University Press, Cambridge, 2006.
6. G. R. Cooper and C. D. MacGillem, *Probabilistic Methods of Signal and System Analysis*, Holt, Rinehart & Winston, New York, 1986.
7. J. A. Cadzow, *Foundations of Digital Signal Processing and Data Analysis*, Macmillan, New York, 1987.
8. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, N.J., 1989.
9. M. Kunt, *Digital Signal Processing*, Artech House, Dedham, Mass., 1986.
10. G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*, Holden Day, San Francisco, 1968.
11. A. Einstein, "Method for the Determination of the Statistical Values of Observations Concerning Quantities Subject to Irregular Observations," reprinted in *IEEE ASSP Magazine,* October 1987, p. 6.
12. P. J. G. Long, "Introduction to Octave," University of Cambridge, September, 2005, available online.

PROBLEMS

Section 10.1: Power Spectral Density

10.1. Let $g(x)$ denote the triangular function shown in Fig. P10.1.

    (a) Find the power spectral density corresponding to $R_X(\tau) = g(\tau/T)$.

    (b) Find the autocorrelation corresponding to the power spectral density
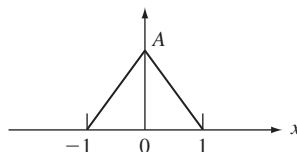$S_X(f) = g(f/W)$.

FIGURE P10.1

10.2. Let $p(x)$ be the rectangular function shown in Fig. P10.2. Is $R_X(\tau) = p(\tau/T)$ a valid autocorrelation function?
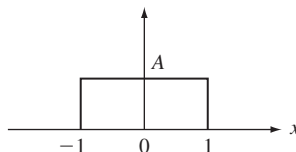

FIGURE P10.2

10.3. (a) Find the power spectral density $S_Y(f)$ of a random process with autocorrelation function $R_X(\tau)\cos(2\pi f_0\tau)$, where $R_X(\tau)$ is itself an autocorrelation function.

    (b) Plot $S_Y(f)$ if $R_X(\tau)$ is as in Problem 10.1a.

10.4. (a) Find the autocorrelation function corresponding to the power spectral density shown in Fig. P10.3.

    (b) Find the total average power.

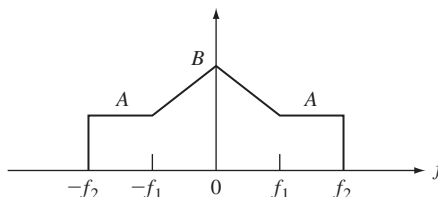   **(c)** Plot the power in the range $|f| > f_0$ as a function of $f_0 > 0$.

FIGURE P10.3

10.5. A random process $X(t)$ has autocorrelation given by $R_X(\tau) = \sigma_X^2 e^{-\tau^2/2\alpha^2}$, $\alpha > 0$.
   (a) Find the corresponding power spectral density.
   (b) Find the amount of power contained in the frequencies $|f| > k/2\pi\alpha$, where $k = 1, 2, 3$.

10.6. Let $Z(t) = X(t) + Y(t)$. Under what conditions does $S_Z(f) = S_X(f) + S_Y(f)$?

10.7. Show that
   (a) $R_{X,Y}(\tau) = R_{Y,X}(-\tau)$.
   (b) $S_{X,Y}(f) = S_{Y,X}^*(f)$.

10.8. Let $Y(t) = X(t) - X(t - d)$.
   (a) Find $R_{X,Y}(\tau)$ and $S_{X,Y}(f)$.
   (b) Find $R_Y(\tau)$ and $S_Y(f)$.

10.9. Do Problem 10.8 if $X(t)$ has the triangular autocorrelation function $g(\tau/T)$ in Problem 10.1 and Fig. P 10.1.

10.10. Let $X(t)$ and $Y(t)$ be independent wide-sense stationary random processes, and define $Z(t) = X(t)Y(t)$.
   (a) Show that $Z(t)$ is wide-sense stationary.
   (b) Find $R_Z(\tau)$ and $S_Z(f)$.

10.11. In Problem 10.10, let $X(t) = a\cos(2\pi f_0 t + \Theta)$ where Θ is a uniform random variable in $(0, 2\pi)$. Find $R_Z(\tau)$ and $S_Z(f)$.

10.12. Let $R_X(k) = 4\alpha^{|k|}$, $|\alpha| < 1$.
   (a) Find $S_X(f)$.
   (b) Plot $S_X(f)$ for $\alpha = 0.25$ and $\alpha = 0.75$, and comment on the effect of the value of α.

10.13. Let $R_X(k) = 4(\alpha)^{|k|} + 16(\beta)^{|k|}$, $\alpha < 1$, $\beta < 1$.
   (a) Find $S_X(f)$.
   (b) Plot $S_X(f)$ for $\alpha = \beta = 0.5$ and $\alpha = 0.75 = 3\beta$ and comment on the effect of value of α/β.

10.14. Let $R_X(k) = 9(1 - |k|/N)$, for $|k| < N$ and 0 elsewhere. Find and plot $S_X(f)$.

10.15. Let $X_n = \cos(2\pi f_0 n + \Theta)$, where Θ is a uniformly distributed random variable in the interval $(0, 2\pi)$. Find and plot $S_X(f)$ for $f_0 = 0.5, 1, 1.75, \pi$.

10.16. Let $D_n = X_n - X_{n-d}$, where d is an integer constant and X_n is a zero-mean, WSS random process.
   (a) Find $R_D(k)$ and $S_D(f)$ in terms of $R_X(k)$ and $S_X(f)$. What is the impact of d?
   (b) Find $E[D_n^2]$.

10.17. Find $R_D(k)$ and $S_D(f)$ in Problem 10.16 if X_n is the moving average process of Example 10.7 with $\alpha = 1$.

10.18. Let X_n be a zero-mean, bandlimited white noise random process with $S_X(f) = 1$ for $|f| < f_c$ and 0 elsewhere, where $f_c < 1/2$.
   (a) Show that $R_X(k) = \sin(2\pi f_c k)/(\pi k)$.
   (b) Find $R_X(k)$ when $f_c = 1/4$.

10.19. Let W_n be a zero-mean white noise sequence, and let X_n be independent of W_n.
   (a) Show that $Y_n = W_n X_n$ is a white sequence, and find σ_Y^2.
   (b) Suppose X_n is a Gaussian random process with autocorrelation $R_X(k) = (1/2)^{|k|}$. Specify the joint pmf's for Y_n.

10.20. Evaluate the periodogram estimate for the random process $X(t) = a\cos(2\pi f_0 t + \Theta)$, where Θ is a uniformly distributed random variable in the interval $(0, 2\pi)$. What happens as $T \to \infty$?

10.21. **(a)** Show how to use the FFT to calculate the periodogram estimate in Eq. (10.32).

   (b) Generate four realizations of an iid zero-mean unit-variance Gaussian sequence of length 128. Calculate the periodogram.

   (c) Calculate 50 periodograms as in part b and show the average of the periodograms after every 10 additional realizations.

Section 10.2: Response of Linear Systems to Random Signals

10.22. Let $X(t)$ be a differentiable WSS random process, and define

$$Y(t) = \frac{d}{dt}X(t).$$

Find an expression for $S_Y(f)$ and $R_Y(\tau)$. *Hint:* For this system, $H(f) = j2\pi f$.

10.23. Let $Y(t)$ be the derivative of $X(t)$, a bandlimited white noise process as in Example 10.3.

   (a) Find $S_Y(f)$ and $R_Y(\tau)$.

   (b) What is the average power of the output?

10.24. Repeat Problem 10.23 if $X(t)$ has $S_X(f) = \beta^2 e^{-\pi f^2}$.

10.25. Let $Y(t)$ be a short-term integration of $X(t)$:

$$Y(t) = \frac{1}{T}\int_{t-T}^{t} X(t')\,dt'.$$

   (a) Find the impulse response $h(t)$ and the transfer function $H(f)$.

   (b) Find $S_Y(f)$ in terms of $S_X(f)$.

10.26. In Problem 10.25, let $R_X(\tau) = (1 - |\tau|/T)$ for $|\tau| < T$ and zero elsewhere.

   (a) Find $S_Y(f)$.

   (b) Find $R_Y(\tau)$.

   (c) Find $E[Y^2(t)]$.

10.27. The input into a filter is zero-mean white noise with noise power density $N_0/2$. The filter has transfer function

$$H(f) = \frac{1}{1 + j2\pi f}.$$

   (a) Find $S_{Y,X}(f)$ and $R_{Y,X}(\tau)$.

   (b) Find $S_Y(f)$ and $R_Y(\tau)$.

   (c) What is the average power of the output?

10.28. A bandlimited white noise process $X(t)$ is input into a filter with transfer function $H(f) = 1 + j2\pi f$.

   (a) Find $S_{Y,X}(f)$ and $R_{Y,X}(\tau)$ in terms of $R_X(\tau)$ and $S_X(f)$.

   (b) Find $S_Y(f)$ and $R_Y(\tau)$ in terms of $R_X(\tau)$ and $S_X(f)$.

   (c) What is the average power of the output?

10.29. **(a)** A WSS process $X(t)$ is applied to a linear system at $t = 0$. Find the mean and autocorrelation function of the output process. Show that the output process becomes WSS as $t \to \infty$.

10.30. Let $Y(t)$ be the output of a linear system with impulse response $h(t)$ and input $X(t)$. Find $R_{Y,X}(\tau)$ when the input is white noise. Explain how this result can be used to estimate the impulse response of a linear system.

10.31. **(a)** A WSS Gaussian random process $X(t)$ is applied to two linear systems as shown in Fig. P10.4. Find an expression for the joint pdf of $Y(t_1)$ and $W(t_2)$.

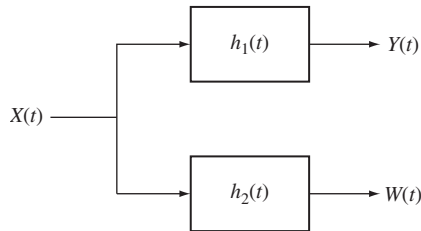   **(b)** Evaluate part a if $X(t)$ is white Gaussian noise.

FIGURE P10.4

10.32. Repeat Problem 10.31b if $h_1(t)$ and $h_2(t)$ are ideal bandpass filters as in Example 10.11. Show that $Y(t)$ and $W(t)$ are independent random processes if the filters have nonoverlapping bands.

10.33. Let $Y(t) = h(t) * X(t)$ and $Z(t) = X(t) - Y(t)$ as shown in Fig. P10.5.

   (a) Find $S_Z(f)$ in terms of $S_X(f)$.
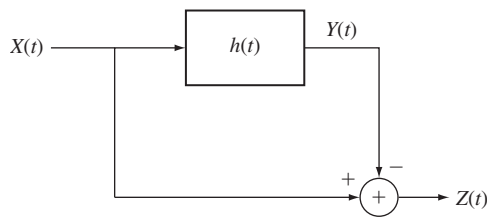
   (b) Find $E[Z^2(t)]$.

FIGURE P10.5

10.34. Let $Y(t)$ be the output of a linear system with impulse response $h(t)$ and input $X(t) + N(t)$. Let $Z(t) = X(t) - Y(t)$.

   (a) Find $R_{X,Y}(\tau)$ and $R_Z(\tau)$.

   (b) Find $S_Z(f)$.

   (c) Find $S_Z(f)$ if $X(t)$ and $N(t)$ are independent random processes.

10.35. A random telegraph signal is passed through an ideal lowpass filter with cutoff frequency W. Find the power spectral density of the difference between the input and output of the filter. Find the average power of the difference signal.

10.36. Let $Y(t) = a\cos(2\pi f_c t + \Theta) + N(t)$ be applied to an ideal bandpass filter that passes the frequencies $|f - f_c| < W/2$. Assume that Θ is uniformly distributed in $(0, 2\pi)$. Find the ratio of signal power to noise power at the output of the filter.

10.37. Let $Y_n = (X_{n+1} + X_n + X_{n-1})/3$ be a "smoothed" version of X_n. Find $R_Y(k)$, $S_Y(f)$, and $E[Y^2{}_n]$.

10.38. Suppose X_n is a white Gaussian noise process in Problem 10.37. Find the joint pmf for (Y_n, Y_{n+1}, Y_{n+2}).

10.39. Let $Y_n = X_n + \beta X_{n-1}$, where X_n is a zero-mean, first-order autoregressive process with autocorrelation $R_X(k) = \sigma^2 \alpha^k$, $|\alpha| < 1$.

   (a) Find $R_{Y,X}(k)$ and $S_{Y,X}(f)$.

   (b) Find $S_Y(f)$, $R_Y(k)$, and $E[Y_n^2]$.

   (c) For what value of β is Y_n a white noise process?

10.40. A zero-mean white noise sequence is input into a cascade of two systems (see Fig. P10.6). System 1 has impulse response $h_n = (1/2)^n u(n)$ and system 2 has impulse response $g_n = (1/4)^n u(n)$ where $u(n) = 1$ for $n \geq 0$ and 0 elsewhere.

   (a) Find $S_Y(f)$ and $S_Z(f)$.

   (b) Find $R_{W,Y}(k)$ and $R_{W,Z}(k)$; find $S_{W,Y}(f)$ and $S_{W,Z}(f)$. *Hint:* Use a partial fraction expansion of $S_{W,Z}(f)$ prior to finding $R_{W,Z}(k)$.
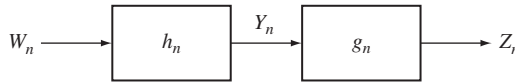
   (c) Find $E[Z_n^2]$.

FIGURE P10.6

10.41. A moving average process X_n is produced as follows:

$$X_n = W_n + \alpha_1 W_{n-1} + \cdots + \alpha_p W_{n-p},$$

where W_n is a zero-mean white noise process.

   (a) Show that $R_X(k) = 0$ for $|k| > p$.

   (b) Find $R_X(k)$ by computing $E[X_{n+k} X_n]$, then find $S_X(f) = \mathcal{F}\{R_X(k)\}$.

   (c) Find the impulse response h_n of the linear system that defines the moving average process. Find the corresponding transfer function $H(f)$, and then $S_X(f)$. Compare your answer to part b.

10.42. Consider the second-order autoregressive process defined by

$$Y_n = \frac{3}{4}Y_{n-1} - \frac{1}{8}Y_{n-2} + W_n,$$

where the input W_n is a zero-mean white noise process.

   (a) Verify that the unit-sample response is $h_n = 2(1/2)^n - (1/4)^n$ for $n \geq 0$, and 0 otherwise.

   (b) Find the transfer function.

   (c) Find $S_Y(f)$ and $R_Y(k) = \mathcal{F}^{-1}\{S_Y(f)\}$.

10.43. Suppose the autoregressive process defined in Problem 10.42 is the input to the following moving average system:

$$Z_n = Y_n - 1/4Y_{n-1}.$$

    (a) Find $S_Z(f)$ and $R_Z(k)$.

    (b) Explain why Z_n is a first-order autoregressive process.

    (c) Find a moving average system that will produce a white noise sequence when Z_n is the input.

10.44. An autoregressive process Y_n is produced as follows:

$$Y_n = \alpha_1 Y_{n-1} + \cdots + \alpha_q Y_{n-q} + W_n,$$

where W_n is a zero-mean white noise process.

    (a) Show that the autocorrelation of Y_n satisfies the following set of equations:

$$R_Y(0) = \sum_{i=1}^{q} \alpha_i R_Y(i) + R_W(0)$$

$$R_Y(k) = \sum_{i=1}^{q} \alpha_i R_Y(k - i).$$

    (b) Use these recursive equations to compute the autocorrelation of the process in Example 10.22.

Section 10.3: Bandlimited Random Processes

10.45. **(a)** Show that the signal $x(t)$ is recovered in Figure 10.10(b) as long as the sampling rate is above the Nyquist rate.

    (b) Suppose that a deterministic signal is sampled at a rate below the Nyquist rate. Use Fig. 10.10(b) to show that the recovered signal contains additional signal components from the adjacent bands. The error introduced by these components is called aliasing.

    (c) Find an expression for the power spectral density of the sampled bandlimited random process $X(t)$.

    (d) Find an expression for the power in the aliasing error components.

    (e) Evaluate the power in the error signal in part c if $S_X(f)$ is as in Problem 10.1b.

10.46. An ideal discrete-time lowpass filter has transfer function:

$$H(f) = \begin{cases} 1 & \text{for} & |f| < f_c < 1/2 \\ 0 & \text{for} & f_c < |f| < 1/2. \end{cases}$$

    (a) Show that $H(f)$ has impulse response $h_n = \sin(2\pi f_c n)/\pi n$.

    (b) Find the power spectral density of $Y(kT)$ that results when the signal in Problem 10.1b is sampled at the Nyquist rate and processed by the filter in part a.

    (c) Let $Y(t)$ be the continuous-time signal that results when the output of the filter in part b is fed to an interpolator operating at the Nyquist rate. Find $S_Y(f)$.

10.47. In order to design a differentiator for bandlimited processes, the filter in Fig. 10.10(c) is designed to have transfer function:

$$H(f) = j2\pi f/T \text{ for } |f| < 1/2.$$

(a) Show that the corresponding impulse response is:

$$h_0 = 0, h_n = \frac{\pi n \cos \pi n - \sin \pi n}{\pi n^2 T} = \frac{(-1)^n}{nT} \quad n \neq 0$$

(b) Suppose that $X(t) = a \cos(2\pi f_0 t + \Theta)$ is sampled at a rate $1/T = 4f_0$ and then input into the above digital filter. Find the output $Y(t)$ of the interpolator.

10.48. Complete the proof of the sampling theorem by showing that the mean square error is zero. *Hint:* First show that $E[(X(t) - (\hat{X}((t)) X(kT)] = 0$, all k.

10.49. Plot the power spectral density of the amplitude modulated signal $Y(t)$ in Example 10.18, assuming $f_c > W$; $f_c < W$. Assume that $A(t)$ is the signal in Problem 10.1b.

10.50. Suppose that a random telegraph signal with transition rate α is the input signal in an amplitude modulation system. Plot the power spectral density of the modulated signal assuming $f_c = \alpha/\pi$ and $f_c = 10\alpha/\pi$.

10.51. Let the input to an amplitude modulation system be $2 \cos(2\pi f_1 + \Phi)$, where Φ is uniformly distributed in $(-\pi, \pi)$. Find the power spectral density of the modulated signal assuming $f_c > f_1$.

10.52. Find the signal-to-noise ratio in the recovered signal in Example 10.18 if $S_N(f) = \alpha f^2$ for $|f \pm f_c| < W$ and zero elsewhere.

10.53. The input signals to a QAM system are independent random processes with power spectral densities shown in Fig. P10.7. Sketch the power spectral density of the QAM signal.

FIGURE P10.7
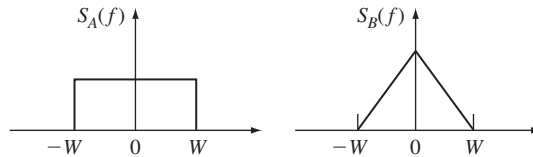
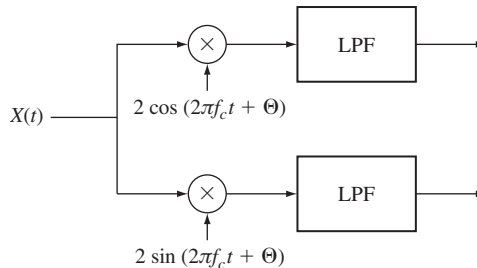**10.54.** Under what conditions does the receiver shown in Fig. P10.8 recover the input signals to a QAM signal?

FIGURE P10.8

10.55. Show that Eq. (10.67b) implies that $S_{B,A}(f)$ is a purely imaginary, odd function of f.

Section 10.4: Optimum Linear Systems

10.56. Let $X_\alpha = Z_\alpha + N_\alpha$ as in Example 10.22, where Z_α is a first-order process with $R_Z(k) = 4(3/4)^{|k|}$ and N_α is white noise with $\sigma_N^2 = 1$.
  (a) Find the optimum $p = 1$ filter for estimating Z_α.
  (b) Find the mean square error of the resulting filter.

10.57. Let $X_\alpha = Z_\alpha + N_\alpha$ as in Example 10.21, where Z_α has $R_Z(k) = \sigma_Z^2(r_1)^{|k|}$ and N_α has $R_N(k) = \sigma_N^2 r_2^{|k|}$, where r_1 and r_2 are less than one in magnitude.
  (a) Find the equation for the optimum filter for estimating Z_α.
  (b) Write the matrix equation for the filter coefficients.
  (c) Solve the $p = 2$ case, if $\sigma_Z^2 = 9$, $r_1 = 2/3$, $\sigma_N^2 = 1$, and $r_2 = 1/3$.
  (d) Find the mean square error for the optimum filter in part c.
  (e) Use the matrix function of Octave to solve parts c and d for $p = 3, 4, 5$.

10.58. Let $X_\alpha = Z_\alpha + N_\alpha$ as in Example 10.21, where Z_α is the first-order moving average process of Example 10.7, and N_α is white noise.
  (a) Find the equation for the optimum filter for estimating Z_α.
  (b) For the $p = 1$ and $p = 2$ cases, write and solve the matrix equation for the filter coefficients.
  (c) Find the mean square error for the optimum filter in part b.

10.59. Let $X_\alpha = Z_\alpha + N_\alpha$ as in Example 10.19, and suppose that an estimator for Z_α uses observations from the following time instants: $I = \{n - p, \ldots, n, \ldots, n + p\}$.
  (a) Solve the $p = 1$ case if Z_α and N_α are as in Problem 10.56.
  (b) Find the mean square error in part a.
  (c) Find the equation for the optimum filter.
  (d) Write the matrix equation for the $2p + 1$ filter coefficients.
  (e) Use the matrix function of Octave to solve parts a and b for $p = 2, 3$.

10.60. Consider the predictor in Eq. (10.86b).
  (a) Find the optimum predictor coefficients in the $p = 2$ case when $R_Z(k) = 9(1/3)^{|k|}$.
  (b) Find the mean square error in part a.
  (c) Use the matrix function of Octave to solve parts a and b for $p = 3, 4, 5$.

10.61. Let $X(t)$ be a WSS, continuous-time process.
  (a) Use the orthogonality principle to find the best estimator for $X(t)$ of the form

$$\hat{X}(t) = aX(t_1) + bX(t_2),$$

  where t_1 and t_2 are given time instants.
  (b) Find the mean square error of the optimum estimator.
  (c) Check your work by evaluating the answer in part b for $t = t_1$ and $t = t_2$. Is the answer what you would expect?

10.62. Find the optimum filter and its mean square error in Problem 10.61 if $t_1 = t - d$ and $t_2 = t + d$.

10.63. Find the optimum filter and its mean square error in Problem 10.61 if $t_1 = t - d$ and $t_2 = t - 2d$, and $R_X(\tau) = e^{-\alpha|\tau|}$ Compare the performance of this filter to the performance of the optimum filter of the form $\hat{X}(t) = aX(t - d)$.

10.64. Modify the system in Problem 10.33 to obtain a model for the estimation error in the optimum infinite-smoothing filter in Example 10.24. Use the model to find an expression for the power spectral density of the error $e(t) = Z(t) - Y(t)$, and then show that the mean square error is given by:

$$E[e^2(t)] = \int_{-\infty}^{\infty} \frac{S_Z(f)S_N(f)}{S_Z(f) + S_N(f)} df.$$

Hint: $E[e^2(t)] = R_e(0)$.

10.65. Solve the infinite-smoothing problem in Example 10.24 if $Z(t)$ is the random telegraph signal with $\alpha = 1/2$ and $N(t)$ is white noise. What is the resulting mean square error?

10.66. Solve the infinite-smoothing problem in Example 10.24 if $Z(t)$ is bandlimited white noise of density $N_1/2$ and $N(t)$ is (infinite-bandwidth) white noise of noise density $N_0/2$. What is the resulting mean square error?

10.67. Solve the infinite-smoothing problem in Example 10.24 if $Z(t)$ and $N(t)$ are as given in Example 10.25. Find the resulting mean square error.

10.68. Let $X_n = Z_n + N_n$, where Z_n and N_n are independent, zero-mean random processes.

   (a) Find the smoothing filter given by Eq. (10.89) when Z_n is a first-order autoregressive process with $\sigma_X^2 = 9$ and $\alpha = 1/2$ and N_n is white noise with $\sigma_N^2 = 4$.

   (b) Use the approach in Problem 10.64 to find the power spectral density of the error $S_e(f)$.

   (c) Find $R_e(k)$ as follows: Let $Z = e^{j2\pi f}$, factor the denominator $S_e(f)$, and take the inverse transform to show that:

$$R_e(k) = \frac{\sigma_X^2 z_1}{\alpha(1 - z_1^2)} z_1^{|k|} \quad \text{where} \quad 0 < z_1 < 1.$$

   (d) Find an expression for the resulting mean square error.

10.69. Find the Wiener filter in Example 10.25 if $N(t)$ is white noise of noise density $N_0/2 = 1/3$ and $Z(t)$ has power spectral density

$$S_z(f) = \frac{4}{4 + 4\pi^2 f^2}.$$

10.70. Find the mean square error for the Wiener filter found in Example 10.25. Compare this with the mean square error of the infinite-smoothing filter found in Problem 10.67.

10.71. Suppose we wish to estimate (predict) $X(t + d)$ by

$$\hat{X}(t + d) = \int_0^{\infty} h(\tau)X(t - \tau)\, d\tau.$$

   (a) Show that the optimum filter must satisfy

$$R_X(\tau + d) = \int_0^{\infty} h(x)R_X(\tau - x)\, dx \qquad \tau \ge 0.$$

   (b) Use the Wiener-Hopf method to find the optimum filter when $R_X(\tau) = e^{-2|\tau|}$.

10.72. Let $X_n = Z_n + N_n$, where Z_n and N_n are independent random processes, N_n is a white noise process with $\sigma_N^2 = 1$, and Z_n is a first-order autoregressive process with $R_Z(k) = 4(1/2)^{|k|}$. We are interested in the optimum filter for estimating Z_n from $X_n, X_{n-1}, \ldots$.

(a) Find $S_X(f)$ and express it in the form:

$$S_X(f) = \frac{\frac{1}{2z_1}\left(1 - \frac{1}{z_1}e^{-j2\pi f}\right)\left(1 - z_1 e^{j2\pi f}\right)}{\left(1 - \frac{1}{2}e^{-j2\pi f}\right)\left(1 - \frac{1}{2}e^{j2\pi f}\right)}.$$

(b) Find the whitening causal filter.

(c) Find the optimal causal filter.

Section 10.5: The Kalman Filter

10.73. If W_n and N_n are Gaussian random processes in Eq. (10.102), are Z_n and X_n Markov processes?

10.74. Derive Eq. (10.120) for the mean square prediction error.

10.75. Repeat Example 10.26 with $a = 0.5$ and $a = 2$.

10.76. Find the Kalman algorithm for the case where the observations are given by

$$X_n = b_n Z_n + N_n$$

where b_n is a sequence of known constants.

*Section 10.6: Estimating the Power Spectral Density

10.77. Verify Eqs. (10.125) and (10.126) for the periodogram and the autocorrelation function estimate.

10.78. Generate a sequence X_n of iid random variables that are uniformly distributed in $(0, 1)$.

(a) Compute several 128-point periodograms and verify the random behavior of the periodogram as a function of f. Does the periodogram vary about the true power spectral density?

(b) Compute the smoothed periodogram based on 10, 20, and 50 independent periodograms. Compare the smoothed periodograms to the true power spectral density.

10.79. Repeat Problem 10.78 with X_n a first-order autoregressive process with autocorrelation function: $R_X(k) = (.9)^{|k|}$; $R_X(k) = (1/2)^{|k|}$; $R_X(k) = (.1)^{|k|}$.

10.80. Consider the following estimator for the autocorrelation function

$$\hat{r}'_k(m) = \frac{1}{k - |m|} \sum_{n=0}^{k-|m|-1} X_n X_{n+m}.$$

Show that if we estimate the power spectrum of X_n by the Fourier transform of $\hat{r}'_k(m)$, the resulting estimator has mean

$$E[\tilde{p}_k(f)] = \sum_{m'=-(k-1)}^{k-1} R_X(m')e^{-j2\pi f m'}.$$

Why is the estimator biased?

Section 10.7: Numerical Techniques for Processing Random Signals

10.81. Let $X(t)$ have power spectral density given by $S_X(f) = \beta^2 e^{-f^2/2W_0^2}/\sqrt{2\pi}$.

(a) Before performing an FFT of $S_X(f)$, you are asked to calculate the power in the aliasing error if the signal is treated as if it were bandlimited with bandwidth kW_0.

What value of W should be used for the FFT if the power in the aliasing error is to be less than 1% of the total power? Assume $W_0 = 1000$ and $\beta = 1$.

(b) Suppose you are to perform $N = 2M$ point FFT of $S_X(f)$. Explore how W, T, and t_0 vary as a function of f_0. Discuss what leeway is afforded by increasing N.

(c) For the value of W in part a, identify the values of the parameters f_0, T, and t_0 for $N = 128, 256, 512, 1024$.

(d) Find the autocorrelation $\{R_X(kt_0)\}$ by applying the FFT to $S_X(f)$. Try the options identified in part c and comment on the accuracy of the results by comparing them to the exact value of $R_X(\tau)$.

10.82. Use the FFT to calculate and plot $S_X(f)$ for the following discrete-time processes:

(a) $R_X(k) = 4\alpha^{|k|}$, for $\alpha = 0.25$ and $\alpha = 0.75$.

(b) $R_X(k) = 4(1/2)^{|k|} + 16(1/4)^{|k|}$.

(c) $X_n = \cos(2\pi f_0 n + \Theta)$, where Θ is a uniformly distributed in $(0, 2\pi]$ and $f_0 = 1000$.

10.83. Use the FFT to calculate and plot $R_X(k)$ for the following discrete-time processes:

(a) $S_X(f) = 1$ for $|f| < f_c$ and 0 elsewhere, where $f_c = 1/8, 1/4, 3/8$.

(b) $S_X(f) = 1/2 + 1/2 \cos 2\pi f$ for $|f| < 1/2$.

10.84. Use the FFT to find the output power spectral density in the following systems:

(a) Input X_n with $R_X(k) = 4\alpha^{|k|}$, for $\alpha = 0.25$, $H(f) = 1$ for $|f| < 1/4$.

(b) Input $X_n = \cos(2\pi f_0 n + \Theta)$, where Θ is a uniformly distributed random variable and $H(f) = j2\pi f$ for $|f| < 1/2$.

(c) Input X_n with $R_X(k)$ as in Problem 10.14 with $N = 3$ and $H(f) = 1$ for $|f| < 1/2$.

10.85. (a) Show that

$$R_X(\tau) = 2\text{Re}\left\{ \int_0^\infty S_X(f)e^{-j2\pi f\tau}\, df \right\}.$$

(b) Use approximations to express the above as a DFT relating N points in the time domain to N points in the frequency domain.

(c) Suppose we meet the $t_0 f_0 = 1/N$ requirement by letting $t_0 = f_0 = 1/\sqrt{N}$. Compare this to the approach leading to Eq. (10.142).

10.86. (a) Generate a sequence of 1024 zero-mean unit-variance Gaussian random variables and pass it through a system with impulse response $h_n = e^{-2n}$ for $n \geq 0$.

(b) Estimate the autocovariance of the output process of the digital filter and compare it to the theoretical autocovariance.

(c) What is the pdf of the continuous-time process that results if the output of the digital filter is fed into an interpolator?

10.87. (a) Use the covariance matrix factorization approach to generate a sequence of 1024 Gaussian samples with autocovariance $h(t) = e^{-2|\tau|}$.

(b) Estimate the autocovariance of the observed sequence and compare to the theoretical result.

Problems Requiring Cumulative Knowledge

10.88. Does the pulse amplitude modulation signal in Example 9.38 have a power spectral density? Explain why or why not. If the answer is yes, find the power spectral density.

10.89. Compare the operation and performance of the Wiener and Kalman filters for the signals discussed in Example 10.26.

10.90. (a) Find the power spectral density of the ARMA process in Example 10.15 by finding the transfer function of the associated linear system.

   (b) For the ARMA process find the cross-power spectral density from $E[Y_nX_m]$, and then the power spectral density from $E[Y_nY_m]$.

10.91. Let $X_1(t)$ and $X_2(t)$ be jointly WSS and jointly Gaussian random processes that are input into two linear time-invariant systems as shown below:

$$X_1(t) \rightarrow \boxed{h_1(t)} \rightarrow Y_1(t)$$

$$X_2(t) \rightarrow \boxed{h_2(t)} \rightarrow Y_2(t)$$

   (a) Find the cross-correlation function of $Y_1(t)$ and $Y_2(t)$. Find the corresponding cross-power spectral density.

   (b) Show that $Y_1(t)$ and $Y_2(t)$ are jointly WSS and jointly Gaussian random processes.

   (c) Suppose that the transfer functions of the above systems are nonoverlapping, that is, $|H_1(f)||H_2(f)| = 0$. Show that $Y_1(t)$ and $Y_2(t)$ are independent random processes.

   (d) Now suppose that $X_1(t)$ and $X_2(t)$ are nonstationary jointly Gaussian random processes. Which of the above results still hold?

10.92. Consider the communication system in Example 9.38 where the transmitted signal $X(t)$ consists of a sequence of pulses that convey binary information. Suppose that the pulses $p(t)$ are given by the impulse response of the ideal lowpass filter in Figure 10.6. The signal that arrives at the receiver is $Y(t) = X(t) + N(t)$ which is to be sampled and processed digitally.

   (a) At what rate should $Y(t)$ be sampled?

   (b) How should the bit carried by each pulse be recovered based on the samples $Y(nT)$?

   (c) What is the probability of error in this system?