5 Joint Probability Distributions

CHAPTER OUTLINE

LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Use joint probability mass functions and joint probability density functions to calculate probabilities
2. Calculate marginal and conditional probability distributions from joint probability distributions
3. Use the multinomial distribution to determine probabilities
4. Interpret and calculate covariances and correlations between random variables
5. Understand properties of a bivariate normal distribution and be able to draw contour plots for the probability density function
6. Calculate means and variance for linear combinations of random variables and calculate probabilities for linear combinations of normally distributed random variables

CD MATERIAL

7. Determine the distribution of a function of one or more random variables
8. Calculate moment generating functions and use them to determine moments for random variables and use the uniqueness property to determine the distribution of a random variable
9. Provide bounds on probabilities for arbitrary distributions based on Chebyshev's inequality

---

Answers for most odd numbered exercises are at the end of the book. Answers to exercises whose numbers are surrounded by a box can be accessed in the e-Text by clicking on the box. Complete worked solutions to certain exercises are also available in the e-Text. These are indicated in the Answers to Selected Exercises section by a box around the exercise number. Exercises are also available for the text sections that appear on CD only. These exercises may be found within the e-Text immediately following the section they accompany.

In Chapters 3 and 4 we studied probability distributions for a single random variable. However, it is often useful to have more than one random variable defined in a random experiment. For example, in the classification of transmitted and received signals, each signal can be classified as high, medium, or low quality. We might define the random variable X to be the number of high-quality signals received and the random variable Y to be the number of low-quality signals received. In another example, the continuous random variable X can denote the length of one dimension of an injection-molded part, and the continuous random variable Y might denote the length of another dimension. We might be interested in probabilities that can be expressed in terms of both X and Y. For example, if the specifications for X and Y are (2.95 to 3.05) and (7.60 to 7.80) millimeters, respectively, we might be interested in the probability that a part satisfies both specifications; that is, $P(2.95 < X < 3.05 \text{ and } 7.60 < Y < 7.80)$.

In general, if X and Y are two random variables, the probability distribution that defines their simultaneous behavior is called a **joint probability distribution.** In this chapter, we investigate some important properties of these joint distributions.

5-1 TWO DISCRETE RANDOM VARIABLES

5-1.1 Joint Probability Distributions

For simplicity, we begin by considering random experiments in which only two random variables are studied. In later sections, we generalize the presentation to the joint probability distribution of more than two random variables.

EXAMPLE 5-1

In the development of a new receiver for the transmission of digital information, each received bit is rated as *acceptable, suspect,* or *unacceptable,* depending on the quality of the received signal, with probabilities 0.9, 0.08, and 0.02, respectively. Assume that the ratings of each bit are independent.

In the first four bits transmitted, let

X denote the number of acceptable bits

Y denote the number of suspect bits

Then, the distribution of X is binomial with $n = 4$ and $p = 0.9$, and the distribution of Y is binomial with $n = 4$ and $p = 0.08$. However, because only four bits are being rated, the possible values of X and Y are restricted to the points shown in the graph in Fig. 5-1. Although the possible values of X are 0, 1, 2, 3, or 4, if $y = 3$, $x = 0$ or 1. By specifying the probability of each of the points in Fig. 5-1, we specify the joint probability distribution of X and Y. Similarly to an individual random variable, we define the range of the random variables (X, Y) to be the set of points (x, y) in two-dimensional space for which the probability that $X = x$ and $Y = y$ is positive.

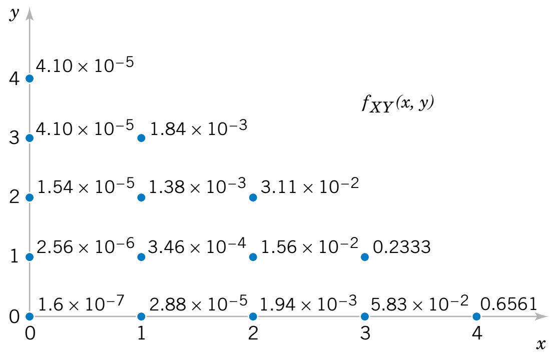

Figure 5-1 Joint probability distribution of X and Y in Example 5-1.

If X and Y are discrete random variables, the joint probability distribution of X and Y is a description of the set of points (x, y) in the range of (X, Y) along with the probability of each point. The joint probability distribution of two random variables is sometimes referred to as the **bivariate probability distribution** or **bivariate distribution** of the random variables. One way to describe the joint probability distribution of two discrete random variables is through a joint probability mass function. Also, $P(X = x \text{ and } Y = y)$ is usually written as $P(X = x, Y = y)$.

Definition

The **joint probability mass function** of the discrete random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies

(1) $f_{XY}(x, y) \geq 0$

(2) $\displaystyle\sum_x \sum_y f_{XY}(x, y) = 1$

(3) $f_{XY}(x, y) = P(X = x, Y = y)$ (5-1)

Subscripts are used to indicate the random variables in the bivariate probability distribution. Just as the probability mass function of a single random variable X is assumed to be zero at all values outside the range of X, so the joint probability mass function of X and Y is assumed to be zero at values for which a probability is not specified.

EXAMPLE 5-2

Probabilities for each point in Fig. 5-1 are determined as follows. For example, $P(X = 2, Y = 1)$ is the probability that exactly two acceptable bits and exactly one suspect bit are received among the four bits transferred. Let a, s, and u denote acceptable, suspect, and unacceptable bits, respectively. By the assumption of independence,

$$P(aasu) = 0.9(0.9)(0.08)(0.02) = 0.0013$$

The number of possible sequences consisting of two a's, one s, and one u is shown in the CD material for Chapter 2:

$$\frac{4!}{2!1!1!} = 12$$

Therefore,

$$P(aasu) = 12(0.0013) = 0.0156$$

and

$$f_{XY}(2, 1) = P(X = 2, Y = 1) = 0.0156$$

The probabilities for all points in Fig. 5-1 are shown next to the point and the figure describes the joint probability distribution of X and Y.

5-1.2 Marginal Probability Distributions

If more than one random variable is defined in a random experiment, it is important to distinguish between the joint probability distribution of X and Y and the probability distribution of each variable individually. The individual probability distribution of a random variable is referred to as its **marginal probability distribution.** In Example 5-1, we mentioned that the marginal probability distribution of X is binomial with $n = 4$ and $p = 0.9$ and the marginal probability distribution of Y is binomial with $n = 4$ and $p = 0.08$.

In general, the marginal probability distribution of X can be determined from the joint probability distribution of X and other random variables. For example, to determine $P(X = x)$, we sum $P(X = x, Y = y)$ over all points in the range of (X, Y) for which $X = x$. Subscripts on the probability mass functions distinguish between the random variables.

EXAMPLE 5-3

The joint probability distribution of X and Y in Fig. 5-1 can be used to find the marginal probability distribution of X. For example,

$$\begin{aligned} P(X = 3) &= P(X = 3, Y = 0) + P(X = 3, Y = 1) \\ &= 0.0583 + 0.2333 = 0.292 \end{aligned}$$

As expected, this probability matches the result obtained from the binomial probability distribution for X; that is, $P(X = 3) = \binom{4}{3}0.9^3 0.1^1 = 0.292$. The marginal probability distribution for X is found by summing the probabilities in each column, whereas the marginal probability distribution for Y is found by summing the probabilities in each row. The results are shown in Fig. 5-2.

Although the marginal probability distribution of X in the previous example can be determined directly from the description of the experiment, in some problems the marginal probability distribution is determined from the joint probability distribution.

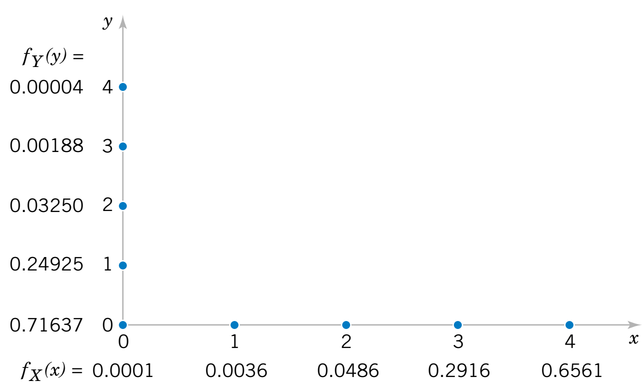

Figure 5-2 Marginal probability distributions of X and Y from Fig. 5-1.

Definition

If X and Y are discrete random variables with joint probability mass function $f_{XY}(x, y)$, then the **marginal probability mass functions** of X and Y are

$$f_X(x) = P(X = x) = \sum_{R_x} f_{XY}(x, y) \quad \text{and} \quad f_Y(y) = P(Y = y) = \sum_{R_y} f_{XY}(x, y) \tag{5-2}$$

where R_x denotes the set of all points in the range of (X, Y) for which $X = x$ and R_y denotes the set of all points in the range of (X, Y) for which $Y = y$

Given a joint probability mass function for random variables X and Y, $E(X)$ and $V(X)$ can be obtained directly from the joint probability distribution of X and Y or by first calculating the marginal probability distribution of X and then determining $E(X)$ and $V(X)$ by the usual method. This is shown in the following equation.

Mean and Variance from Joint Distribution

If the marginal probability distribution of X has the probability mass function $f_X(x)$, then

$$\begin{aligned} E(X) = \mu_X &= \sum_x x f_X(x) = \sum_x x \left(\sum_{R_x} f_{XY}(x, y) \right) = \sum_x \sum_{R_x} x f_{XY}(x, y) \\ &= \sum_R x f_{XY}(x, y) \end{aligned} \tag{5-3}$$

and

$$\begin{aligned} V(X) = \sigma_X^2 &= \sum_x (x - \mu_X)^2 f_X(x) = \sum_x (x - \mu_X)^2 \sum_{R_x} f_{XY}(x, y) \\ &= \sum_x \sum_{R_x} (x - \mu_X)^2 f_{XY}(x, y) = \sum_R (x - \mu_X)^2 f_{XY}(x, y) \end{aligned}$$

where R_x denotes the set of all points in the range of (X, Y) for which $X = x$ and R denotes the set of all points in the range of (X, Y)

EXAMPLE 5-4

In Example 5-1, $E(X)$ can be found as

$$\begin{aligned} E(X) = {} & 0[f_{XY}(0, 0) + f_{XY}(0, 1) + f_{XY}(0, 2) + f_{XY}(0, 3) + f_{XY}(0, 4)] \\ & + 1[f_{XY}(1, 0) + f_{XY}(1, 1) + f_{XY}(1, 2) + f_{XY}(1, 3)] \\ & + 2[f_{XY}(2, 0) + f_{XY}(2, 1) + f_{XY}(2, 2)] \\ & + 3[f_{XY}(3, 0) + f_{XY}(3, 1)] \\ & + 4[f_{XY}(4, 0)] \\ = {} & 0[0.0001] + 1[0.0036] + 2[0.0486] + 3[0.02916] + 4[0.6561] = 3.6 \end{aligned}$$

Alternatively, because the marginal probability distribution of X is binomial,

$$E(X) = np = 4(0.9) = 3.6$$

The calculation using the joint probability distribution can be used to determine $E(X)$ even in cases in which the marginal probability distribution of X is not known. As practice, you can use the joint probability distribution to verify that $E(Y) = 0.32$ in Example 5-1.

Also,

$$V(X) = np(1 - p) = 4(0.9)(1 - 0.9) = 0.36$$

Verify that the same result can be obtained from the joint probability distribution of X and Y.

5-1.3 Conditional Probability Distributions

When two random variables are defined in a random experiment, knowledge of one can change the probabilities that we associate with the values of the other. Recall that in Example 5-1, X denotes the number of acceptable bits and Y denotes the number of suspect bits received by a receiver. Because only four bits are transmitted, if $X = 4$, Y must equal 0. Using the notation for conditional probabilities from Chapter 2, we can write this result as $P(Y = 0 | X = 4) = 1$. If $X = 3$, Y can only equal 0 or 1. Consequently, the random variables X and Y can be considered to be dependent. Knowledge of the value obtained for X changes the probabilities associated with the values of Y.

Recall that the definition of conditional probability for events A and B is $P(B | A) = P(A \cap B)/P(A)$. This definition can be applied with the event A defined to be $X = x$ and event B defined to be $Y = y$.

EXAMPLE 5-5

For Example 5-1, X and Y denote the number of acceptable and suspect bits received, respectively. The remaining bits are unacceptable.

$$\begin{aligned} P(Y = 0 | X = 3) &= P(X = 3, Y = 0)/P(X = 3) \\ &= f_{XY}(3, 0)/f_X(3) = 0.05832/0.2916 = 0.200 \end{aligned}$$

The probability that $Y = 1$ given that $X = 3$ is

$$\begin{aligned} P(Y = 1 | X = 3) &= P(X = 3, Y = 1)/P(X = 3) \\ &= f_{XY}(3, 1)/f_X(3) = 0.2333/0.2916 = 0.800 \end{aligned}$$

Given that $X = 3$, the only possible values for Y are 0 and 1. Notice that $P(Y = 0 | X = 3) + P(Y = 1 | X = 3) = 1$. The values 0 and 1 for Y along with the probabilities 0.200 and 0.800 define the conditional probability distribution of Y given that $X = 3$.

Example 5-5 illustrates that the conditional probabilities that $Y = y$ given that $X = x$ can be thought of as a new probability distribution. The following definition generalizes these ideas.

Definition

Given discrete random variables X and Y with joint probability mass function $f_{XY}(x, y)$ the **conditional probability mass function** of Y given $X = x$ is

$$f_{Y|x}(y) = f_{XY}(x, y)/f_X(x) \qquad \text{for } f_X(x) > 0 \tag{5-4}$$

The function $f_{Y|x}(y)$ is used to find the probabilities of the possible values for Y given that $X = x$. That is, it is the probability mass function for the possible values of Y given that $X = x$. More precisely, let R_x denote the set of all points in the range of (X, Y) for which $X = x$. The conditional probability mass function provides the conditional probabilities for the values of Y in the set R_x.

Because a conditional probability mass function $f_{Y|x}(y)$ is a probability mass function for all y in R_x, the following properties are satisfied:

(1) $f_{Y|x}(y) \geq 0$

(2) $\sum_{R_x} f_{Y|x}(y) = 1$

(3) $P(Y = y | X = x) = f_{Y|x}(y)$ (5-5)

EXAMPLE 5-6

For the joint probability distribution in Fig. 5-1, $f_{Y|x}(y)$ is found by dividing each $f_{XY}(x, y)$ by $f_X(x)$. Here, $f_X(x)$ is simply the sum of the probabilities in each column of Fig. 5-1. The function $f_{Y|x}(y)$ is shown in Fig. 5-3. In Fig. 5-3, each column sums to one because it is a probability distribution.

Properties of random variables can be extended to a conditional probability distribution of Y given $X = x$. The usual formulas for mean and variance can be applied to a conditional probability mass function.

Definition

Let R_x denote the set of all points in the range of (X, Y) for which $X = x$. The **conditional mean** of Y given $X = x$, denoted as $E(Y|x)$ or $\mu_{Y|x}$, is

$$E(Y|x) = \sum_{R_x} y f_{Y|x}(y) \tag{5-6}$$

and the **conditional variance** of Y given $X = x$, denoted as $V(Y|x)$ or $\sigma^2_{Y|x}$, is

$$V(Y|x) = \sum_{R_x} (y - \mu_{Y|x})^2 f_{Y|x}(y) = \sum_{R_x} y^2 f_{Y|x}(y) - \mu^2_{Y|x}$$

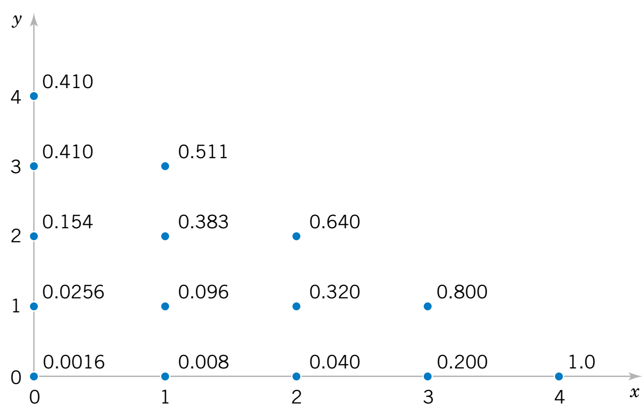

Figure 5-3 Conditional probability distributions of Y given $X = x$, $f_{Y|x}(y)$ in Example 5-6.

EXAMPLE 5-7

For the random variables in Example 5-1, the conditional mean of Y given $X = 2$ is obtained from the conditional distribution in Fig. 5-3:

$$E(Y|2) = \mu_{Y|2} = 0(0.040) + 1(0.320) + 2(0.640) = 1.6$$

The conditional mean is interpreted as the expected number of acceptable bits given that two of the four bits transmitted are suspect. The conditional variance of Y given $X = 2$ is

$$V(Y|2) = (0 - \mu_{Y|2})^2(0.040) + (1 - \mu_{Y|2})^2(0.320) + (2 - \mu_{Y|2})^2(0.640) = 0.32$$

5-1.4 Independence

In some random experiments, knowledge of the values of X does not change any of the probabilities associated with the values for Y.

EXAMPLE 5-8

In a plastic molding operation, each part is classified as to whether it conforms to color and length specifications. Define the random variable X and Y as

$$X = \begin{cases} 1 & \text{if the part conforms to color specifications} \\ 0 & \text{otherwise} \end{cases}$$
$$Y = \begin{cases} 1 & \text{if the part conforms to length specifications} \\ 0 & \text{otherwise} \end{cases}$$

Assume the joint probability distribution of X and Y is defined by $f_{XY}(x, y)$ in Fig. 5-4(a). The marginal probability distributions of X and Y are also shown in Fig. 5-4(a). Note that $f_{XY}(x, y) = f_X(x)\, f_Y(y)$. The conditional probability mass function $f_{Y|x}(y)$ is shown in Fig. 5-4(b). Notice that for any x, $f_{Y|x}(y) = f_Y(y)$. That is, knowledge of whether or not the part meets color specifications does not change the probability that it meets length specifications.

By analogy with independent events, we define two random variables to be **independent** whenever $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for all x and y. Notice that independence implies that $f_{XY}(x, y) = f_X(x) f_Y(y)$ for *all* x and y. If we find one pair of x and y in which the equality fails, X and Y are not independent. If two random variables are independent, then

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y)$$

With similar calculations, the following equivalent statements can be shown.

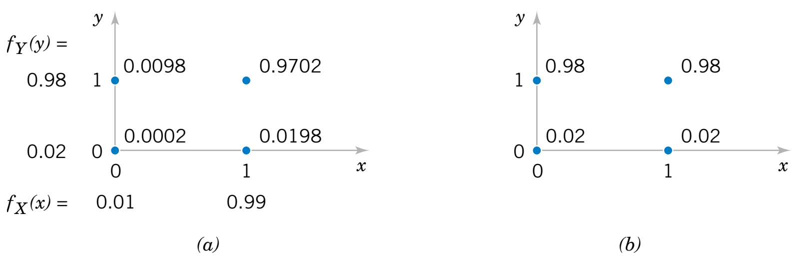

Figure 5-4 (a) Joint and marginal probability distributions of X and Y in Example 5-8. (b) Conditional probability distribution of Y given $X = x$ in Example 5-8.

For discrete random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are **independent.**

(1) $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y

(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y, respectively. (5-7)

Rectangular Range for (X, Y)!

If the set of points in two-dimensional space that receive positive probability under $f_{XY}(x, y)$ does not form a rectangle, X and Y are not independent because knowledge of X can restrict the range of values of Y that receive positive probability. In Example 5-1 knowledge that $X = 3$ implies that Y can equal only 0 or 1. Consequently, the marginal probability distribution of Y does not equal the conditional probability distribution $f_{Y|3}(y)$ for $X = 3$. Using this idea, we know immediately that the random variables X and Y with joint probability mass function in Fig. 5-1 are not independent. If the set of points in two-dimensional space that receives positive probability under $f_{XY}(x, y)$ forms a rectangle, independence is possible but not demonstrated. One of the conditions in Equation 5-7 must still be verified.

Rather than verifying independence from a joint probability distribution, knowledge of the random experiment is often used to assume that two random variables are independent. Then, the joint probability mass function of X and Y is computed from the product of the marginal probability mass functions.

EXAMPLE 5-9

In a large shipment of parts, 1% of the parts do not conform to specifications. The supplier inspects a random sample of 30 parts, and the random variable X denotes the number of parts in the sample that do not conform to specifications. The purchaser inspects another random sample of 20 parts, and the random variable Y denotes the number of parts in this sample that do not conform to specifications. What is the probability that $X \leq 1$ and $Y \leq 1$?

Although the samples are typically selected without replacement, if the shipment is large, relative to the sample sizes being used, approximate probabilities can be computed by assuming the sampling is with replacement and that X and Y are independent. With this assumption, the marginal probability distribution of X is binomial with $n = 30$ and $p = 0.01$, and the marginal probability distribution of Y is binomial with $n = 20$ and $p = 0.01$.

If independence between X and Y were not assumed, the solution would have to proceed as follows:

$$\begin{aligned} P(X \leq 1, Y \leq 1) &= P(X = 0, Y = 0) + P(X = 1, Y = 0) \\ &\quad + P(X = 0, Y = 1) + P(X = 1, Y = 1) \\ &= f_{XY}(0, 0) + f_{XY}(1, 0) + f_{XY}(0, 1) + f_{XY}(1, 1) \end{aligned}$$

However, with independence, property (4) of Equation 5-7 can be used as

$$P(X \leq 1, Y \leq 1) = P(X \leq 1)\, P(Y \leq 1)$$

and the binomial distributions for X and Y can be used to determine these probabilities as $P(X \le 1) = 0.9639$ and $P(Y \le 1) = 0.9831$. Therefore, $P(X \le 1, Y \le 1) = 0.948$.

Consequently, the probability that the shipment is accepted for use in manufacturing is 0.948 even if 1% of the parts do not conform to specifications. If the supplier and the purchaser change their policies so that the shipment is acceptable only if zero nonconforming parts are found in the sample, the probability that the shipment is accepted for production is still quite high. That is,

$$P(X = 0, Y = 0) = P(X = 0)P(Y = 0) = 0.605$$

This example shows that inspection is not an effective means of achieving quality.

EXERCISES FOR SECTION 5-1

5-1. Show that the following function satisfies the properties of a joint probability mass function.

x	y	$f_{XY}(x, y)$
1	1	1/4
1.5	2	1/8
1.5	3	1/4
2.5	4	1/4
3	5	1/8

5-2. Continuation of Exercise 5-1. Determine the following probabilities:
(a) $P(X < 2.5, Y < 3)$ (b) $P(X < 2.5)$
(c) $P(Y < 3)$ (d) $P(X > 1.8, Y > 4.7)$

5-3. Continuation of Exercise 5-1. Determine $E(X)$ and $E(Y)$.

5-4. Continuation of Exercise 5-1. Determine
(a) The marginal probability distribution of the random variable X.
(b) The conditional probability distribution of Y given that $X = 1.5$.
(c) The conditional probability distribution of X given that $Y = 2$.
(d) $E(Y|X = 1.5)$
(e) Are X and Y independent?

5-5. Determine the value of c that makes the function $f(x, y) = c(x + y)$ a joint probability mass function over the nine points with $x = 1, 2, 3$ and $y = 1, 2, 3$.

5-6. Continuation of Exercise 5-5. Determine the following probabilities:
(a) $P(X = 1, Y < 4)$ (b) $P(X = 1)$
(c) $P(Y = 2)$ (d) $P(X < 2, Y < 2)$

5-7. Continuation of Exercise 5-5. Determine $E(X)$, $E(Y)$, $V(X)$, and $V(Y)$.

5-8. Continuation of Exercise 5-5. Determine
(a) The marginal probability distribution of the random variable X.
(b) The conditional probability distribution of Y given that $X = 1$.
(c) The conditional probability distribution of X given that $Y = 2$.
(d) $E(Y|X = 1)$
(e) Are X and Y independent?

5-9. Show that the following function satisfies the properties of a joint probability mass function.

x	y	$f_{XY}(x, y)$
−1	−2	1/8
−0.5	−1	1/4
0.5	1	1/2
1	2	1/8

5-10. Continuation of Exercise 5-9. Determine the following probabilities:
(a) $P(X < 0.5, Y < 1.5)$ (b) $P(X < 0.5)$
(c) $P(Y < 1.5)$ (d) $P(X > 0.25, Y < 4.5)$

5-11. Continuation of Exercise 5-9. Determine $E(X)$ and $E(Y)$.

5-12. Continuation of Exercise 5-9. Determine
(a) The marginal probability distribution of the random variable X.
(b) The conditional probability distribution of Y given that $X = 1$.
(c) The conditional probability distribution of X given that $Y = 1$.
(d) $E(X|y = 1)$
(e) Are X and Y independent?

5-13. Four electronic printers are selected from a large lot of damaged printers. Each printer is inspected and classified as containing either a major or a minor defect. Let the random variables X and Y denote the number of printers with major and minor defects, respectively. Determine the range of the joint probability distribution of X and Y.

5-14. In the transmission of digital information, the probability that a bit has high, moderate, and low distortion is 0.01, 0.10, and 0.95, respectively. Suppose that three bits are transmitted and that the amount of distortion of each bit is assumed to be independent. Let X and Y denote the number of bits with high and moderate distortion out of the three, respectively. Determine

(a) $f_{XY}(x, y)$ (b) $f_X(x)$
(c) $E(X)$ (d) $f_{Y|1}(y)$
(e) $E(Y|X = 1)$ (f) Are X and Y independent?

5-15. A small-business Web site contains 100 pages and 60%, 30%, and 10% of the pages contain low, moderate, and high graphic content, respectively. A sample of four pages is selected without replacement, and X and Y denote the number of pages with moderate and high graphics output in the sample. Determine

(a) $f_{XY}(x, y)$ (b) $f_X(x)$
(c) $E(X)$ (d) $f_{Y|3}(y)$
(e) $E(Y|X = 3)$ (f) $V(Y|X = 3)$
(g) Are X and Y independent?

5-16. A manufacturing company employs two inspecting devices to sample a fraction of their output for quality control purposes. The first inspection monitor is able to accurately detect 99.3% of the defective items it receives, whereas the second is able to do so in 99.7% of the cases. Assume that four defective items are produced and sent out for inspection. Let X and Y denote the number of items that will be identified as defective by inspecting devices 1 and 2, respectively. Assume the devices are independent. Determine

(a) $f_{XY}(x, y)$ (b) $f_X(x)$
(c) $E(X)$ (d) $f_{Y|2}(y)$
(e) $E(Y|X = 2)$ (f) $V(Y|X = 2)$
(g) Are X and Y independent?

5-2 MULTIPLE DISCRETE RANDOM VARIABLES

5-2.1 Joint Probability Distributions

EXAMPLE 5-10 In some cases, more than two random variables are defined in a random experiment, and the concepts presented earlier in the chapter can easily be extended. The notation can be cumbersome and if doubts arise, it is helpful to refer to the equivalent concept for two random variables. Suppose that the quality of each bit received in Example 5-1 is categorized even more finely into one of the four classes, excellent, good, fair, or poor, denoted by E, G, F, and P, respectively. Also, let the random variables X_1, X_2, X_3, and X_4 denote the number of bits that are E, G, F, and P, respectively, in a transmission of 20 bits. In this example, we are interested in the joint probability distribution of four random variables. Because each of the 20 bits is categorized into one of the four classes, only values for x_1, x_2, x_3, and x_4 such that $x_1 + x_2 + x_3 + x_4 = 20$ receive positive probability in the probability distribution.

In general, given discrete random variables $X_1, X_2, X_3, \ldots, X_p$, the joint probability distribution of $X_1, X_2, X_3, \ldots, X_p$ is a description of the set of points $(x_1, x_2, x_3, \ldots, x_p)$ in the range of $X_1, X_2, X_3, \ldots, X_p$, along with the probability of each point. A joint probability mass function is a simple extension of a bivariate probability mass function.

Definition

The **joint probability mass function** of $X_1, X_2, \ldots, X_p$ is

$$f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p) = P(X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p) \qquad (5\text{-}8)$$

for all points $(x_1, x_2, \ldots, x_p)$ in the range of $X_1, X_2, \ldots, X_p$.

A marginal probability distribution is a simple extension of the result for two random variables.

Definition

If $X_1, X_2, X_3, \ldots, X_p$ are discrete random variables with joint probability mass function $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$, the **marginal probability mass function** of any X_i is

$$f_{X_i}(x_i) = P(X_i = x_i) = \sum_{R_{x_i}} f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p) \tag{5-9}$$

where R_{x_i} denotes the set of points in the range of $(X_1, X_2, \ldots, X_p)$ for which $X_i = x_i$.

EXAMPLE 5-11

Points that have positive probability in the joint probability distribution of three random variables X_1, X_2, X_3 are shown in Fig. 5-5. The range is the nonnegative integers with $x_1 + x_2 + x_3 = 3$. The marginal probability distribution of X_2 is found as follows.

$$P(X_2 = 0) = f_{X_1X_2X_3}(3, 0, 0) + f_{X_1X_2X_3}(0, 0, 3) + f_{X_1X_2X_3}(1, 0, 2) + f_{X_1X_2X_3}(2, 0, 1)$$
$$P(X_2 = 1) = f_{X_1X_2X_3}(2, 1, 0) + f_{X_1X_2X_3}(0, 1, 2) + f_{X_1X_2X_3}(1, 1, 1)$$
$$P(X_2 = 2) = f_{X_1X_2X_3}(1, 2, 0) + f_{X_1X_2X_3}(0, 2, 1)$$
$$P(X_2 = 3) = f_{X_1X_2X_3}(0, 3, 0)$$

Furthermore, $E(X_i)$ and $V(X_i)$ for $i = 1, 2, \ldots, p$ can be determined from the marginal probability distribution of X_i or from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows.

Mean and Variance from Joint Distribution

$$E(X_i) = \sum_R x_i f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$$

and

$$V(X_i) = \sum_R (x_i - \mu_{X_i})^2 f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p) \tag{5-10}$$

where R is the set of all points in the range of $X_1, X_2, \ldots, X_p$.

With several random variables, we might be interested in the probability distribution of some subset of the collection of variables. The probability distribution of $X_1, X_2, \ldots, X_k$, $k < p$ can be obtained from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows.

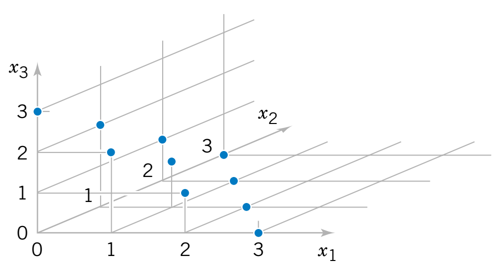

Figure 5-5 Joint probability distribution of X_1, X_2, and X_3.

Distribution of a Subset of Random Variables

If $X_1, X_2, X_3, \dots, X_p$ are discrete random variables with joint probability mass function $f_{X_1X_2\dots X_p}(x_1, x_2, \dots, x_p)$, the joint **probability mass function** of $X_1, X_2, \dots, X_k$, $k < p$, is

$$f_{X_1X_2\dots X_k}(x_1, x_2, \dots, x_k) = P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$$
$$= \sum_{R_{x_1x_2\dots x_k}} P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) \qquad \text{(5-11)}$$

where $R_{x_1x_2\dots x_k}$ denotes the set of all points in the range of $X_1, X_2, \dots, X_p$ for which $X_1 = x_1, X_2 = x_2, \dots, X_k = x_k$.

That is, $P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k)$ is the sum of the probabilities over all points in the range of $X_1, X_2, X_3, \dots, X_p$ for which $X_1 = x_1$, $X_2 = x_2, \dots,$ and $X_k = x_k$. An example is presented in the next section. Any k random variables can be used in the definition. The first k simplifies the notation.

Conditional Probability Distributions

Conditional probability distributions can be developed for multiple discrete random variables by an extension of the ideas used for two discrete random variables. For example, the conditional joint probability mass function of X_1, X_2, X_3 given X_4, X_5 is

$$f_{X_1X_2X_3|x_4x_5}(x_1, x_2, x_3) = \frac{f_{X_1X_2X_3X_4X_5}(x_1, x_2, x_3, x_4, x_5)}{f_{X_4X_5}(x_4, x_5)}$$

for $f_{X_4X_5}(x_4, x_5) > 0$. The conditional joint probability mass function of X_1, X_2, X_3 given X_4, X_5 provides the conditional probabilities at all points in the range of X_1, X_2, X_3, X_4, X_5 for which $X_4 = x_4$ and $X_5 = x_5$.

The concept of independence can be extended to multiple discrete random variables.

Definition

Discrete variables $X_1, X_2, \dots, X_p$ are **independent** if and only if

$$f_{X_1X_2\dots X_p}(x_1, x_2, \dots, x_p) = f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_p}(x_p) \qquad \text{(5-12)}$$

for *all* $x_1, x_2, \dots, x_p$.

Similar to the result for bivariate random variables, independence implies that Equation 5-12 holds for *all* $x_1, x_2, \dots, x_p$. If we find one point for which the equality fails, $X_1, X_2, \dots, X_p$ are not independent. It can be shown that if $X_1, X_2, \dots, X_p$ are independent,

$$P(X_1 \in A_1, X_2 \in A_2, \dots, X_p \in A_p) = P(X_1 \in A_1) P(X_2 \in A_2) \dots P(X_p \in A_p)$$

for *any* sets $A_1, A_2, \dots, A_p$.

5-2.2 Multinomial Probability Distribution

A joint probability distribution for multiple discrete random variables that is quite useful is an extension of the binomial. The random experiment that generates the probability distribution consists of a series of independent trials. However, the results from each trial can be categorized into one of k classes.

EXAMPLE 5-12

We might be interested in a probability such as the following. Of the 20 bits received, what is the probability that 14 are excellent, 3 are good, 2 are fair, and 1 is poor? Assume that the classifications of individual bits are independent events and that the probabilities of E, G, F, and P are 0.6, 0.3, 0.08, and 0.02, respectively. One sequence of 20 bits that produces the specified numbers of bits in each class can be represented as

$$EEEEEEEEEEEEEEGGGFFP$$

Using independence, we find that the probability of this sequence is

$$P(EEEEEEEEEEEEEEGGGFFP) = 0.6^{14}0.3^{3}0.08^{2}0.02^{1} = 2.708 \times 10^{-9}$$

Clearly, all sequences that consist of the same numbers of E's, G's, F's, and P's have the same probability. Consequently, the requested probability can be found by multiplying 2.708×10^{-9} by the number of sequences with 14 E's, three G's, two F's, and one P. The number of sequences is found from the CD material for Chapter 2 to be

$$\frac{20!}{14!3!2!1!} = 2325600$$

Therefore, the requested probability is

$$P(14E\text{'s, three } G\text{'s, two } F\text{'s, and one } P) = 2325600(2.708 \times 10^{-9}) = 0.0063$$

Example 5-12 leads to the following generalization of a binomial experiment and a binomial distribution.

Multinomial Distribution

Suppose a random experiment consists of a series of n trials. Assume that

(1) The result of each trial is classified into one of k classes.

(2) The probability of a trial generating a result in class 1, class 2, ..., class k is constant over the trials and equal to $p_1, p_2, \ldots, p_k$, respectively.

(3) The trials are independent.

The random variables $X_1, X_2, \ldots, X_k$ that denote the number of trials that result in class 1, class 2, ..., class k, respectively, have a **multinomial distribution** and the joint probability mass function is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} \qquad \text{(5-13)}$$

for $x_1 + x_2 + \cdots + x_k = n$ and $p_1 + p_2 + \cdots + p_k = 1$.

The multinomial distribution is considered a multivariable extension of the binomial distribution.

EXAMPLE 5-13

In Example 5-12, let the random variables X_1, X_2, X_3, and X_4 denote the number of bits that are E, G, F, and P, respectively, in a transmission of 20 bits. The probability that 12 of the bits received are E, 6 are G, 2 are F, and 0 are P is

$$P(X_1 = 12, X_2 = 6, X_3 = 2, X_4 = 0) = \frac{20!}{12!6!2!0!}\, 0.6^{12}0.3^{6}0.08^{2}0.02^{0} = 0.0358$$

Each trial in a multinomial random experiment can be regarded as either generating or not generating a result in class i, for each $i = 1, 2, \ldots, k$. Because the random variable X_i is the number of trials that result in class i, X_i has a binomial distribution.

> If $X_1, X_2, \ldots, X_k$ have a multinomial distribution, the marginal probability distribution of X_i is binomial with
>
> $$E(X_i) = np_i \quad \text{and} \quad V(X_i) = np_i(1 - p_i) \qquad (5\text{-}14)$$

EXAMPLE 5-14

In Example 5-13, the marginal probability distribution of X_2 is binomial with $n = 20$ and $p = 0.3$. Furthermore, the joint marginal probability distribution of X_2 and X_3 is found as follows. The $P(X_2 = x_2, X_3 = x_3)$ is the probability that exactly x_2 trials result in G and that x_3 result in F. The remaining $n - x_2 - x_3$ trials must result in either E or P. Consequently, we can consider each trial in the experiment to result in one of three classes, $\{G\}$, $\{F\}$, or $\{E, P\}$, with probabilities 0.3, 0.08, and $0.6 + 0.02 = 0.62$, respectively. With these new classes, we can consider the trials to comprise a new multinomial experiment. Therefore,

$$\begin{aligned} f_{X_2X_3}(x_2, x_3) &= P(X_2 = x_2, X_3 = x_3) \\ &= \frac{n!}{x_2!x_3!\,(n - x_2 - x_3)!}(0.3)^{x_2}(0.08)^{x_3}(0.62)^{n-x_2-x_3} \end{aligned}$$

The joint probability distribution of other sets of variables can be found similarly.

EXERCISES FOR SECTION 5-2

5-17. Suppose the random variables X, Y, and Z have the following joint probability distribution

x	y	z	$f(x, y, z)$
1	1	1	0.05
1	1	2	0.10
1	2	1	0.15
1	2	2	0.20
2	1	1	0.20
2	1	2	0.15
2	2	1	0.10
2	2	2	0.05

Determine the following:
(a) $P(X = 2)$ (b) $P(X = 1, Y = 2)$
(c) $P(Z < 1.5)$ (d) $P(X = 1$ or $Z = 2)$
(e) $E(X)$

5-18. Continuation of Exercise 5-17. Determine the following:
(a) $P(X = 1|Y = 1)$ (b) $P(X = 1, Y = 1|Z = 2)$
(c) $P(X = 1|Y = 1, Z = 2)$

5-19. Continuation of Exercise 5-17. Determine the conditional probability distribution of X given that $Y = 1$ and $Z = 2$.

5-20. Based on the number of voids, a ferrite slab is classified as either high, medium, or low. Historically, 5% of the slabs are classified as high, 85% as medium, and 10% as low.

A sample of 20 slabs is selected for testing. Let X, Y, and Z denote the number of slabs that are independently classified as high, medium, and low, respectively.
(a) What is the name and the values of the parameters of the joint probability distribution of X, Y, and Z?
(b) What is the range of the joint probability distribution of X, Y, Z?
(c) What is the name and the values of the parameters of the marginal probability distribution of X?
(d) Determine $E(X)$and $V(X)$.

5-21. Continuation of Exercise 5-20. Determine the following:
(a) $P(X = 1, Y = 17, Z = 3)$
(b) $P(X \leq 1, Y = 17, Z = 3)$
(c) $P(X \leq 1)$
(d) $E(X)$

5-22. Continuation of Exercise 5-20. Determine the following:
(a) $P(X = 2, Z = 3 \mid Y = 17)$ (b) $P(X = 2 \mid Y = 17)$
(c) $E(X \mid Y = 17)$

5-23. An order of 15 printers contains four with a graphics-enhancement feature, five with extra memory, and six with both features. Four printers are selected at random, without replacement, from this set. Let the random variables X, Y, and Z denote the number of printers in the sample with graphics enhancement only, extra memory only, and both, respectively.
(a) Describe the range of the joint probability distribution of X, Y, and Z.
(b) Is the probability distribution of X, Y, and Z multinomial? Why or why not?

5-24. Continuation of Exercise 5-23. Determine the conditional probability distribution of X given that $Y = 2$.

5-25. Continuation of Exercise 5-23. Determine the following:
(a) $P(X = 1, Y = 2, Z = 1)$ (b) $P(X = 1, Y = 1)$
(c) $E(X)$ and $V(X)$

5-26. Continuation of Exercise 5-23. Determine the following:
(a) $P(X = 1, Y = 2 \mid Z = 1)$ (b) $P(X = 2 \mid Y = 2)$
(c) The conditional probability distribution of X given that $Y = 0$ and $Z = 3$.

5-27. Four electronic ovens that were dropped during shipment are inspected and classified as containing either a major, a minor, or no defect. In the past, 60% of dropped ovens had a major defect, 30% had a minor defect, and 10% had no defect. Assume that the defects on the four ovens occur independently.
(a) Is the probability distribution of the count of ovens in each category multinomial? Why or why not?
(b) What is the probability that, of the four dropped ovens, two have a major defect and two have a minor defect?
(c) What is the probability that no oven has a defect?

5-28. Continuation of Exercise 5-27. Determine the following:
(a) The joint probability mass function of the number of ovens with a major defect and the number with a minor defect.
(b) The expected number of ovens with a major defect.
(c) The expected number of ovens with a minor defect.

5-29. Continuation of Exercise 5-27. Determine the following:
(a) The conditional probability that two ovens have major defects given that two ovens have minor defects
(b) The conditional probability that three ovens have major defects given that two ovens have minor defects
(c) The conditional probability distribution of the number of ovens with major defects given that two ovens have minor defects
(d) The conditional mean of the number of ovens with major defects given that two ovens have minor defects

5-30. In the transmission of digital information, the probability that a bit has high, moderate, or low distortion is 0.01, 0.04, and 0.95, respectively. Suppose that three bits are transmitted and that the amount of distortion of each bit is assumed to be independent.
(a) What is the probability that two bits have high distortion and one has moderate distortion?
(b) What is the probability that all three bits have low distortion?

5-31. Continuation of Exercise 5-30. Let X and Y denote the number of bits with high and moderate distortion out of the three transmitted, respectively. Determine the following:
(a) The probability distribution, mean and variance of X.
(b) The conditional probability distribution, conditional mean and conditional variance of X given that $Y = 2$.

5-32. A marketing company performed a risk analysis for a manufacturer of synthetic fibers and concluded that new competitors present no risk 13% of the time (due mostly to the diversity of fibers manufactured), moderate risk 72% of the time (some overlapping of products), and very high risk (competitor manufactures the exact same products) 15% of the time. It is known that 12 international companies are planning to open new facilities for the manufacture of synthetic fibers within the next three years. Assume the companies are independent. Let X, Y, and Z denote the number of new competitors that will pose no, moderate, and very high risk for the interested company, respectively.
(a) What is the range of the joint probability distribution of X, Y, and Z?
(b) Determine $P(X = 1, Y = 3, Z = 1)$
(c) Determine $P(Z \leq 2)$

5-33. Continuation of Exercise 5-32. Determine the following:
(a) $P(Z = 2 \mid Y = 1, X = 10)$ (b) $P(Z \leq 1 \mid X = 10)$
(c) $P(Y \leq 1, Z \leq 1 \mid X = 10)$ (d) $E(Z \mid X = 10)$

5-3 TWO CONTINUOUS RANDOM VARIABLES

5-3.1 Joint Probability Distributions

Our presentation of the joint probability distribution of two continuous random variables is similar to our discussion of two discrete random variables. As an example, let the continuous random variable X denote the length of one dimenson of an injection-molded part, and let the continuous random variable Y denote the length of another dimension. The sample space of the random experiment consists of points in two dimensions.

We can study each random variable separately. However, because the two random variables are measurements from the same part, small disturbances in the injection-molding process, such as pressure and temperature variations, might be more likely to generate values for X and Y in specific regions of two-dimensional space. For example, a small pressure increase might generate parts such that both X and Y are greater than their respective targets and a small pressure decrease might generate parts such that X and Y are both less than their respective targets. Therefore, based on pressure variations, we expect that the probability of a part with X much greater than its target and Y much less than its target is small. Knowledge of the joint probability distribution of X and Y provides information that is not obvious from the marginal probability distributions.

The joint probability distribution of two continuous random variables X and Y can be specified by providing a method for calculating the probability that X and Y assume a value in any region R of two-dimensional space. Analogous to the probability density function of a single continuous random variable, a **joint probability density function** can be defined over two-dimensional space. The double integral of $f_{XY}(x, y)$ over a region R provides the probability that (X, Y) assumes a value in R. This integral can be interpreted as the volume under the surface $f_{XY}(x, y)$ over the region R.

A joint probability density function for X and Y is shown in Fig. 5-6. The probability that (X, Y) assumes a value in the region R equals the volume of the shaded region in Fig. 5-6. In this manner, a joint probability density function is used to determine probabilities for X and Y.

Definition

A **joint probability density function** for the continuous random variables X and Y, denoted as $f_{XY}(x, y)$, satisfies the following properties:

(1) $f_{XY}(x, y) \geq 0$ for all x, y

(2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\, dy = 1$

(3) For any region R of two-dimensional space

$$P([X, Y] \in R) = \iint_R f_{XY}(x, y)\, dx\, dy \quad (5\text{-}15)$$

Typically, $f_{XY}(x, y)$ is defined over all of two-dimensional space by assuming that $f_{XY}(x, y) = 0$ for all points for which $f_{XY}(x, y)$ is not specified.

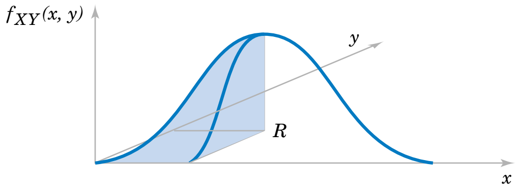

Figure 5-6 Joint probability density function for random variables X and Y.

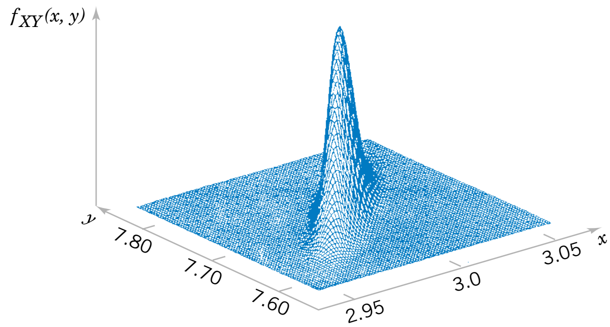

Figure 5-7 Joint probability density function for the lengths of different dimensions of an injection-molded part.

At the start of this chapter, the lengths of different dimensions of an injection-molded part were presented as an example of two random variables. Each length might be modeled by a normal distribution. However, because the measurements are from the same part, the random variables are typically not independent. A probability distribution for two normal random variables that are not independent is important in many applications and it is presented later in this chapter. If the specifications for X and Y are 2.95 to 3.05 and 7.60 to 7.80 millimeters, respectively, we might be interested in the probability that a part satisfies both specifications; that is, $P(2.95 < X < 3.05, 7.60 < Y < 7.80)$. Suppose that $f_{XY}(x, y)$ is shown in Fig. 5-7. The required probability is the volume of $f_{XY}(x, y)$ within the specifications. Often a probability such as this must be determined from a numerical integration.

EXAMPLE 5-15

Let the random variable X denote the time until a computer server connects to your machine (in milliseconds), and let Y denote the time until the server authorizes you as a valid user (in milliseconds). Each of these random variables measures the wait from a common starting time and $X < Y$. Assume that the joint probability density function for X and Y is

$$f_{XY}(x, y) = 6 \times 10^{-6} \exp(-0.001x - 0.002y) \quad \text{for } x < y$$

Reasonable assumptions can be used to develop such a distribution, but for now, our focus is only on the joint probability density function.

The region with nonzero probability is shaded in Fig. 5-8. The property that this joint probability density function integrates to 1 can be verified by the integral of $f_{XY}(x, y)$ over this region as follows:

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x, y)\, dy\, dx &= \int_{0}^{\infty}\left(\int_{x}^{\infty} 6 \times 10^{-6} e^{-0.001x - 0.002y}\, dy\right) dx \\
&= 6 \times 10^{-6} \int_{0}^{\infty}\left(\int_{x}^{\infty} e^{-0.002y}\, dy\right) e^{-0.001x}\, dx \\
&= 6 \times 10^{-6} \int_{0}^{\infty}\left(\frac{e^{-0.002x}}{0.002}\right) e^{-0.001x}\, dx \\
&= 0.003\left(\int_{0}^{\infty} e^{-0.003x}\, dx\right) = 0.003\left(\frac{1}{0.003}\right) = 1
\end{aligned}$$

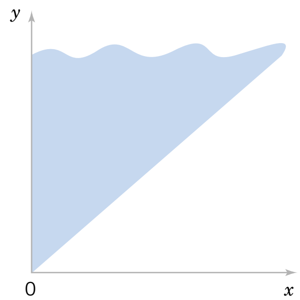

Figure 5-8 The joint probability density function of X and Y is nonzero over the shaded region.

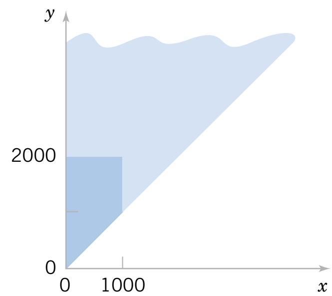

Figure 5-9 Region of integration for the probability that $X < 1000$ and $Y < 2000$ is darkly shaded.

The probability that $X < 1000$ and $Y < 2000$ is determined as the integral over the darkly shaded region in Fig. 5-9.

$$
\begin{aligned}
P(X \le 1000, Y \le 2000) &= \int_0^{1000} \int_x^{2000} f_{XY}(x, y)\, dy\, dx \\
&= 6 \times 10^{-6} \int_0^{1000} \left(\int_x^{2000} e^{-0.002y}\, dy \right) e^{-0.001x}\, dx \\
&= 6 \times 10^{-6} \int_0^{1000} \left(\frac{e^{-0.002x} - e^{-4}}{0.002} \right) e^{-0.001x}\, dx \\
&= 0.003 \int_0^{1000} e^{-0.003x} - e^{-4} e^{-0.001x}\, dx \\
&= 0.003 \left[\left(\frac{1 - e^{-3}}{0.003} \right) - e^{-4} \left(\frac{1 - e^{-1}}{0.001} \right) \right] \\
&= 0.003\,(316.738 - 11.578) = 0.915
\end{aligned}
$$

5-3.2 Marginal Probability Distributions

Similar to joint discrete random variables, we can find the marginal probability distributions of X and Y from the joint probability distribution.

Definition

If the joint probability density function of continuous random variables X and Y is $f_{XY}(x, y)$, the **marginal probability density functions** of X and Y are

$$
f_X(x) = \int_{R_x} f_{XY}(x, y)\, dy \quad \text{and} \quad f_Y(y) = \int_{R_y} f_{XY}(x, y)\, dx \tag{5-16}
$$

where R_x denotes the set of all points in the range of (X, Y) for which $X = x$ and R_y denotes the set of all points in the range of (X, Y) for which $Y = y$

A probability involving only one random variable, say, for example, $P(a < X < b)$, can be found from the marginal probability distribution of X or from the joint probability distribution of X and Y. For example, $P(a < X < b)$ equals $P(a < X < b, -\infty < Y < \infty)$. Therefore,

$$P(a < X < b) = \int_a^b \int_{R_x} f_{XY}(x, y)\, dy\, dx = \int_a^b \left(\int_{R_x} f_{XY}(x, y)\, dy \right) dx = \int_a^b f_X(x)\, dx$$

Similarly, $E(X)$ and $V(X)$ can be obtained directly from the joint probability distribution of X and Y or by first calculating the marginal probability distribution of X. The details, shown in the following equations, are similar to those used for discrete random variables.

Mean and Variance from Joint Distribution

$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_{-\infty}^{\infty} x \left[\int_{R_x} f_{XY}(x, y)\, dy \right] dx$$

$$= \iint_R x f_{XY}(x, y)\, dx\, dy \tag{5-17}$$

and

$$V(X) = \sigma_X^2 \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx = \int_{-\infty}^{\infty} (x - \mu_X)^2 \left[\int_{R_x} f_{XY}(x, y)\, dy \right] dx$$

$$= \iint_R (x - \mu_X)^2 f_{XY}(x, y)\, dx\, dy$$

where R_X denotes the set of all points in the range of (X, Y) for which $X = x$ and R_Y denotes the set of all points in the range of (X, Y)

EXAMPLE 5-16 For the random variables that denote times in Example 5-15, calculate the probability that Y exceeds 2000 milliseconds.

This probability is determined as the integral of $f_{XY}(x, y)$ over the darkly shaded region in Fig. 5-10. The region is partitioned into two parts and different limits of integration are determined for each part.

$$P(Y > 2000) = \int_0^{2000} \left(\int_{2000}^{\infty} 6 \times 10^{-6} e^{-0.001x - 0.002y}\, dy \right) dx$$

$$+ \int_{2000}^{\infty} \left(\int_x^{\infty} 6 \times 10^{-6} e^{-0.001x - 0.002y}\, dy \right) dx$$

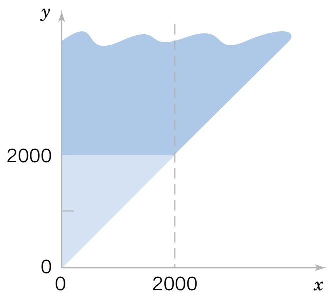

Figure 5-10 Region of integration for the probability that $Y < 2000$ is darkly shaded and it is partitioned into two regions with $x < 2000$ and $x > 2000$.

The first integral is

$$6 \times 10^{-6} \int_0^{2000} \left(\frac{e^{-0.002y}}{-0.002} \Big|_{2000}^{\infty} \right) e^{-0.001x}\, dx = \frac{6 \times 10^{-6}}{0.002} e^{-4} \int_0^{2000} e^{-0.001x}\, dx$$

$$= \frac{6 \times 10^{-6}}{0.002} e^{-4} \left(\frac{1 - e^{-2}}{0.001} \right) = 0.0475$$

The second integral is

$$6 \times 10^{-6} \int_{2000}^{\infty} \left(\frac{e^{-0.002y}}{-0.002} \Big|_{x}^{\infty} \right) e^{-0.001x}\, dx = \frac{6 \times 10^{-6}}{0.002} \int_{2000}^{\infty} e^{-0.003x}\, dx$$

$$= \frac{6 \times 10^{-6}}{0.002} \left(\frac{e^{-6}}{0.003} \right) = 0.0025$$

Therefore,

$$P(Y > 2000) = 0.0475 + 0.0025 = 0.05.$$

Alternatively, the probability can be calculated from the marginal probability distribution of Y as follows. For $y > 0$

$$f_Y(y) = \int_0^y 6 \times 10^{-6} e^{-0.001x - 0.002y}\, dx = 6 \times 10^{-6} e^{-0.002y} \int_0^y e^{-0.001x}\, dx$$

$$= 6 \times 10^{-6} e^{-0.002y} \left(\frac{e^{-0.001x}}{-0.001} \Big|_0^y \right) = 6 \times 10^{-6} e^{-0.002y} \left(\frac{1 - e^{-0.001y}}{0.001} \right)$$

$$= 6 \times 10^{-3} e^{-0.002y} (1 - e^{-0.001y}) \quad \text{for } y > 0$$

We have obtained the marginal probability density function of Y. Now,

$$P(Y > 2000) = 6 \times 10^{-3} \int_{2000}^{\infty} e^{-0.002y}(1 - e^{-0.001y})\, dy$$

$$= 6 \times 10^{-3} \left[\left(\frac{e^{-0.002y}}{-0.002} \bigg|_{2000}^{\infty} \right) - \left(\frac{e^{-0.003y}}{-0.003} \bigg|_{2000}^{\infty} \right) \right]$$

$$= 6 \times 10^{-3} \left[\frac{e^{-4}}{0.002} - \frac{e^{-6}}{0.003} \right] = 0.05$$

5-3.3 Conditional Probability Distributions

Analogous to discrete random variables, we can define the conditional probability distribution of Y given $X = x$.

Definition

Given continuous random variables X and Y with joint probability density function $f_{XY}(x, y)$, the **conditional probability density function** of Y given $X = x$ is

$$f_{Y|x}(y) = \frac{f_{XY}(x, y)}{f_X(x)} \quad \text{for} \quad f_X(x) > 0 \tag{5-18}$$

The function $f_{Y|x}(y)$ is used to find the probabilities of the possible values for Y given that $X = x$. Let R_x denote the set of all points in the range of (X, Y) for which $X = x$. The conditional probability density function provides the conditional probabilities for the values of Y in the set R_x.

Because the conditional probability density function $f_{Y|x}(y)$ is a probability density function for all y in R_x, the following properties are satisfied:

(1) $f_{Y|x}(y) \geq 0$

(2) $\int_{R_x} f_{Y|x}(y)\, dy = 1$

(3) $P(Y \in B | X = x) = \int_B f_{Y|x}(y)\, dy$ for any set B in the range of Y

(5-19)

It is important to state the region in which a joint, marginal, or conditional probability density function is not zero. The following example illustrates this.

EXAMPLE 5-17 For the random variables that denote times in Example 5-15, determine the conditional probability density function for Y given that $X = x$.

First the marginal density function of x is determined. For $x > 0$

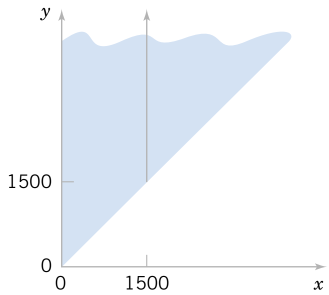

Figure 5-11 The conditional probability density function for Y, given that $x = 1500$, is nonzero over the solid line.

$$f_X(x) = \int_x^\infty 6 \times 10^{-6} e^{-0.001x - 0.002y}\, dy = 6 \times 10^{-6} e^{-0.001x} \left(\frac{e^{-0.002y}}{-0.002} \bigg|_x^\infty \right)$$

$$= 6 \times 10^{-6} e^{-0.001x} \left(\frac{e^{-0.002x}}{0.002} \right) = 0.003 e^{-0.003x} \quad \text{for} \quad x > 0$$

This is an exponential distribution with $\lambda = 0.003$. Now, for $0 < x$ and $x < y$ the conditional probability density function is

$$f_{Y|x}(y) = f_{XY}(x, y)/f_x(x) = \frac{6 \times 10^{-6} e^{-0.001x - 0.002y}}{0.003 e^{-0.003x}}$$

$$= 0.002 e^{0.002x - 0.002y} \quad \text{for } 0 < x \quad \text{and} \quad x < y$$

The conditional probability density function of Y, given that $x = 1500$, is nonzero on the solid line in Fig. 5-11.

Determine the probability that Y exceeds 2000, given that $x = 1500$. That is, determine $P(Y > 2000 \mid x = 1500)$.The conditional probability density function is integrated as follows:

$$P(Y > 2000|x = 1500) = \int_{2000}^\infty f_{Y|1500}(y)\, dy = \int_{2000}^\infty 0.002 e^{0.002(1500) - 0.002y}\, dy$$

$$= 0.002 e^3 \left(\frac{e^{-0.002y}}{-0.002} \bigg|_{2000}^\infty \right) = 0.002 e^3 \left(\frac{e^{-4}}{0.002} \right) = 0.368$$

Definition

Let R_x denote the set of all points in the range of (X, Y) for which $X = x$. The **conditional mean** of Y given $X = x$, denoted as $E(Y|x)$ or $\mu_{Y|x}$, is

$$E(Y|x) = \int_{R_x} y f_{Y|x}(y)\, dy$$

and the **conditional variance** of Y given $X = x$, denoted as $V(Y|x)$ or $\sigma^2_{Y|x}$, is

$$V(Y|x) = \int_{R_x} (y - \mu_{Y|x})^2 f_{Y|x}(y)\, dy = \int_{R_x} y^2 f_{Y|x}(y)\, dy - \mu^2_{Y|x} \tag{5-20}$$

EXAMPLE 5-18

For the random variables that denote times in Example 5-15, determine the conditional mean for Y given that $x = 1500$.

The conditional probability density function for Y was determined in Example 5-17. Because $f_{Y|1500}(y)$ is nonzero for $y > 1500$,

$$E(Y|x = 1500) = \int_{1500}^{\infty} y\left(0.002e^{0.002(1500)-0.002y}\right) dy = 0.002e^{3} \int_{1500}^{\infty} ye^{-0.002y}\, dy$$

Integrate by parts as follows:

$$\int_{1500}^{\infty} ye^{-0.002y}\, dy = y\frac{e^{-0.002y}}{-0.002}\bigg|_{1500}^{\infty} - \int_{1500}^{\infty}\left(\frac{e^{-0.002y}}{-0.002}\right) dy$$

$$= \frac{1500}{0.002}e^{-3} - \left(\frac{e^{-0.002y}}{(-0.002)(-0.002)}\bigg|_{1500}^{\infty}\right)$$

$$= \frac{1500}{0.002}e^{-3} + \frac{e^{-3}}{(0.002)(0.002)} = \frac{e^{-3}}{0.002}(2000)$$

With the constant $0.002e^{3}$ reapplied

$$E(Y|x = 1500) = 2000$$

5-3.4 Independence

The definition of independence for continuous random variables is similar to the definition for discrete random variables. If $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for all x and y, X and Y are **independent.** Independence implies that $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for *all* x and y. If we find one pair of x and y in which the equality fails, X and Y are not independent.

Definition

For continuous random variables X and Y, if any one of the following properties is true, the others are also true, and X and Y are said to be **independent.**

(1) $f_{XY}(x, y) = f_X(x)\, f_Y(y)$ for all x and y

(2) $f_{Y|x}(y) = f_Y(y)$ for all x and y with $f_X(x) > 0$

(3) $f_{X|y}(x) = f_X(x)$ for all x and y with $f_Y(y) > 0$

(4) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ for any sets A and B in the range of X and Y, respectively. (5-21)

EXAMPLE 5-19

For the joint distribution of times in Example 5-15, the

- Marginal distribution of Y was determined in Example 5-16.
- Conditional distribution of Y given $X = x$ was determined in Example 5-17.

Because the marginal and conditional probability densities are not the same for all values of x, property (2) of Equation 5-20 implies that the random variables are not independent. The fact that these variables are not independent can be determined quickly by noticing that the range of (X, Y), shown in Fig. 5-8, is not rectangular. Consequently, knowledge of X changes the interval of values for Y that receives nonzero probability.

EXAMPLE 5-20 Suppose that Example 5-15 is modified so that the joint probability density function of X and Y is $f_{XY}(x, y) = 2 \times 10^{-6} e^{-0.001x - 0.002y}$ for $x \geq 0$ and $y \geq 0$. Show that X and Y are independent and determine $P(X > 1000, Y < 1000)$.

The marginal probability density function of X is

$$f_X(x) = \int_0^\infty 2 \times 10^{-6} e^{-0.001x - 0.002y}\, dy = 0.001 e^{-0.001x} \quad \text{for } x > 0$$

The marginal probability density function of y is

$$f_Y(y) = \int_0^\infty 2 \times 10^{-6} e^{-0.001x - 0.002y}\, dx = 0.002 e^{-0.002y} \quad \text{for } y > 0$$

Therefore, $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y and X and Y are independent.

To determine the probability requested, property (4) of Equation 5-21 and the fact that each random variable has an exponential distribution can be applied.

$$P(X > 1000, Y < 1000) = P(X > 1000)P(Y < 1000) = e^{-1}(1 - e^{-2}) = 0.318$$

Often, based on knowledge of the system under study, random variables are assumed to be independent. Then, probabilities involving both variables can be determined from the marginal probability distributions.

EXAMPLE 5-21 Let the random variables X and Y denote the lengths of two dimensions of a machined part, respectively. Assume that X and Y are independent random variables, and further assume that the distribution of X is normal with mean 10.5 millimeters and variance 0.0025 (millimeter)^2 and that the distribution of Y is normal with mean 3.2 millimeters and variance 0.0036 (millimeter)^2. Determine the probability that $10.4 < X < 10.6$ and $3.15 < Y < 3.25$.

Because X and Y are independent,

$$P(10.4 < X < 10.6, 3.15 < Y < 3.25) = P(10.4 < X < 10.6)P(3.15 < Y < 3.25)$$
$$= P\left(\frac{10.4 - 10.5}{0.05} < Z < \frac{10.6 - 10.5}{0.05}\right) P\left(\frac{3.15 - 3.2}{0.06} < Z < \frac{3.25 - 3.2}{0.06}\right)$$
$$= P(-2 < Z < 2)P(-0.833 < Z < 0.833) = 0.566$$

where Z denotes a standard normal random variable.

EXERCISES FOR SECTION 5-3

5-34. Determine the value of c such that the function $f(x, y) = cxy$ for $0 < x < 3$ and $0 < y < 3$ satisfies the properties of a joint probability density function.

5-35. Continuation of Exercise 5-34. Determine the following:

(a) $P(X < 2, Y < 3)$ (b) $P(X < 2.5)$
(c) $P(1 < Y < 2.5)$ (d) $P(X > 1.8, 1 < Y < 2.5)$
(e) $E(X)$ (f) $P(X < 0, Y < 4)$

5-36. Continuation of Exercise 5-34. Determine the following:
(a) Marginal probability distribution of the random variable X
(b) Conditional probability distribution of Y given that $X = 1.5$
(c) $E(Y|X) = 1.5)$
(d) $P(Y < 2|X = 1.5)$
(e) Conditional probability distribution of X given that $Y = 2$

5-37. Determine the value of c that makes the function $f(x, y) = c(x + y)$ a joint probability density function over the range $0 < x < 3$ and $x < y < x + 2$.

5-38. Continuation of Exercise 5-37. Determine the following:
(a) $P(X < 1, Y < 2)$ (b) $P(1 < X < 2)$
(c) $P(Y > 1)$ (d) $P(X < 2, Y < 2)$
(e) $E(X)$

5-39. Continuation of Exercise 5-37. Determine the following:
(a) Marginal probability distribution of X
(b) Conditional probability distribution of Y given that $X = 1$
(c) $E(Y|X = 1)$
(d) $P(Y > 2|X = 1)$
(e) Conditional probability distribution of X given that $Y = 2$

5-40. Determine the value of c that makes the function $f(x, y) = cxy$ a joint probability density function over the range $0 < x < 3$ and $0 < y < x$.

5-41. Continuation of Exercise 5-40. Determine the following:
(a) $P(X < 1, Y < 2)$ (b) $P(1 < X < 2)$
(c) $P(Y > 1)$ (d) $P(X < 2, Y < 2)$
(e) $E(X)$ (f) $E(Y)$

5-42. Continuation of Exercise 5-40. Determine the following:
(a) Marginal probability distribution of X
(b) Conditional probability distribution of Y given $X = 1$
(c) $E(Y|X = 1)$
(d) $P(Y > 2|X = 1)$
(e) Conditional probability distribution of X given $Y = 2$

5-43. Determine the value of c that makes the function $f(x, y) = ce^{-2x-3y}$ a joint probability density function over the range $0 < x$ and $0 < y < x$.

5-44. Continuation of Exercise 5-43. Determine the following:
(a) $P(X < 1, Y < 2)$ (b) $P(1 < X < 2)$
(c) $P(Y > 3)$ (d) $P(X < 2, Y < 2)$
(e) $E(X)$ (f) $E(Y)$

5-45. Continuation of Exercise 5-43. Determine the following:
(a) Marginal probability distribution of X
(b) Conditional probability distribution of Y given $X = 1$
(c) $E(Y|X = 1)$
(d) Conditional probability distribution of X given $Y = 2$

5-46. Determine the value of c that makes the function $f(x, y) = ce^{-2x-3y}$ a joint probability density function over the range $0 < x$ and $x < y$.

5-47. Continuation of Exercise 5-46. Determine the following:
(a) $P(X < 1, Y < 2)$ (b) $P(1 < X < 2)$
(c) $P(Y > 3)$ (d) $P(X < 2, Y < 2)$
(e) $E(X)$ (f) $E(Y)$

5-48. Continuation of Exercise 5-46. Determine the following:
(a) Marginal probability distribution of X
(b) Conditional probability distribution of Y given $X = 1$
(c) $E(Y|X = 1)$
(d) $P(Y < 2|X = 1)$
(e) Conditional probability distribution of X given $Y = 2$

5-49. Two methods of measuring surface smoothness are used to evaluate a paper product. The measurements are recorded as deviations from the nominal surface smoothness in coded units. The joint probability distribution of the two measurements is a uniform distribution over the region $0 < x < 4$, $0 < y$, and $x - 1 < y < x + 1$. That is, $f_{XY}(x, y) = c$ for x and y in the region. Determine the value for c such that $f_{XY}(x, y)$ is a joint probability density function.

5-50. Continuation of Exercise 5-49. Determine the following:
(a) $P(X < 0.5, Y < 0.5)$ (b) $P(X < 0.5)$
(c) $E(X)$ (d) $E(Y)$

5-51. Continuation of Exercise 5-49. Determine the following:
(a) Marginal probability distribution of X
(b) Conditional probability distribution of Y given $X = 1$
(c) $E(Y|X = 1)$
(d) $P(Y < 0.5|X = 1)$

5-52. The time between surface finish problems in a galvanizing process is exponentially distributed with a mean of 40 hours. A single plant operates three galvanizing lines that are assumed to operate independently.
(a) What is the probability that none of the lines experiences a surface finish problem in 40 hours of operation?
(b) What is the probability that all three lines experience a surface finish problem between 20 and 40 hours of operation?
(c) Why is the joint probability density function not needed to answer the previous questions?

5-53. A popular clothing manufacturer receives Internet orders via two different routing systems. The time between orders for each routing system in a typical day is known to be exponentially distributed with a mean of 3.2 minutes. Both systems operate independently.
(a) What is the probability that no orders will be received in a 5 minute period? In a 10 minute period?
(b) What is the probability that both systems receive two orders between 10 and 15 minutes after the site is officially open for business?

(c) Why is the joint probability distribution not needed to answer the previous questions?

5-54. The conditional probability distribution of Y given $X = x$ is $f_{Y|x}(y) = xe^{-xy}$ for $y > 0$ and the marginal probability distribution of X is a continuous uniform distribution over 0 to 10.

(a) Graph $f_{Y|X}(y) = xe^{-xy}$ for $y > 0$ for several values of x. Determine

(b) $P(Y < 2|X = 2)$ (c) $E(Y|X = 2)$

(d) $E(Y|X = x)$ (e) $f_{XY}(x, y)$

(f) $f_Y(y)$

5-4 MULTIPLE CONTINUOUS RANDOM VARIABLES

As for discrete random variables, in some cases, more than two continuous random variables are defined in a random experiment.

EXAMPLE 5-22 Many dimensions of a machined part are routinely measured during production. Let the random variables, X_1, X_2, X_3, and X_4 denote the lengths of four dimensions of a part. Then, at least four random variables are of interest in this study.

The joint probability distribution of continuous random variables, $X_1, X_2, X_3 \ldots, X_p$ can be specified by providing a method of calculating the probability that $X_1, X_2, X_3, \ldots, X_p$ assume a value in a region R of p-dimensional space. A joint probability density function $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$ is used to determine the probability that $(X_1, X_2, X_3, \ldots, X_p)$ assume a value in a region R by the multiple integral of $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$ over the region R.

Definition

A **joint probability density function** for the continuous random variables $X_1, X_2, X_3, \ldots, X_p$, denoted as $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$, satisfies the following properties:

(1) $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p) \geq 0$

(2) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p = 1$

(3) For any region B of p-dimensional space

$$P[(X_1, X_2, \ldots, X_p) \in B] = \iint_B \cdots \int f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p \quad (5\text{-}22)$$

Typically, $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$ is defined over all of p-dimensional space by assuming that $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p) = 0$ for all points for which $f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)$ is not specified.

EXAMPLE 5-23 In an electronic assembly, let the random variables X_1, X_2, X_3, X_4 denote the lifetimes of four components in hours. Suppose that the joint probability density function of these variables is

$$f_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4) = 9 \times 10^{-2} e^{-0.001x_1 - 0.002x_2 - 0.0015x_3 - 0.003x_4}$$
$$\text{for } x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0$$

What is the probability that the device operates for more than 1000 hours without any failures?

The requested probability is $P(X_1 > 1000, X_2 > 1000, X_3 > 1000, X_4 > 1000)$, which equals the multiple integral of $f_{X_1X_2X_3X_4}(x_1, x_2, x_3, x_4)$ over the region $x_1 > 1000$, $x_2 > 1000$, $x_3 > 1000$, $x_4 > 1000$. The joint probability density function can be written as a product of exponential functions, and each integral is the simple integral of an exponential function. Therefore,

$$P(X_1 > 1000, X_2 > 1000, X_3 > 1000, X_4 > 1000) = e^{-1-2-1.5-3} = 0.00055$$

Suppose that the joint probability density function of several continuous random variables is a constant, say c over a region R (and zero elsewhere). In this special case,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p = c \times (\text{volume of region } R) = 1$$

by property (2) of Equation 5-22. Therefore, $c = 1/\text{volume}\ (R)$. Furthermore, by property (3) of Equation 5-22.

$$\begin{aligned} &P[(X_1, X_2, \ldots, X_p) \in B] \\ &= \int\int_B \cdots \int f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p = c \times \text{volume}\ (B \cap R) \\ &= \frac{\text{volume}\ (B \cap R)}{\text{volume}\ (R)} \end{aligned}$$

When the joint probability density function is constant, the probability that the random variables assume a value in the region B is just the ratio of the volume of the region $B \cap R$ to the volume of the region R for which the probability is positive.

EXAMPLE 5-24

Suppose the joint probability density function of the continuous random variables X and Y is constant over the region $x^2 + y^2 \le 4$. Determine the probability that $X^2 + Y^2 \le 1$.

The region that receives positive probability is a circle of radius 2. Therefore, the area of this region is 4π. The area of the region $x^2 + y^2 \le 1$ is π. Consequently, the requested probability is $\pi/(4\pi) = 1/4$.

Definition

If the joint probability density function of continuous random variables $X_1, X_2, \ldots, X_p$ is $f_{X_1X_2\ldots X_p}(x_1, x_2 \ldots, x_p)$ the **marginal probability density function** of X_i is

$$f_{X_i}(x_i) = \int\int_{R_{x_i}} \cdots \int f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_{i-1}\, dx_{i+1} \ldots dx_p \qquad (5\text{-}23)$$

where R_{x_i} denotes the set of all points in the range of $X_1, X_2, \ldots, X_p$ for which $X_i = x_i$.

As for two random variables, a probability involving only one random variable, say, for example $P(a < X_i < b)$, can be determined from the marginal probability distribution of X_i or from the joint probability distribution of $X_1, X_2, \ldots, X_p$. That is,

$$P(a < X_i < b) = P(-\infty < X_1 < \infty, \ldots, -\infty < X_{i-1} < \infty, a < X_i < b,$$
$$-\infty < X_{i+1} < \infty, \ldots, -\infty < X_p < \infty)$$

Furthermore, $E(X_i)$ and $V(X_i)$, for $i = 1, 2, \ldots, p$, can be determined from the marginal probability distribution of X_i or from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows.

Mean and Variance from Joint Distribution

$$E(X_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_i f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p$$

and (5-24)

$$V(X_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (x_i - \mu_{X_i})^2 f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p$$

The probability distribution of a subset of variables such as $X_1, X_2, \ldots, X_k$, $k < p$, can be obtained from the joint probability distribution of $X_1, X_2, X_3, \ldots, X_p$ as follows.

Distribution of a Subset of Random Variables

If the joint probability density function of continuous random variables $X_1, X_2, \ldots, X_p$ is $f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)$, the **probability density function** of $X_1, X_2, \ldots, X_k$, $k < p$, is

$$f_{X_1 X_2 \ldots X_k}(x_1, x_2, \ldots, x_k)$$
$$= \int\int_{R_{x_1 x_2 \ldots x_k}}\cdots\int f_{X_1 X_2 \ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_{k+1}\, dx_{k+2} \ldots dx_p \qquad (5\text{-}25)$$

where $R_{x_1 x_2 \ldots x_k}$ denotes the set of all points in the range of $X_1, X_2, \ldots, X_k$ for which $X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k$.

Conditional Probability Distribution

Conditional probability distributions can be developed for multiple continuous random variables by an extension of the ideas used for two continuous random variables.

$$f_{X_1 X_2 X_3 | x_4 x_5}(x_1, x_2, x_3) = \frac{f_{X_1 X_2 X_3 X_4 X_5}(x_1, x_2, x_3, x_4, x_5)}{f_{X_4 X_5}(x_4, x_5)}$$

for $f_{X_4 X_5}(x_4, x_5) > 0$.

The concept of independence can be extended to multiple continuous random variables.

Definition

Continuous random variables $X_1, X_2, \ldots, X_p$ are **independent** if and only if

$$f_{X_1 X_2 \ldots X_p}(x_1, x_2 \ldots, x_p) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_p}(x_p) \quad \text{for } all\ x_1, x_2, \ldots, x_p \tag{5-26}$$

Similar to the result for only two random variables, independence implies that Equation 5-26 holds for *all* $x_1, x_2, \ldots, x_p$. If we find one point for which the equality fails, $X_1, X_2, \ldots, X_p$ are not independent. It is left as an exercise to show that if $X_1, X_2, \ldots, X_p$ are independent,

$$P(X_1 \in A_1, X_2 \in A_2, \ldots, X_p \in A_p) = P(X_1 \in A_1) P(X_2 \in A_2) \ldots P(X_p \in A_p)$$

for *any* regions $A_1, A_2, \ldots, A_p$ in the range of $X_1, X_2, \ldots, X_p$, respectively.

EXAMPLE 5-25

In Chapter 3, we showed that a negative binomial random variable with parameters p and r can be represented as a sum of r geometric random variables $X_1, X_2, \ldots, X_r$. Each geometric random variable represents the additional trials required to obtain the next success. Because the trials in a binomial experiment are independent, $X_1, X_2, \ldots, X_r$ are independent random variables.

EXAMPLE 5-26

Suppose X_1, X_2, and X_3 represent the thickness in micrometers of a substrate, an active layer, and a coating layer of a chemical product. Assume that X_1, X_2, and X_3 are independent and normally distributed with $\mu_1 = 10000$, $\mu_2 = 1000$, $\mu_3 = 80$, $\sigma_1 = 250$, $\sigma_2 = 20$, and $\sigma_3 = 4$, respectively. The specifications for the thickness of the substrate, active layer, and coating layer are $9200 < x_1 < 10800$, $950 < x_2 < 1050$, and $75 < x_3 < 85$, respectively. What proportion of chemical products meets all thickness specifications? Which one of the three thicknesses has the least probability of meeting specifications?

The requested probability is $P(9200 < X_1 < 10800, 950 < X_2 < 1050, 75 < X_3 < 85$. Because the random variables are independent,

$$\begin{aligned} &P(9200 < X_1 < 10800, 950 < X_2 < 1050, 75 < X_3 < 85) \\ &= P(9200 < X_1 < 10800) P(950 < X_2 < 1050) P(75 < X_3 < 85) \end{aligned}$$

After standardizing, the above equals

$$P(-3.2 < Z < 3.2) P(-2.5 < Z < 2.5) P(-1.25 < Z < 1.25)$$

where Z is a standard normal random variable. From the table of the standard normal distribution, the above equals

$$(0.99862)(0.98758)(0.78870) = 0.7778$$

The thickness of the coating layer has the least probability of meeting specifications. Consequently, a priority should be to reduce variability in this part of the process.

EXERCISES FOR SECTION 5–4

5-55. Suppose the random variables X, Y, and Z have the joint probability density function $f(x, y, z) = 8xyz$ for $0 < x < 1$, $0 < y < 1$, and $0 < z < 1$. Determine the following:
(a) $P(X < 0.5)$ (b) $P(X < 0.5, Y < 0.5)$
(c) $P(Z < 2)$ (d) $P(X < 0.5 \text{ or } Z < 2)$
(e) $E(X)$

5-56. Continuation of Exercise 5-55. Determine the following:
(a) $P(X < 0.5 \mid Y = 0.5)$
(b) $P(X < 0.5, Y < 0.5 \mid Z = 0.8)$

5-57. Continuation of Exercise 5-55. Determine the following:
(a) Conditional probability distribution of X given that $Y = 0.5$ and $Z = 0.8$
(b) $P(X < 0.5 \mid Y = 0.5, Z = 0.8)$

5-58. Suppose the random variables X, Y, and Z have the joint probability density function $f_{XYZ}(x, y, z) = c$ over the cylinder $x^2 + y^2 < 4$ and $0 < z < 4$. Determine the following.
(a) The constant c so that $f_{XYZ}(x, y, z)$ is a probability density function
(b) $P(X^2 + Y^2 < 2)$
(c) $P(Z < 2)$
(d) $E(X)$

5-59. Continuation of Exercise 5-58. Determine the following:
(a) $P(X < 1 \mid Y = 1)$ (b) $P(X^2 + Y^2 < 1 \mid Z = 1)$

5-60. Continuation of Exercise 5-58. Determine the conditional probability distribution of Z given that $X = 1$ and $Y = 1$.

5-61. Determine the value of c that makes $f_{XYZ}(x, y, z) = c$ a joint probability density function over the region $x > 0$, $y > 0$, $z > 0$, and $x + y + z < 1$.

5-62. Continuation of Exercise 5-61. Determine the following:
(a) $P(X < 0.5, Y < 0.5, Z < 0.5)$
(b) $P(X < 0.5, Y < 0.5)$
(c) $P(X < 0.5)$
(d) $E(X)$

5-63. Continuation of Exercise 5-61. Determine the following:
(a) Marginal distribution of X
(b) Joint distribution of X and Y
(c) Conditional probability distribution of X given that $Y = 0.5$ and $Z = 0.5$
(d) Conditional probability distribution of X given that $Y = 0.5$

5-64. The yield in pounds from a day's production is normally distributed with a mean of 1500 pounds and standard deviation of 100 pounds. Assume that the yields on different days are independent random variables.
(a) What is the probability that the production yield exceeds 1400 pounds on each of five days next week?
(b) What is the probability that the production yield exceeds 1400 pounds on at least four of the five days next week?

5-65. The weights of adobe bricks used for construction are normally distributed with a mean of 3 pounds and a standard deviation of 0.25 pound. Assume that the weights of the bricks are independent and that a random sample of 20 bricks is selected.
(a) What is the probability that all the bricks in the sample exceed 2.75 pounds?
(b) What is the probability that the heaviest brick in the sample exceeds 3.75 pounds?

5-66. A manufacturer of electroluminescent lamps knows that the amount of luminescent ink deposited on one of its products is normally distributed with a mean of 1.2 grams and a standard deviation of 0.03 grams. Any lamp with less than 1.14 grams of luminescent ink will fail to meet customer's specifications. A random sample of 25 lamps is collected and the mass of luminescent ink on each is measured.
(a) What is the probability that at least 1 lamp fails to meet specifications?
(b) What is the probability that 5 lamps or fewer fail to meet specifications?
(c) What is the probability that all lamps conform to specifications?
(d) Why is the joint probability distribution of the 25 lamps not needed to answer the previous questions?

5-5 COVARIANCE AND CORRELATION

When two or more random variables are defined on a probability space, it is useful to describe how they vary together; that is, it is useful to measure the relationship between the variables. A common measure of the relationship between two random variables is the **covariance.** To define the covariance, we need to describe the expected value of a function of two random variables $h(X, Y)$. The definition simply extends that used for a function of a single random variable.

Definition

$$E[h(X, Y)] = \begin{cases} \displaystyle\sum\sum_R h(x, y) f_{XY}(x, y) & X, Y \text{ discrete} \\ \displaystyle\iint_R h(x, y) f_{XY}(x, y)\, dx\, dy & X, Y \text{ continuous} \end{cases} \tag{5-27}$$

That is, $E[h(X, Y)]$ can be thought of as the weighted average of $h(x, y)$ for each point in the range of (X, Y). The value of $E[h(X, Y)]$ represents the average value of $h(X, Y)$ that is expected in a long sequence of repeated trials of the random experiment.

EXAMPLE 5-27 For the joint probability distribution of the two random variables in Fig. 5-12, calculate $E[(X - \mu_X)(Y - \mu_Y)]$.

The result is obtained by multiplying $x - \mu_X$ times $y - \mu_Y$, times $f_{XY}(x, y)$ for each point in the range of (X, Y). First, μ_X and μ_Y are determined from Equation 5-3 as

$$\mu_X = 1 \times 0.3 + 3 \times 0.7 = 2.4$$

and

$$\mu_Y = 1 \times 0.3 + 2 \times 0.4 + 3 \times 0.3 = 2.0$$

Therefore,

$$\begin{aligned} E[(X - \mu_X)(Y - \mu_Y)] = {} & (1 - 2.4)(1 - 2.0) \times 0.1 \\ & + (1 - 2.4)(2 - 2.0) \times 0.2 + (3 - 2.4)(1 - 2.0) \times 0.2 \\ & + (3 - 2.4)(2 - 2.0) \times 0.2 + (3 - 2.4)(3 - 2.0) \times 0.3 = 0.2 \end{aligned}$$

The covariance is defined for both continuous and discrete random variables by the same formula.

Definition

The **covariance** between the random variables X and Y, denoted as $\text{cov}(X, Y)$ or σ_{XY}, is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y \tag{5-28}$$

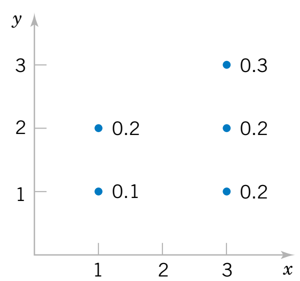

Figure 5-12 Joint distribution of X and Y for Example 5-27.

If the points in the joint probability distribution of X and Y that receive positive probability tend to fall along a line of positive (or negative) slope, σ_{XY} is positive (or negative). If the points tend to fall along a line of positive slope, X tends to be greater than μ_X when Y is greater than μ_Y. Therefore, the product of the two terms $x - \mu_X$ and $y - \mu_Y$ tends to be positive. However, if the points tend to fall along a line of negative slope, $x - \mu_X$ tends to be positive when $y - \mu_Y$ is negative, and vice versa. Therefore, the product of $x - \mu_X$ and $y - \mu_Y$ tends to be negative. In this sense, the covariance between X and Y describes the variation between the two random variables. Figure 5-13 shows examples of pairs of random variables with positive, negative, and zero covariance.

Covariance is a measure of **linear relationship** between the random variables. If the relationship between the random variables is nonlinear, the covariance might not be sensitive to the relationship. This is illustrated in Fig. 5-13(d). The only points with nonzero probability are the points on the circle. There is an identifiable relationship between the variables. Still, the covariance is zero.

The equality of the two expressions for covariance in Equation 5-28 is shown for continuous random variables as follows. By writing the expectations as integrals,

$$
\begin{aligned}
E[(Y-\mu_Y)(X-\mu_X)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)\, f_{XY}(x,y)\, dx\, dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [xy - \mu_X y - x\mu_Y + \mu_X\mu_Y]\, f_{XY}(x,y)\, dx\, dy
\end{aligned}
$$

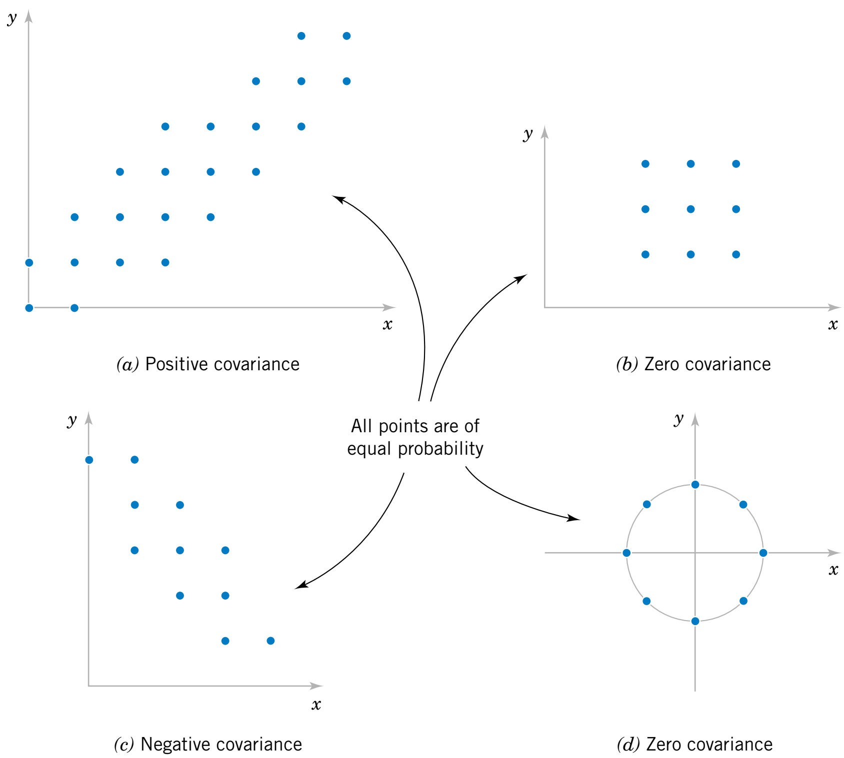

Figure 5-13 Joint probability distributions and the sign of covariance between X and Y.

Now

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mu_X y\, f_{XY}(x, y)\, dx\, dy = \mu_X \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{XY}(x, y)\, dx\, dy \right] = \mu_X \mu_Y$$

Therefore,

$$\begin{aligned} E[(X - \mu_X)(Y - \mu_Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x, y)\, dx\, dy - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x, y)\, dx\, dy - \mu_X\mu_Y = E(XY) - \mu_X\mu_Y \end{aligned}$$

EXAMPLE 5-28 In Example 5-1, the random variables X and Y are the number of acceptable and suspect bits among four bits received during a digital communication, respectively. Is the covariance between X and Y positive or negative?

Because X and Y are the number of acceptable and suspect bits out of the four received, $X + Y \leq 4$. If X is near 4, Y must be near 0. Therefore, X and Y have a negative covariance. This can be verified from the joint probability distribution in Fig. 5-1.

There is another measure of the relationship between two random variables that is often easier to interpret than the covariance.

Definition

The **correlation** between random variables X and Y, denoted as ρ_{XY}, is

$$\rho_{XY} = \frac{\operatorname{cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{5-29}$$

Because $\sigma_X > 0$ and $\sigma_Y > 0$, if the covariance between X and Y is positive, negative, or zero, the correlation between X and Y is positive, negative, or zero, respectively. The following result can be shown.

For any two random variables X and Y

$$-1 \leq \rho_{XY} \leq +1 \tag{5-30}$$

The correlation just scales the covariance by the standard deviation of each variable. Consequently, the correlation is a dimensionless quantity that can be used to compare the linear relationships between pairs of variables in different units.

If the points in the joint probability distribution of X and Y that receive positive probability tend to fall along a line of positive (or negative) slope, ρ_{XY} is near $+1$ (or -1). If ρ_{XY} equals $+1$ or -1, it can be shown that the points in the joint probability distribution that receive positive probability fall exactly along a straight line. Two random variables with nonzero correlation are said to be **correlated.** Similar to covariance, the correlation is a measure of the **linear relationship** between random variables.

EXAMPLE 5-29

For the discrete random variables X and Y with the joint distribution shown in Fig. 5-14, determine σ_{XY} and ρ_{XY}.

The calculations for $E(XY)$, $E(X)$, and $V(X)$ are as follows.

$$\begin{aligned} E(XY) &= 0 \times 0 \times 0.2 + 1 \times 1 \times 0.1 + 1 \times 2 \times 0.1 + 2 \times 1 \times 0.1 \\ &\quad + 2 \times 2 \times 0.1 + 3 \times 3 \times 0.4 = 4.5 \\ E(X) &= 0 \times 0.2 + 1 \times 0.2 + 2 \times 0.2 + 3 \times 0.4 = 1.8 \\ V(X) &= (0 - 1.8)^2 \times 0.2 + (1 - 1.8)^2 \times 0.2 + (2 - 1.8)^2 \times 0.2 \\ &\quad + (3 - 1.8)^2 \times 0.4 = 1.36 \end{aligned}$$

Because the marginal probability distribution of Y is the same as for X, $E(Y) = 1.8$ and $V(Y) = 1.36$. Consequently,

$$\sigma_{XY} = E(XY) - E(X)E(Y) = 4.5 - (1.8)(1.8) = 1.26$$

Furthermore,

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{1.26}{(\sqrt{1.36})(\sqrt{1.36})} = 0.926$$

EXAMPLE 5-30

Suppose that the random variable X has the following distribution: $P(X = 1) = 0.2$, $P(X = 2) = 0.6$, $P(X = 3) = 0.2$. Let $Y = 2X + 5$. That is, $P(Y = 7) = 0.2$, $P(Y = 9) = 0.6$, $P(Y = 11) = 0.2$. Determine the correlation between X and Y. Refer to Fig. 5-15.

Because X and Y are linearly related, $\rho = 1$. This can be verified by direct calculations: Try it.

For independent random variables, we do not expect any relationship in their joint probability distribution. The following result is left as an exercise.

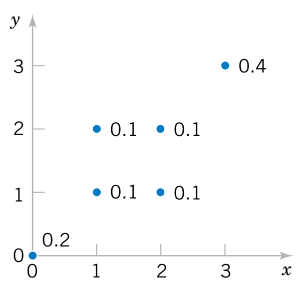

Figure 5-14 Joint distribution for Example 5-29.

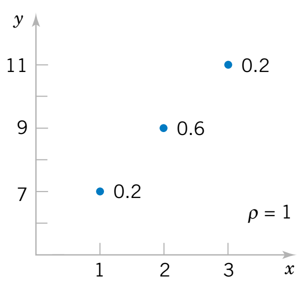

Figure 5-15 Joint distribution for Example 5-30.

If X and Y are independent random variables,

$$\sigma_{XY} = \rho_{XY} = 0 \tag{5-31}$$

EXAMPLE 5-31 For the two random variables in Fig. 5-16, show that $\sigma_{XY} = 0$.

The two random variables in this example are continuous random variables. In this case $E(XY)$ is defined as the double integral over the range of (X, Y). That is,

$$E(XY) = \int_0^4\int_0^2 xyf_{XY}(x,y)\,dx\,dy = \frac{1}{16}\int_0^4\left[\int_0^2 x^2y^2\,dx\right]dy = \frac{1}{16}\int_0^4 y^2\left[x^3/3\Big|_0^2\right]$$

$$= \frac{1}{16}\int_0^4 y^2[8/3]\,dy = \frac{1}{6}\left[y^3/3\Big|_0^4\right] = \frac{1}{6}[64/3] = 32/9$$

Also,

$$E(X) = \int_0^4\int_0^2 xf_{XY}(x,y)\,dx\,dy = \frac{1}{16}\int_0^4\left[\int_0^2 x^2\,dx\right]dy = \frac{1}{16}\int_0^4\left[x^3/3\Big|_0^2\right]dy$$

$$= \frac{1}{16}\left[y^2/2\Big|_0^4\right][8/3] = \frac{1}{6}[16/2] = 4/3$$

$$E(Y) = \int_0^4\int_0^2 yf_{XY}(x,y)\,dx\,dy = \frac{1}{16}\int_0^4 y^2\left[\int_0^2 x\,dx\right]dy = \frac{1}{16}\int_0^4 y^2\left[x^2/2\Big|_0^2\right]dy$$

$$= \frac{2}{16}\left[y^3/3\Big|_0^4\right] = \frac{1}{8}[64/3] = 8/3$$

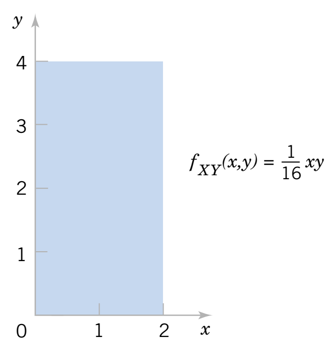

Figure 5-16 Random variables with zero covariance from Example 5-31.

Thus,

$$E(XY) - E(X)E(Y) = 32/9 - (4/3)(8/3) = 0$$

It can be shown that these two random variables are independent. You can check that $f_{XY}(x, y) = f_X(x) f_Y(y)$ for all x and y.

However, if the correlation between two random variables is zero, we *cannot* immediately conclude that the random variables are independent. Figure 5-13(d) provides an example.

EXERCISES FOR SECTION 5-5

5-67. Determine the covariance and correlation for the following joint probability distribution:

x	1	1	2	4
y	3	4	5	6
$f_{XY}(x, y)$	1/8	1/4	1/2	1/8

5-68. Determine the covariance and correlation for the following joint probability distribution:

x	-1	-0.5	0.5	1
y	-2	-1	1	2
$f_{XY}(x, y)$	1/8	1/4	1/2	1/8

5-69. Determine the value for c and the covariance and correlation for the joint probability mass function $f_{XY}(x, y) = c(x + y)$ for $x = 1, 2, 3$ and $y = 1, 2, 3$.

5-70. Determine the covariance and correlation for the joint probability distribution shown in Fig. 5-4(a) and described in Example 5-8.

5-71. Determine the covariance and correlation for X_1 and X_2 in the joint distribution of the multinomial random variables X_1, X_2 and X_3 in with $p_1 = p_2 = p_3 = \frac{1}{3}$ and $n = 3$. What can you conclude about the sign of the correlation between two random variables in a multinomial distribution?

5-72. Determine the value for c and the covariance and correlation for the joint probability density function $f_{XY}(x, y) = cxy$ over the range $0 < x < 3$ and $0 < y < x$.

5-73. Determine the value for c and the covariance and correlation for the joint probability density function $f_{XY}(x, y) = c$ over the range $0 < x < 5$, $0 < y$, and $x - 1 < y < x + 1$.

5-74. Determine the covariance and correlation for the joint probability density function $f_{XY}(x, y) = 6 \times 10^{-6} e^{-0.001x - 0.002y}$ over the range $0 < x$ and $x < y$ from Example 5-15.

5-75. Determine the covariance and correlation for the joint probability density function $f_{XY}(x, y) = e^{-x-y}$ over the range $0 < x$ and $0 < y$.

5-76. Suppose that the correlation between X and Y is ρ. For constants a, b, c, and d, what is the correlation between the random variables $U = aX + b$ and $V = cY + d$?

5-77. The joint probability distribution is

x	-1	0	0	1
y	0	-1	1	0
$f_{XY}(x, y)$	1/4	1/4	1/4	1/4

Show that the correlation between X and Y is zero, but X and Y are not independent.

5-78. Suppose X and Y are independent continuous random variables. Show that $\sigma_{XY} = 0$.

5-6 BIVARIATE NORMAL DISTRIBUTION

An extension of a normal distribution to two random variables is an important bivariate probability distribution.

EXAMPLE 5-32

At the start of this chapter, the length of different dimensions of an injection-molded part was presented as an example of two random variables. Each length might be modeled by a normal distribution. However, because the measurements are from the same part, the random variables are typically not independent. A probability distribution for two normal random variables that are not independent is important in many applications. As stated at the start of the chapter, if the specifications for X and Y are 2.95 to 3.05 and 7.60 to 7.80 millimeters, respectively, we might be interested in the probability that a part satisfies both specifications; that is, $P(2.95 < X < 3.05, 7.60 < Y < 7.80)$.

Definition

The probability density function of a **bivariate normal distribution** is

$$f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\} \quad (5\text{-}32)$$

for $-\infty < x < \infty$ and $-\infty < y < \infty$, with parameters $\sigma_X > 0$, $\sigma_Y > 0$, $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, and $-1 < \rho < 1$.

The result that $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$ integrates to 1 is left as an exercise. Also, the bivariate normal probability density function is positive over the entire plane of real numbers.

Two examples of bivariate normal distributions are illustrated in Fig. 5-17 along with corresponding contour plots. Each curve on the contour plots is a set of points for which the probability density function is constant. As seen in the contour plots, the bivariate normal probability density function is constant on ellipses in the (x, y) plane. (We can consider a circle to be a special case of an ellipse.) The center of each ellipse is at the point (μ_X, μ_Y). If $\rho > 0$ ($\rho < 0$), the major axis of each ellipse has positive (negative) slope, respectively. If $\rho = 0$, the major axis of the ellipse is aligned with either the x or y coordinate axis.

EXAMPLE 5-33

The joint probability density function $f_{XY}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-0.5(x^2+y^2)}$ is a special case of a bivariate normal distribution with $\sigma_X = 1$, $\sigma_Y = 1$, $\mu_X = 0$, $\mu_Y = 0$, and $\rho = 0$. This probability density function is illustrated in Fig. 5-18. Notice that the contour plot consists of concentric circles about the origin.

By completing the square in the exponent, the following results can be shown. The details are left as an exercise.

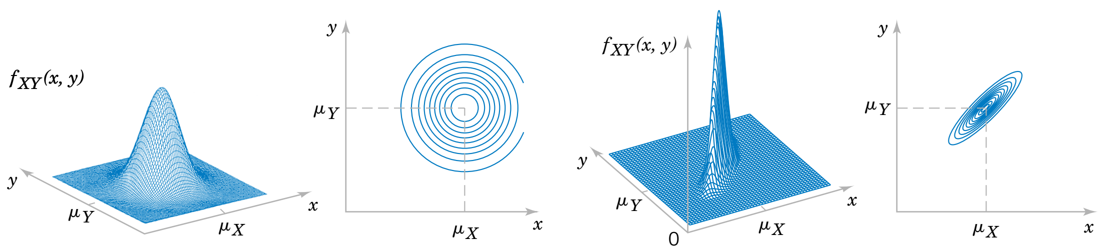

Figure 5-17 Examples of bivariate normal distributions.

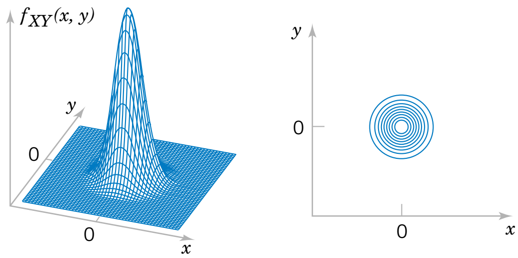

Figure 5-18 Bivariate normal probability density function with $\sigma_X = 1$, $\sigma_Y = 1$, $\rho = 0$, $\mu_X = 0$, and $\mu_Y = 0$.

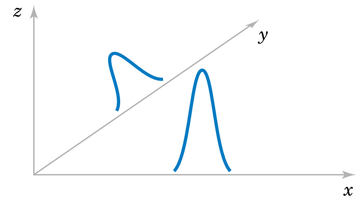

Figure 5-19 Marginal probability density functions of a bivariate normal distribution.

Marginal Distributions of Bivariate Normal Random Variables

If X and Y have a bivariate normal distribution with joint probability density $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$, the **marginal probability distributions** of X and Y are normal with means μ_X and μ_Y and standard deviations σ_X and σ_Y, respectively. (5-33)

Figure 5-19 illustrates that the marginal probability distributions of X and Y are normal. Furthermore, as the notation suggests, ρ represents the correlation between X and Y. The following result is left as an exercise.

If X and Y have a bivariate normal distribution with joint probability density function $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$, the correlation between X and Y is ρ. (5-34)

The contour plots in Fig. 5-17 illustrate that as ρ moves from zero (left graph) to 0.9 (right graph), the ellipses narrow around the major axis. The probability is more concentrated about a line in the (x, y) plane and graphically displays greater correlation between the variables. If $\rho = -1$ or $+1$, all the probability is concentrated on a line in the (x, y) plane. That is, the probability that X and Y assume a value that is not on the line is zero. In this case, the bivariate normal probability density is not defined.

In general, zero correlation does not imply independence. But in the special case that X and Y have a bivariate normal distribution, if $\rho = 0$, X and Y are independent. The details are left as an exercise.

If X and Y have a bivariate normal distribution with $\rho = 0$, X and Y are independent. (5-35)

An important use of the bivariate normal distribution is to calculate probabilities involving two correlated normal random variables.

EXAMPLE 5-34

Suppose that the X and Y dimensions of an injection-molded part have a bivariate normal distribution with $\sigma_X = 0.04$, $\sigma_Y = 0.08$. $\mu_X = 3.00$. $\mu_Y = 7.70$, and $\rho = 0.8$. Then, the probability that a part satisfies both specifications is

$$P(2.95 < X < 3.05, 7.60 < Y < 7.80)$$

This probability can be obtained by integrating $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X\ \mu_Y, \rho)$ over the region $2.95 < x < 3.05$ and $7.60 < y < 7.80$, as shown in Fig. 5-7. Unfortunately, there is often no closed-form solution to probabilities involving bivariate normal distributions. In this case, the integration must be done numerically.

EXERCISES FOR SECTION 5-6

5-79. Let X and Y represent concentration and viscosity of a chemical product. Suppose X and Y have a bivariate normal distribution with $\sigma_X = 4$, $\sigma_Y = 1$, $\mu_X = 2$, and $\mu_Y = 1$. Draw a rough contour plot of the joint probability density function for each of the following values for ρ:
(a) $\rho = 0$ (b) $\rho = 0.8$
(c) $\rho = -0.8$

5-80. Let X and Y represent two dimensions of an injection molded part. Suppose X and Y have a bivariate normal distribution with $\sigma_X = 0.04$, $\sigma_Y = 0.08$, $\mu_X = 3.00$, $\mu_Y = 7.70$, and $\rho_Y = 0$. Determine $P(2.95 < X < 3.05, 7.60 < Y < 7.80)$.

5-81. In the manufacture of electroluminescent lamps, several different layers of ink are deposited onto a plastic substrate. The thickness of these layers is critical if specifications regarding the final color and intensity of light of the lamp are to be met. Let X and Y denote the thickness of two different layers of ink. It is known that X is normally distributed with a mean of 0.1 millimeter and a standard deviation of 0.00031 millimeter, and Y is also normally distributed with a mean of 0.23 millimeter and a standard deviation of 0.00017 millimeter. The value of ρ for these variables is equal to zero. Specifications call for a lamp to have a thickness of the ink corresponding to X in the range of 0.099535 to 0.100465 millimeters and Y in the range of 0.22966 to 0.23034 millimeters. What is the probability that a randomly selected lamp will conform to specifications?

5-82. Suppose that X and Y have a bivariate normal distribution with joint probability density function $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$.
(a) Show that the conditional distribution of Y, given that $X = x$ is normal.
(b) Determine $E(Y|X = x)$.
(c) Determine $V(Y|X = x)$.

5-83. If X and Y have a bivariate normal distribution with $\rho = 0$, show that X and Y are independent.

5-84. Show that the probability density function $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$ of a bivariate normal distribution integrates to one. [*Hint:* Complete the square in the exponent and use the fact that the integral of a normal probability density function for a single variable is 1.]

5-85. If X and Y have a bivariate normal distribution with joint probability density $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$, show that the marginal probability distribution of X is normal with mean μ_X and standard deviation σ_X. [*Hint:* Complete the square in the exponent and use the fact that the integral of a normal probability density function for a single variable is 1.]

5-86. If X and Y have a bivariate normal distribution with joint probability density $f_{XY}(x, y; \sigma_X, \sigma_Y, \mu_X, \mu_Y, \rho)$, show that the correlation between X and Y is ρ. [*Hint:* Complete the square in the exponent].

5-7 LINEAR COMBINATIONS OF RANDOM VARIABLES

A random variable is sometimes defined as a function of one or more random variables. The CD material presents methods to determine the distributions of general functions of random variables. Furthermore, moment-generating functions are introduced on the CD and used to determine the distribution of a sum of random variables. In this section, results for linear functions are highlighted because of their importance in the remainder of the book. References are made to the CD material as needed. For example, if the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part, $Y = 2X_1 + 2X_2$ is a random variable that represents the perimeter of the part. As another example, recall that the negative binomial random variable was represented as the sum of several geometric random variables.

In this section, we develop results for random variables that are linear combinations of random variables.

Definition

Given random variables $X_1, X_2, \ldots, X_p$ and constants $c_1, c_2, \ldots, c_p$,

$$Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p \tag{5-36}$$

is a **linear combination** of $X_1, X_2, \ldots, X_p$.

Now, $E(Y)$ can be found from the joint probability distribution of $X_1, X_2, \ldots, X_p$ as follows. Assume $X_1, X_2, \ldots, X_p$ are continuous random variables. An analogous calculation can be used for discrete random variables.

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} (c_1x_1 + c_2x_2 + \cdots + c_px_p) f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p \\
&= c_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1 f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p \\
&\quad + c_2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_2 f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p + , \ldots , \\
&\quad + c_p \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_p f_{X_1X_2\ldots X_p}(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p
\end{aligned}$$

By using Equation 5-24 for each of the terms in this expression, we obtain the following.

Mean of a Linear Combination

If $Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$,

$$E(Y) = c_1E(X_1) + c_2E(X_2) + \cdots + c_pE(X_p) \tag{5-37}$$

Furthermore, it is left as an exercise to show the following.

Variance of a Linear Combination

If $X_1, X_2, \ldots, X_p$ are random variables, and $Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$, then in general

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) + 2\sum_{i<j}\sum c_ic_j \operatorname{cov}(X_i, X_j) \quad (5\text{-}38)$$

If $X_1, X_2, \ldots, X_p$ are **independent,**

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) \quad (5\text{-}39)$$

Note that the result for the variance in Equation 5-39 requires the random variables to be independent. To see why the independence is important, consider the following simple example. Let X_1 denote any random variable and define $X_2 = -X_1$. Clearly, X_1 and X_2 are not independent. In fact, $\rho_{XY} = -1$. Now, $Y = X_1 + X_2$ is 0 with probability 1. Therefore, $V(Y) = 0$, regardless of the variances of X_1 and X_2.

EXAMPLE 5-35 In Chapter 3, we found that if Y is a negative binomial random variable with parameters p and r, $Y = X_1 + X_2 + \cdots + X_r$, where each X_i is a geometric random variable with parameter p and they are independent. Therefore, $E(X_i) = 1/p$ and $E(X_i) = (1 - p)/p^2$. From Equation 5-37, $E(Y) = r/p$ and from Equation 5-39, $V(Y) = r(1 - p)/p^2$.

An approach similar to the one applied in the above example can be used to verify the formulas for the mean and variance of an Erlang random variable in Chapter 4.

EXAMPLE 5-36 Suppose the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part. Assume $E(X_1) = 2$ centimeters with standard deviation 0.1 centimeter and $E(X_2) = 5$ centimeters with standard deviation 0.2 centimeter. Also, assume that the covariance between X_1 and X_2 is -0.005. Then, $Y = 2X_1 + 2X_2$ is a random variable that represents the perimeter of the part. From Equation 5-36,

$$E(Y) = 2(2) + 2(5) = 14 \text{ centimeters}$$

and from Equation 5-38

$$\begin{aligned} V(Y) &= 2^2(0.1^2) + 2^2(0.2^2) + 2 \times 2 \times 2(-0.005) \\ &= 0.04 + 0.16 - 0.04 = 0.16 \text{ centimeters squared} \end{aligned}$$

Therefore, the standard deviation of Y is $0.16^{1/2} = 0.4$ centimeters.

The particular linear combination that represents the average of p random variables, with identical means and variances, is used quite often in the subsequent chapters. We highlight the results for this special case.

Mean and Variance of an Average

If $\overline{X} = (X_1 + X_2 + \cdots + X_p)/p$ with $E(X_i) = \mu$ for $i = 1, 2, \ldots, p$

$$E(\overline{X}) = \mu \tag{5-40a}$$

if $X_1, X_2, \ldots, X_p$ are also independent with $V(X_i) = \sigma^2$ for $i = 1, 2, \ldots, p$,

$$V(\overline{X}) = \frac{\sigma^2}{p} \tag{5-40b}$$

The conclusion for $V(\overline{X})$ is obtained as follows. Using Equation 5-39, with $c_i = 1/p$ and $V(X_i) = \sigma^2$, yields

$$V(\overline{X}) = \underbrace{(1/p)^2\sigma^2 + \cdots + (1/p)^2\sigma^2}_{p \text{ terms}} = \sigma^2/p$$

Another useful result concerning linear combinations of random variables is a **reproductive property** that holds for independent, normal random variables.

Reproductive Property of the Normal Distribution

If $X_1, X_2, \ldots, X_p$ are independent, normal random variables with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$, for $i = 1, 2, \ldots, p$,

$$Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$$

is a normal random variable with

$$E(Y) = c_1\mu_1 + c_2\mu_2 + \cdots + c_p\mu_p$$

and

$$V(Y) = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_p^2\sigma_p^2 \tag{5-41}$$

The mean and variance of Y follow from Equations 5-37 and 5-39. The fact that Y has a normal distribution can be obtained from moment-generating functions discussed in Section 5-9 in the CD material.

EXAMPLE 5-37

Let the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part. Assume that X_1 is normal with $E(X_1) = 2$ centimeters and standard deviation 0.1 centimeter and that X_2 is normal with $E(X_2) = 5$ centimeters and standard deviation 0.2 centimeter. Also, assume that X_1 and X_2 are independent. Determine the probability that the perimeter exceeds 14.5 centimeters.

Then, $Y = 2X_1 + 2X_2$ is a normal random variable that represents the perimeter of the part. We obtain, $E(Y) = 14$ centimeters and the variance of Y is

$$V(Y) = 4 \times 0.1^2 + 4 \times 0.2^2 = 0.0416$$

Now,

$$\begin{aligned} P(Y > 14.5) &= P[(Y - \mu_Y)/\sigma_Y > (14.5 - 14)/\sqrt{0.0416}] \\ &= P(Z > 1.12) = 0.13 \end{aligned}$$

EXAMPLE 5-38 Soft-drink cans are filled by an automated filling machine. The mean fill volume is 12.1 fluid ounces, and the standard deviation is 0.1 fluid ounce. Assume that the fill volumes of the cans are independent, normal random variables. What is the probability that the average volume of 10 cans selected from this process is less than 12 fluid ounces?

Let $X_1, X_2, \ldots, X_{10}$ denote the fill volumes of the 10 cans. The average fill volume (denoted as $\overline{X}$) is a normal random variable with

$$E(\overline{X}) = 12.1 \quad \text{and} \quad V(\overline{X}) = \frac{0.1^2}{10} = 0.001$$

Consequently,

$$\begin{aligned} P(\overline{X} < 12) &= P\left[\frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{12 - 12.1}{\sqrt{0.001}}\right] \\ &= P(Z < -3.16) = 0.00079 \end{aligned}$$

EXERCISES FOR SECTION 5-7

5-87. If X and Y are independent, normal random variables with $E(X) = 0$, $V(X) = 4$, $E(Y) = 10$, and $V(Y) = 9$. Determine the following:
(a) $E(2X + 3Y)$ (b) $V(2X + 3Y)$
(c) $P(2X + 3Y < 30)$ (d) $P(2X + 3Y < 40)$

5-88. Suppose that the random variable X represents the length of a punched part in centimeters. Let Y be the length of the part in millimeters. If $E(X) = 5$ and $V(X) = 0.25$, what are the mean and variance of Y?

5-89. A plastic casing for a magnetic disk is composed of two halves. The thickness of each half is normally distributed with a mean of 2 millimeters and a standard deviation of 0.1 millimeter and the halves are independent.
(a) Determine the mean and standard deviation of the total thickness of the two halves.
(b) What is the probability that the total thickness exceeds 4.3 millimeters?

5-90. In the manufacture of electroluminescent lamps, several different layers of ink are deposited onto a plastic substrate. The thickness of these layers is critical if specifications regarding the final color and intensity of light of the lamp are to be met. Let X and Y denote the thickness of two different layers of ink. It is known that X is normally distributed with a mean of 0.1 millimeter and a standard deviation of 0.00031 millimeter and Y is also normally distributed with a mean of 0.23 millimeter and a standard deviation of 0.00017 millimeter. Assume that these variables are independent.
(a) If a particular lamp is made up of these two inks only, what is the probability that the total ink thickness is less than 0.2337 millimeter?
(b) A lamp with a total ink thickness exceeding 0.2405 millimeters lacks the uniformity of color demanded by the customer. Find the probability that a randomly selected lamp fails to meet customer specifications.

5-91. The width of a casing for a door is normally distributed with a mean of 24 inches and a standard deviation of 1/8 inch. The width of a door is normally distributed with a mean of 23 and 7/8 inches and a standard deviation of 1/16 inch. Assume independence.
(a) Determine the mean and standard deviation of the difference between the width of the casing and the width of the door.

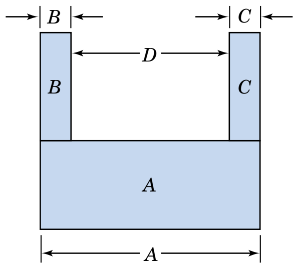

Figure 5-20 Figure for the U-shaped component.

(b) What is the probability that the width of the casing minus the width of the door exceeds 1/4 inch?
(c) What is the probability that the door does not fit in the casing?

5-92. A U-shaped component is to be formed from the three parts A, B, and C. The picture is shown in Fig. 5-20. The length of A is normally distributed with a mean of 10 millimeters and a standard deviation of 0.1 millimeter. The thickness of parts B and C is normally distributed with a mean of 2 millimeters and a standard deviation of 0.05 millimeter. Assume all dimensions are independent.
(a) Determine the mean and standard deviation of the length of the gap D.
(b) What is the probability that the gap D is less than 5.9 millimeters?

5-93. Soft-drink cans are filled by an automated filling machine and the standard deviation is 0.5 fluid ounce. Assume that the fill volumes of the cans are independent, normal random variables.
(a) What is the standard deviation of the average fill volume of 100 cans?
(b) If the mean fill volume is 12.1 ounces, what is the probability that the average fill volume of the 100 cans is below 12 fluid ounces?
(c) What should the mean fill volume equal so that the probability that the average of 100 cans is below 12 fluid ounces is 0.005?
(d) If the mean fill volume is 12.1 fluid ounces, what should the standard deviation of fill volume equal so that the probability that the average of 100 cans is below 12 fluid ounces is 0.005?
(e) Determine the number of cans that need to be measured such that the probability that the average fill volume is less than 12 fluid ounces is 0.01.

5-94. The photoresist thickness in semiconductor manufacturing has a mean of 10 micrometers and a standard deviation of 1 micrometer. Assume that the thickness is normally distributed and that the thicknesses of different wafers are independent.
(a) Determine the probability that the average thickness of 10 wafers is either greater than 11 or less than 9 micrometers.
(b) Determine the number of wafers that needs to be measured such that the probability that the average thickness exceeds 11 micrometers is 0.01.
(c) If the mean thickness is 10 micrometers, what should the standard deviation of thickness equal so that the probability that the average of 10 wafers is either greater than 11 or less than 9 micrometers is 0.001?

5-95. Assume that the weights of individuals are independent and normally distributed with a mean of 160 pounds and a standard deviation of 30 pounds. Suppose that 25 people squeeze into an elevator that is designed to hold 4300 pounds.
(a) What is the probability that the load (total weight) exceeds the design limit?
(b) What design limit is exceeded by 25 occupants with probability 0.0001?

5-8 FUNCTIONS OF RANDOM VARIABLES (CD ONLY)

5-9 MOMENT GENERATING FUNCTION (CD ONLY)

5-10 CHEBYSHEV'S INEQUALITY (CD ONLY)

Supplemental Exercises

5-96. Show that the following function satisfies the properties of a joint probability mass function:

x	y	$f(x, y)$
0	0	1/4
0	1	1/8
1	0	1/8
1	1	1/4
2	2	1/4

5-97. Continuation of Exercise 5-96. Determine the following probabilities:
(a) $P(X < 0.5, Y < 1.5)$ (b) $P(X \leq 1)$
(c) $P(X < 1.5)$ (d) $P(X > 0.5, Y < 1.5)$
(e) Determine $E(X)$, $E(Y)$, $V(X)$, and $V(Y)$.

5-98. Continuation of Exercise 5-96. Determine the following:
(a) Marginal probability distribution of the random variable X
(b) Conditional probability distribution of Y given that $X = 1$
(c) $E(Y|X = 1)$
(d) Are X and Y independent? Why or why not?
(e) Calculate the correlation between X and Y.

5-99. The percentage of people given an antirheumatoid medication who suffer severe, moderate, or minor side effects are 10, 20, and 70%, respectively. Assume that people react independently and that 20 people are given the medication. Determine the following:

(a) The probability that 2, 4, and 14 people will suffer severe, moderate, or minor side effects, respectively
(b) The probability that no one will suffer severe side effects
(c) The mean and variance of the number of people that will suffer severe side effects
(d) What is the conditional probability distribution of the number of people who suffer severe side effects given that 19 suffer minor side effects?
(e) What is the conditional mean of the number of people who suffer severe side effects given that 19 suffer minor side effects?

5-100. The backoff torque required to remove bolts in a steel plate is rated as high, moderate, or low. Historically, the probability of a high, moderate, or low rating is 0.6, 0.3, or 0.1, respectively. Suppose that 20 bolts are evaluated and that the torque ratings are independent.

(a) What is the probability that 12, 6, and 2 bolts are rated as high, moderate, and low, respectively?
(b) What is the marginal distribution of the number of bolts rated low?
(c) What is the expected number of bolts rated low?
(d) What is the probability that the number of bolts rated low is greater than two?

5-101. Continuation of Exercise 5-100

(a) What is the conditional distribution of the number of bolts rated low given that 16 bolts are rated high?
(b) What is the conditional expected number of bolts rated low given that 16 bolts are rated high?
(c) Are the numbers of bolts rated high and low independent random variables?

5-102. To evaluate the technical support from a computer manufacturer, the number of rings before a call is answered by a service representative is tracked. Historically, 70% of the calls are answered in two rings or less, 25% are answered in three or four rings, and the remaining calls require five rings or more. Suppose you call this manufacturer 10 times and assume that the calls are independent.

(a) What is the probability that eight calls are answered in two rings or less, one call is answered in three or four rings, and one call requires five rings or more?
(b) What is the probability that all 10 calls are answered in four rings or less?
(c) What is the expected number of calls answered in four rings or less?

5-103. Continuation of Exercise 5-102

(a) What is the conditional distribution of the number of calls requiring five rings or more given that eight calls are answered in two rings or less?
(b) What is the conditional expected number of calls requiring five rings or more given that eight calls are answered in two rings or less?
(c) Are the number of calls answered in two rings or less and the number of calls requiring five rings or more independent random variables?

5-104. Determine the value of c such that the function $f(x, y) = cx^2y$ for $0 < x < 3$ and $0 < y < 2$ satisfies the properties of a joint probability density function.

5-105. Continuation of Exercise 5-104. Determine the following:

(a) $P(X < 1, Y < 1)$ (b) $P(X < 2.5)$
(c) $P(1 < Y < 2.5)$ (d) $P(X > 2, 1 < Y < 1.5)$
(e) $E(X)$ (f) $E(Y)$

5-106. Continuation of Exercise 5-104.

(a) Determine the marginal probability distribution of the random variable X.
(b) Determine the conditional probability distribution of Y given that $X = 1$.
(c) Determine the conditional probability distribution of X given that $Y = 1$.

5-107. The joint distribution of the continuous random variables X, Y, and Z is constant over the region $x^2 + y^2 \le 1$, $0 < z < 4$.

(a) Determine $P(X^2 + Y^2 \le 0.5)$
(b) Determine $P(X^2 + Y^2 \le 0.5, Z < 2)$
(c) What is the joint conditional probability density function of X and Y given that $Z = 1$?
(d) What is the marginal probability density function of X?

5-108. Continuation of Exercise 5-107.

(a) Determine the conditional mean of Z given that $X = 0$ and $Y = 0$.
(b) In general, determine the conditional mean of Z given that $X = x$ and $Y = y$.

5-109. Suppose that X and Y are independent, continuous uniform random variables for $0 < x < 1$ and $0 < y < 1$. Use the joint probability density function to determine the probability that $|X - Y| < 0.5$.

5-110. The lifetimes of six major components in a copier are independent exponential random variables with means of 8000, 10,000, 10,000, 20,000, 20,000, and 25,000 hours, respectively.

(a) What is the probability that the lifetimes of all the components exceed 5000 hours?
(b) What is the probability that at least one component lifetime exceeds 25,000 hours?

5-111. Contamination problems in semiconductor manufacturing can result in a functional defect, a minor defect, or no defect in the final product. Suppose that 20, 50, and 30% of the contamination problems result in functional, minor, and no defects, respectively. Assume that the effects of 10 contamination problems are independent.

(a) What is the probability that the 10 contamination problems result in two functional defects and five minor defects?
(b) What is the distribution of the number of contamination problems that result in no defects?

(c) What is the expected number of contamination problems that result in no defects?

5-112. The weight of adobe bricks for construction is normally distributed with a mean of 3 pounds and a standard deviation of 0.25 pound. Assume that the weights of the bricks are independent and that a random sample of 25 bricks is chosen.

(a) What is the probability that the mean weight of the sample is less than 2.95 pounds?

(b) What value will the mean weight exceed with probability 0.99?

5-113. The length and width of panels used for interior doors (in inches) are denoted as X and Y, respectively. Suppose that X and Y are independent, continuous uniform random variables for $17.75 < x < 18.25$ and $4.75 < y < 5.25$, respectively.

(a) By integrating the joint probability density function over the appropriate region, determine the probability that the area of a panel exceeds 90 squared inches.

(b) What is the probability that the perimeter of a panel exceeds 46 inches?

5-114. The weight of a small candy is normally distributed with a mean of 0.1 ounce and a standard deviation of 0.01 ounce. Suppose that 16 candies are placed in a package and that the weights are independent.

(a) What are the mean and variance of package net weight?

(b) What is the probability that the net weight of a package is less than 1.6 ounces?

(c) If 17 candies are placed in each package, what is the probability that the net weight of a package is less than 1.6 ounces?

5-115. The time for an automated system in a warehouse to locate a part is normally distributed with a mean of 45 seconds and a standard deviation of 30 seconds. Suppose that independent requests are made for 10 parts.

(a) What is the probability that the average time to locate 10 parts exceeds 60 seconds?

(b) What is the probability that the total time to locate 10 parts exceeds 600 seconds?

5-116. A mechanical assembly used in an automobile engine contains four major components. The weights of the components are independent and normally distributed with the following means and standard deviations (in ounces):

Component	Mean	Standard Deviation
Left case	4	0.4
Right case	5.5	0.5
Bearing assembly	10	0.2
Bolt assembly	8	0.5

(a) What is the probability that the weight of an assembly exceeds 29.5 ounces?

(b) What is the probability that the mean weight of eight independent assemblies exceeds 29 ounces?

5-117. Suppose X and Y have a bivariate normal distribution with $\sigma_X = 4$, $\sigma_Y = 1$, $\mu_X = 4$, $\mu_Y = 4$, and $\rho = -0.2$. Draw a rough contour plot of the joint probability density function.

5-118. If
$$f_{XY}(x, y) = \frac{1}{1.2\pi} \exp\left\{\frac{-1}{0.72}\left[(x-1)^2 - 1.6(x-1)(y-2) + (y-2)^2\right]\right\}$$
determine $E(X)$, $E(Y)$, $V(X)$, $V(Y)$, and ρ by recorganizing the parameters in the joint probability density function.

5-119. The permeability of a membrane used as a moisture barrier in a biological application depends on the thickness of two integrated layers. The layers are normally distributed with means of 0.5 and 1 millimeters, respectively. The standard deviations of layer thickness are 0.1 and 0.2 millimeters, respectively. The correlation between layers is 0.7.

(a) Determine the mean and variance of the total thickness of the two layers.

(b) What is the probability that the total thickness is less than 1 millimeter?

(c) Let X_1 and X_2 denote the thickness of layers 1 and 2, respectively. A measure of performance of the membrane is a function $2X_1 + 3X_2$ of the thickness. Determine the mean and variance of this performance measure.

5-120. The permeability of a membrane used as a moisture barrier in a biological application depends on the thickness of three integrated layers. Layers 1, 2, and 3 are normally distributed with means of 0.5, 1, and 1.5 millimeters, respectively. The standard deviations of layer thickness are 0.1, 0.2, and 0.3, respectively. Also, the correlation between layers 1 and 2 is 0.7, between layers 2 and 3 is 0.5, and between layers 1 and 3 is 0.3.

(a) Determine the mean and variance of the total thickness of the three layers.

(b) What is the probability that the total thickness is less than 1.5 millimeters?

5-121. A small company is to decide what investments to use for cash generated from operations. Each investment has a mean and standard deviation associated with the percentage gain. The first security has a mean percentage gain of 5% with a standard deviation of 2%, and the second security provides the same mean of 5% with a standard deviation of 4%. The securities have a correlation of −0.5, so there is a negative correlation between the percentage returns. If the company invests two million dollars with half in each security, what is the mean and standard deviation of the percentage return? Compare the standard deviation of this strategy to one that invests the two million dollars into the first security only.

MIND-EXPANDING EXERCISES

5-122. Show that if $X_1, X_2, \ldots, X_p$ are independent, continuous random variables, $P(X_1 \in A_1, X_2 \in A_2, \ldots, X_p \in A_p) = P(X_1 \in A_1)P(X_2 \in A_2)\ldots P(X_p \in A_p)$ for any regions $A_1, A_2, \ldots, A_p$ in the range of $X_1, X_2, \ldots, X_p$ respectively.

5-123. Show that if $X_1, X_2, \ldots, X_p$ are independent random variables and $Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$,

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p)$$

You can assume that the random variables are continuous.

5-124. Suppose that the joint probability function of the continuous random variables X and Y is constant on the rectangle $0 < x < a$, $0 < y < b$. Show that X and Y are independent.

5-125. Suppose that the range of the continuous variables X and Y is $0 < x < a$ and $0 < y < b$. Also suppose that the joint probability density function $f_{XY}(x, y) = g(x)h(y)$, where $g(x)$ is a function only of x and $h(y)$ is a function only of y. Show that X and Y are independent.

IMPORTANT TERMS AND CONCEPTS

In the E-book, click on any term or concept below to go to that subject.

Bivariate normal distribution
Conditional mean
Conditional probability density function
Conditional probability mass function
Conditional variance
Contour plots
Correlation
Covariance
Independence
Joint probability density function
Joint probability mass function
Linear combinations of random variables
Marginal probability distribution
Multinomial distribution
Reproductive property of the normal distribution

CD MATERIAL

Convolution
Functions of random variables
Jacobian of a transformation

Moment generating function
Uniqueness property of moment generating function
Chebyshev's inequality

5-8 FUNCTIONS OF RANDOM VARIABLES (CD ONLY)

In many situations in statistics, it is necessary to derive the probability distribution of a function of one or more random variables. In this section, we present some results that are helpful in solving this problem.

Suppose that X is a discrete random variable with probability distribution $f_X(x)$. Let $Y = h(X)$ be a function of X that defines a one-to-one transformation between the values of X and Y, and we wish to find the probability distribution of Y. By a one-to-one transformation, we mean that each value x is related to one and only one value of $y = h(x)$ and that each value of y is related to one and only one value of x, say, $x = u(y)$, where $u(y)$ is found by solving $y = h(x)$ for x in terms of y.

Now, the random variable Y takes on the value y when X takes on the value $u(y)$. Therefore, the probability distribution of Y is

$$f_Y(y) = P(Y = y) = P[X = u(y)] = f_X[u(y)]$$

We may state this result as follows.

> Suppose that X is a **discrete** random variable with probability distribution $f_X(x)$. Let $Y = h(X)$ define a one-to-one transformation between the values of X and Y so that the equation $y = h(x)$ can be solved uniquely for x in terms of y. Let this solution be $x = u(y)$. Then the probability distribution of the random variable Y is
>
> $$f_Y(y) = f_X[u(y)] \tag{S5-1}$$

EXAMPLE S5-1 Let X be a geometric random variable with probability distribution

$$f_X(x) = p(1-p)^{x-1}, \qquad x = 1, 2, \ldots$$

Find the probability distribution of $Y = X^2$.

Since $X \geq 0$, the transformation is one to one; that is, $y = x^2$ and $x = \sqrt{y}$. Therefore, Equation S5-1 indicates that the distribution of the random variable Y is

$$f_Y(y) = f(\sqrt{y}) = p(1-p)^{\sqrt{y}-1}, \qquad y = 1, 4, 9, 16, \ldots$$

Now suppose that we have two discrete random variables X_1 and X_2 with joint probability distribution $f_{X_1X_2}(x_1, x_2)$ and we wish to find the joint probability distribution $f_{Y_1Y_2}(y_1, y_2)$ of two new random variables $Y_1 = h_1(X_1, X_2)$ and $Y_2 = h_2(X_1, X_2)$. We assume that the functions h_1 and h_2 define a one-to-one transformation between (x_1, x_2) and (y_1, y_2). Solving the equations $y_1 = h_1(x_1, x_2)$ and $y_2 = h_2(x_1, x_2)$ simultaneously, we obtain the unique solution $x_1 = u_1(y_1, y_2)$ and $x_2 = u_2(y_1, y_2)$. Therefore, the random variables Y_1 and Y_2 take on the values y_1 and y_2 when X_1 takes on the value $u_1(y_1, y_2)$ and X_2 takes the value $u_2(y_1, y_2)$. The joint probability distribution of Y_1 and Y_2 is

$$\begin{aligned} f_{Y_1Y_2}(y_1, y_2) &= P(Y_1 = y_1, Y_2 = y_2) \\ &= P[X_1 = u_1(y_1, y_2), X_2 = u_2(y_1, y_2)] \\ &= f_{X_1X_2}[u_1(y_1, y_2), u_2(y_1, y_2)] \end{aligned}$$

We will also state this result as follows.

Suppose that X_1 and X_2 are **discrete** random variables with joint probability distribution $f_{X_1X_2}(x_1, x_2)$, and let $Y_1 = h_1(X_1, X_2)$ and $Y_2 = h_2(X_1, X_2)$ define one-to-one transformations between the points (x_1, x_2) and (y_1, y_2) so that the equations $y_1 = h_1(x_1, x_2)$ and $y_2 = h_2(x_1, x_2)$ can be solved uniquely for x_1 and x_2 in terms of y_1 and y_2. Let this solution be $x_1 = u_1(y_1, y_2)$ and $x_2 = u_2(y_1, y_2)$. Then the joint probability distribution of Y_1 and Y_2 is

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}[u_1(y_1, y_2), u_2(y_1, y_2)] \tag{S5-2}$$

A very important application of Equation S5-2 is in finding the distribution of a random variable Y_1 that is a function of two other random variables X_1 and X_2. That is, let $Y_1 = h_1(X_1, X_2)$ where X_1 and X_2 are discrete random variables with joint distribution $f_{X_1X_2}(x_1, x_2)$. We want to find the probability distribution of Y_1, say, $f_{Y_1}(y_1)$. To do this, we define a second function $Y_2 = h_2(X_1, X_2)$ so that the one-to-one correspondence between the points (x_1, x_2) and (y_1, y_2) is maintained, and we use the result in Equation S5-2 to find the joint probability distribution of Y_1 and Y_2. Then the distribution of Y_1 alone is found by summing over the y_2 values in this joint distribution. That is, $f_{Y_1}(y_1)$ is just the **marginal probability distribution** of Y_1, or

$$f_{Y_1}(y_1) = \sum_{y_2} f_{Y_1Y_2}(y_1, y_2)$$

EXAMPLE S5-2 Consider the case where X_1 and X_2 are independent Poisson random variables with parameters λ_1 and λ_2, respectively. We will find the distribution of the random variable $Y_1 = X_1 + X_2$.

The joint distribution of X_1 and X_2 is

$$\begin{aligned} f_{X_1X_2}(x_1, x_2) &= f_{X_1}(x_1) f_{X_2}(x_2) \\ &= \frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!} \frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{x_1}\lambda_2^{x_2}}{x_1!\, x_2!}, \qquad x_1 = 0, 1, \ldots, \qquad x_2 = 0, 1, \ldots \end{aligned}$$

because X_1 and X_2 are independent. Now to use Equation S5-2 we need to define a second function $Y_2 = h_2(X_1, X_2)$. Let this function be $Y_2 = X_2$. Now the solutions for x_1 and x_2 are $x_1 = y_1 - y_2$ and $x_2 = y_2$. Thus, from Equation S5-2 the joint probability distribution of Y_1 and Y_2 is

$$f_{Y_1Y_2}(y_1, y_2) = \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{(y_1-y_2)}\lambda_2^{y_2}}{(y_1 - y_2)!\, y_2!}, \qquad y_1 = 0, 1, 2, \ldots, \qquad y_2 = 0, 1, \ldots, y_1$$

Because $x_1 \geq 0$, the transformation $x_1 = y_1 - y_2$ requires that $x_2 = y_2$ must always be less than or equal to y_1. Thus, the values of y_2 are $0, 1, \ldots, y_1$, and the marginal probability distribution of Y_1 is obtained as follows:

$$f_{Y_1}(y_1) = \sum_{y_2=0}^{y_1} f_{Y_1Y_2}(y_1, y_2) = \sum_{y_2=0}^{y_1} \frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{(y_1-y_2)}\lambda_2^{y_2}}{(y_1 - y_2)!\, y_2!}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{y_1!} \sum_{y_2=0}^{y_1} \frac{y_1!\,\lambda_1^{(y_1-y_2)}\lambda_2^{y_2}}{(y_1-y_2)!\,y_2!} = \frac{e^{-(\lambda_1+\lambda_2)}}{y_1!} \sum_{y_2=0}^{y_1} \binom{y_1}{y_2} \lambda_1^{(y_1-y_2)}\lambda_2^{y_2}$$

The summation in this last expression is the binomial expansion of $(\lambda_1 + \lambda_2)^{y_1}$, so

$$f_{Y_1}(y_1) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^{y_1}}{y_1!}, \qquad y_1 = 0, 1, \ldots$$

We recognize this as a Poisson distribution with parameter $\lambda_1 + \lambda_2$. Therefore, we have shown that the sum of two independent Poisson random variables with parameters λ_1 and λ_2 has a Poisson distribution with parameter $\lambda_1 + \lambda_2$.

We now consider the situation where the random variables are continuous. Let $Y = h(X)$, with X continuous and the transformation is one to one.

Suppose that X is a **continuous** random variable with probability distribution $f_X(x)$. The function $Y = h(X)$ is a one-to-one transformation between the values of Y and X so that the equation $y = h(x)$ can be uniquely solved for x in terms of y. Let this solution be $x = u(y)$. The probability distribution of Y is

$$f_Y(y) = f_X[u(y)]\,|J| \tag{S5-3}$$

where $J = u'(y)$ is called the **Jacobian** of the transformation and the absolute value of J is used.

Equation S5-3 is shown as follows. Let the function $y = h(x)$ be an increasing function of x. Now

$$\begin{aligned} P(Y \le a) &= P[X \le u(a)] \\ &= \int_{-\infty}^{u(a)} f_X(x)\,dx \end{aligned}$$

If we change the variable of integration from x to y by using $x = u(y)$, we obtain $dx = u'(y)\,dy$ and then

$$P(Y \le a) = \int_{-\infty}^{a} f_X[u(y)]u'(y)\,dy$$

Since the integral gives the probability that $Y \le a$ for all values of a contained in the feasible set of values for y, $f_X[u(y)]u'(y)$ must be the probability density of Y. Therefore, the probability distribution of Y is

$$f_Y(y) = f_X[u(y)]u'(y) = f_X[u(y)]J$$

If the function $y = h(x)$ is a decreasing function of x, a similar argument holds.

EXAMPLE S5-3 Let X be a continuous random variable with probability distribution

$$f_X(x) = \frac{x}{8}, \quad 0 \le x < 4$$

Find the probability distribution of $Y = h(X) = 2X + 4$.

Note that $y = h(x) = 2x + 4$ is an increasing function of x. The inverse solution is $x = u(y) = (y - 4)/2$, and from this we find the Jacobian to be $J = u'(y) = dx/dy = 1/2$. Therefore, from S5-3 the probability distribution of Y is

$$f_Y(y) = \frac{(y-4)/2}{8}\left(\frac{1}{2}\right) = \frac{y-4}{32}, \qquad 4 \le y \le 12$$

We now consider the case where X_1 and X_2 are continuous random variables and we wish to find the joint probability distribution of $Y_1 = h_1(X_1, X_2)$ and $Y_2 = h_2(X_1, X_2)$ where the transformation is one to one. The application of this will typically be in finding the probability distribution of $Y_1 = h_1(X_1, X_2)$, analogous to the discrete case discussed above. We will need the following result.

Suppose that X_1 and X_2 are **continuous** random variables with joint probability distribution $f_{X_1X_2}(x_1, x_2)$, and let $Y_1 = h_1(X_1, X_2)$ and $Y_2 = h_2(X_1, X_2)$ define a one-to-one transformation between the points (x_1, x_2) and (y_1, y_2). Let the equations $y_1 = h_1(x_1, x_2)$ and $y_2 = h_2(x_1, x_2)$ be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 as $x_1 = u_1(y_1, y_2)$ and $x_2 = u_2(y_1, y_2)$. Then the joint probability of Y_1 and Y_2 is

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}[u_1(y_1, y_2), u_2(y_1, y_2)]\,|J| \tag{S5-4}$$

where J is the *Jacobian* and is given by the following determinant:

$$J = \begin{vmatrix} \partial x_1/\partial y_1, & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1, & \partial x_2/\partial y_2 \end{vmatrix}$$

and the absolute value of the determinant is used.

This result can be used to find $f_{Y_1Y_2}(y_1, y_2)$, the joint probability distribution of Y_1 and Y_2. Then the probability distribution of Y_1 is

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{Y_1Y_2}(y_1, y_2)\, dy_2$$

That is, $f_{Y_1}(y_1)$ is the **marginal probability distribution** of Y_1.

EXAMPLE S5-4 Suppose that X_1 and X_2 are independent exponential random variables with $f_{X_1}(x_1) = 2e^{-2x_1}$ and $f_{X_2}(x_2) = 2e^{-2x_2}$. Find the probability distribution of $Y = X_1/X_2$.

The joint probability distribution of X_1 and X_2 is

$$f_{X_1X_2}(x_1, x_2) = 4e^{-2(x_1+x_2)}, \qquad x_1 \geq 0, \qquad x_2 \geq 0$$

because X_1 and X_2 are independent. Let $Y_1 = h_1(X_1, X_2) = X_1/X_2$ and $Y_2 = h_2(X_1, X_2) = X_1 + X_2$. The inverse solutions of $y_1 = x_1/x_2$ and $y_2 = x_1 + x_2$ are $x_1 = y_1y_2/(1 + y_1)$ and $x_2 = y_2/(1 + y_1)$, and it follows that

$$\frac{\partial x_1}{\partial y_1} = y_2\left[\frac{1}{(1+y_1)^2}\right], \qquad \frac{\partial x_1}{\partial y_2} = \left[\frac{y_1}{(1+y_1)}\right]$$

$$\frac{\partial x_2}{\partial y_1} = y_2\left[\frac{-1}{(1+y_1)^2}\right], \qquad \frac{\partial x_2}{\partial y_2} = \left[\frac{1}{(1+y_1)}\right]$$

Therefore

$$J = \begin{vmatrix} \dfrac{y_2}{(1+y_1)^2} & \dfrac{y_1}{(1+y_1)} \\ \dfrac{-y_2}{(1+y_1)^2} & \dfrac{1}{(1+y_1)} \end{vmatrix} = \frac{y_2}{(1+y_1)^2}$$

and from Equation S5-4 the joint probability distribution of Y_1 and Y_2 is

$$\begin{aligned} f_{Y_1Y_2}(y_1, y_2) &= f_{X_1X_2}[u_1(y_1, y_2), u_2(y_1, y_2)]\,|J| \\ &= 4e^{-2[y_1y_2/(1+y_1)+y_2/(1+y_1)]}\left|\frac{y_2}{(1+y_1)^2}\right| \\ &= 4e^{-2y_2}\,y_2/(1+y_1)^2 \end{aligned}$$

for $y_1 > 0, y_2 > 0$. We need to find the distribution of $Y_1 = X_1/X_2$. This is the marginal probability distribution of Y_1, or

$$\begin{aligned} f_{Y_1}(y_1) &= \int_0^\infty f_{Y_1Y_2}(y_1, y_2)\,dy_2 \\ &= \int_0^\infty 4e^{-2y_2}\,[y_2/(1+y_1)^2]\,dy_2 \\ &= \frac{1}{(1+y_1)^2}, \qquad y_1 > 0 \end{aligned}$$

An important application of Equation S5-4 is to obtain the distribution of the sum of two *independent* random variables X_1 and X_2. Let $Y_1 = X_1 + X_2$ and let $Y_2 = X_2$. The inverse solutions are $x_1 = y_1 - y_2$ and $x_2 = y_2$. Therefore,

$$\frac{\partial x_1}{\partial y_1} = 1 \qquad \frac{\partial x_1}{\partial y_2} = -1$$

$$\frac{\partial x_2}{\partial y_1} = 0 \qquad \frac{\partial x_2}{\partial y_2} = 1$$

and $|J| = 1$. From Equation S5-4, the joint probability density function of Y_1 and Y_2 is

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1}(y_1 - y_2)f_{X_2}(y_2)$$

Therefore, the marginal probability density function of Y_1 is

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{X_1}(y_1 - y_2)f_{X_2}(y_2)\, dy_2$$

The notation is simpler if the variable of integration y_2 is replaced with x and y_1 is replaced with y. Then the following result is obtained.

Convolution of X_1 and X_2

If X_1 and X_2 are independent random variables with probability density functions $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$, respectively, the probability density function of $Y = X_1 + X_2$ is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - x)f_{X_2}(x)\, dx \qquad \text{(S5-5)}$$

The probability density function of Y in Equation S5-5 is referred to as the **convolution** of the probability density functions for X_1 and X_2. This concept is commonly used for transformations (such as Fourier transformations) in mathematics. This integral may be evaluated numerically to obtain the probability density function of Y, even for complex probability density functions for X_1 and X_2. A similar result can be obtained for discrete random variables with the integral replaced with a sum.

In some problems involving transformations, we need to find the probability distribution of the random variable $Y = h(X)$ when X is a continuous random variable, but the transformation is not one to one. The following result is helpful.

Suppose that X is a **continuous** random variable with probability distribution $f_X(x)$, and $Y = h(X)$ is a transformation that is not one to one. If the interval over which X is defined can be partitioned into m mutually exclusive disjoint sets such that each of the inverse functions $x_1 = u_1(y), x_2 = u_2(y), \ldots, x_m = u_m(y)$ of $y = u(x)$ is one to one, the probability distribution of Y is

$$f_Y(y) = \sum_{i=1}^{m} f_X[u_i(y)]\,|J_i| \qquad \text{(S5-6)}$$

where $J_i = u_i'(y)$, $i = 1, 2, \ldots, m$ and the absolute values are used.

To illustrate how this equation is used, suppose that X is a normal random variable with mean μ and variance σ^2, and we wish to show that the distribution of $Y = (X - \mu)^2/\sigma^2$ is a chi-squared distribution with one degree of freedom. Let $Z = (X - \mu)/\sigma$, and $Y = Z^2$. The probability distribution of Z is the standard normal; that is,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty$$

The inverse solutions of $y = z^2$ are $z = \pm\sqrt{y}$, so the transformation is not one to one. Define $z_1 = -\sqrt{y}$ and $z_2 = +\sqrt{y}$ so that $J_1 = -(1/2)/\sqrt{y}$ and $J_2 = (1/2)/\sqrt{y}$. Then by Equation S5-6, the probability distribution of Y is

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-y/2} \left| \frac{-1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-y/2} \left| \frac{1}{2\sqrt{y}} \right| \\ &= \frac{1}{2^{1/2}\sqrt{\pi}} y^{1/2-1} e^{-y/2}, \qquad y > 0 \end{aligned}$$

Now it can be shown that $\sqrt{\pi} = \Gamma(1/2)$, so we may write $f(y)$ as

$$f_Y(y) = \frac{1}{2^{1/2}\,\Gamma\left(\frac{1}{2}\right)} y^{1/2-1} e^{-y/2}, \qquad y > 0$$

which is the chi-squared distribution with 1 degree of freedom.

EXERCISES FOR SECTION 5-8

S5-1. Suppose that X is a random variable with probability distribution

$$f_X(x) = 1/4, \qquad x = 1, 2, 3, 4$$

Find the probability distribution of the random $Y = 2X + 1$.

S5-2. Let X be a binomial random variable with $p = 0.25$ and $n = 3$. Find the probability distribution of the random variable $Y = X^2$.

S5-3. Suppose that X is a continuous random variable with probability distribution

$$f_X(x) = \frac{x}{18}, \qquad 0 \le x \le 6$$

(a) Find the probability distribution of the random variable $Y = 2X + 10$.
(b) Find the expected value of Y.

S5-4. Suppose that X has a uniform probability distribution

$$f_X(x) = 1, \qquad 0 \le x \le 1$$

Show that the probability distribution of the random variable $Y = -2 \ln X$ is chi-squared with two degrees of freedom.

S5-5. A current of I amperes flows through a resistance of R ohms according to the probability distribution

$$f_I(i) = 2i, \qquad 0 \le i \le 1$$

Suppose that the resistance is also a random variable with probability distribution

$$f_R(r) = 1, \qquad 0 \le r \le 1$$

Assume that I and R are independent.
(a) Find the probability distribution for the power (in watts) $P = I^2R$.
(b) Find $E(P)$.

S5-6. A random variable X has the following probability distribution:

$$f_X(x) = e^{-x}, \qquad x \ge 0$$

(a) Find the probability distribution for $Y = X^2$.
(b) Find the probability distribution for $Y = X^{1/2}$.
(c) Find the probability distribution for $Y = \ln X$.

S5-7. The velocity of a particle in a gas is a random variable V with probability distribution

$$f_V(v) = av^2 e^{-bv} \qquad v > 0$$

where b is a constant that depends on the temperature of the gas and the mass of the particle.
(a) Find the value of the constant a.
(b) The kinetic energy of the particle is $W = mV^2/2$. Find the probability distribution of W.

S5-8. Suppose that X has the probability distribution

$$f_X(x) = 1, \qquad 1 \le x \le 2$$

Find the probability distribution of the random variable $Y = e^X$.

S5-9. Prove that Equation S5-3 holds when $y = h(x)$ is a decreasing function of x.

S5-10. The random variable X has the probability distribution

$$f_X(x) = \frac{x}{8}, \qquad 0 \le x \le 4$$

Find the probability distribution of $Y = (X - 2)^2$.

S5-11. Consider a rectangle with sides of length S_1 and S_2, where S_1 and S_2 are independent random variables. The probability distributions of S_1 and S_2 are

$$f_{S_1}(s_1) = 2s_1, \qquad 0 \le s_1 \le 1$$

and

$$f_{S_2}(s_2) = \frac{s_2}{8}, \qquad 0 \le s_2 \le 4$$

(a) Find the joint distribution of the area of the rectangle $A = S_1 S_2$ and the random variable $Y = S_1$.
(b) Find the probability distribution of the area A of the rectangle.

S5-12. Suppose we have a simple electrical circuit in which Ohm's law $V = IR$ holds. We are interested in the probability distribution of the resistance R given that V and I are independent random variables with the following distributions:

$$f_V(v) = e^{-v}, \qquad v \ge 0$$

and

$$f_I(i) = 1, \qquad 1 \le i \le 2$$

Find the probability distribution of R.

5-9 MOMENT GENERATING FUNCTIONS (CD ONLY)

Suppose that X is a random variable with mean μ. Throughout this book we have used the idea of the expected value of the random variable X, and in fact $E(X) = \mu$. Now suppose that we are interested in the expected value of a particular function of X, say, $g(X) = X^r$. The expected value of this function, or $E[g(X)] = E(X^r)$, is called the rth moment about the origin of the random variable X, which we will denote by μ'_r.

Definition

The rth **moment about the origin** of the random variable X is

$$\mu'_r = E(X^r) = \begin{cases} \sum_x x^r f(x), & X \text{ discrete} \\ \int_{-\infty}^{\infty} x^r f(x)\,dx, & X \text{ continuous} \end{cases} \qquad \text{(S5-7)}$$

Notice that the first moment about the origin is just the mean, that is, $E(X) = \mu'_1 = \mu$. Furthermore, since the second moment about the origin is $E(X^2) = \mu'_2$, we can write the variance of a random variable in terms of origin moments as follows:

$$\sigma^2 = E(X^2) - [E(X)]^2 = \mu'_2 - \mu^2$$

The moments of a random variable can often be determined directly from the definition in Equation S5-7, but there is an alternative procedure that is frequently useful that makes use of a special function.

Definition

The **moment generating function** of the random variable X is the expected value of e^{tX} and is denoted by $M_X(t)$. That is,

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} f(x), & X \text{ discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x)\, dx, & X \text{ continuous} \end{cases} \tag{S5-8}$$

The moment generating function $M_X(t)$ will exist only if the sum or integral in the above definition converges. If the moment generating function of a random variable does exist, it can be used to obtain all the origin moments of the random variable.

Let X be a random variable with moment generating function $M_X(t)$. Then

$$\mu'_r = \left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} \tag{S5-9}$$

Assuming that we can differentiate inside the summation and integral signs,

$$\frac{d^r M_X(t)}{dt^r} = \begin{cases} \sum_x x^r e^{tx} f(x), & X \quad \textit{discrete} \\ \int_{-\infty}^{\infty} x^r e^{tx} f(x)\, dx, & X \quad \textit{continuous} \end{cases}$$

Now if we set $t = 0$ in this expression, we find that

$$\left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0} = E(X^r)$$

EXAMPLE S5-5

Suppose that X has a binomial distribution, that is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n$$

Determine the moment generating function and use it to verify that the mean and variance of the binomial random variable are $\mu = np$ and $\sigma^2 = np(1-p)$.

From the definition of a moment generating function, we have

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}$$

This last summation is the binomial expansion of $[pe^t + (1 - p)]^n$, so

$$M_X(t) = [pe^t + (1 - p)]^n$$

Taking the first and second derivatives, we obtain

$$M'_X(t) = \frac{dM_X(t)}{dt} = npe^t[1 + p(e^t - 1)]^{n-1}$$

and

$$M''_X(t) = \frac{d^2M_X(t)}{dt^2} = npe^t(1 - p + npe^t)[1 + p(e^t - 1)]^{n-2}$$

If we set $t = 0$ in $M'_X(t)$, we obtain

$$M'_X(t)\big|_{t=0} = \mu'_1 = \mu = np$$

which is the mean of the binomial random variable X. Now if we set $t = 0$ in $M''_X(t)$,

$$M''_X(t)\,\big|_{\,t=0} = \mu'_2 = np(1 - p + np)$$

Therefore, the variance of the binomial random variable is

$$\sigma^2 = \mu'_2 - \mu^2 = np(1 - p + np) - (np)^2 = np - np^2 = np(1 - p)$$

EXAMPLE S5-6 Find the moment generating function of the normal random variable and use it to show that the mean and variance of this random variable are μ and σ^2, respectively.

The moment generating function is

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sigma\sqrt{2\pi}}\, e^{-(x-\mu)^2/(2\sigma^2)}\, dx$$
$$= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-[x^2-2(\mu+t\sigma^2)x+\mu^2]/(2\sigma^2)}\, dx$$

If we complete the square in the exponent, we have

$$x^2 - 2(\mu + t\sigma^2)x + \mu^2 = [x - (\mu + t\sigma^2)]^2 - 2\mu t\sigma^2 - t^2\sigma^4$$

and then

$$M_X(t) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\{[x-(\mu+t\sigma^2)]^2-2\mu t\sigma^2-t^2\sigma^4\}/(2\sigma^2)}\, dx$$
$$= e^{\mu t+\sigma^2t^2/2}\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(1/2)[x-(\mu+t\sigma^2)]^2/\sigma^2} dx$$

Let $u = [x - (\mu + t\sigma^2)]/\sigma$. Then $dx = \sigma\, du$, and this last expression above becomes

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$$

Now the integral is just the total area under a standard normal density, which is 1, so the moment generating function of a normal random variable is

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}$$

Differentiating this function twice with respect to t and setting $t = 0$ in the result, we find

$$\left.\frac{dM_X(t)}{dt}\right|_{t=0} = \mu_1' = \mu \qquad \text{and} \qquad \left.\frac{d^2 M_X(t)}{dt^2}\right|_{t=0} = \mu_2' = \sigma^2 + \mu^2$$

Therefore, the variance of the normal random variable is

$$\sigma^2 = \mu_2' - \mu^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

Moment generating functions have many important and useful properties. One of the most important of these is the **uniqueness property.** That is, the moment generating function of a random variable is unique when it exists, so if we have two random variables X and Y, say, with moment generating functions $M_X(t)$ and $M_Y(t)$, then if $M_X(t) = M_Y(t)$ for all values of t, both X and Y have the same probability distribution. Some of the other useful properties of the moment generating function are summarized as follows.

Properties of Moment Generating Functions

If X is a random variable and a is a constant, then

(1) $M_{X+a}(t) = e^{at} M_X(t)$

(2) $M_{aX}(t) = M_X(at)$

If $X_1, X_2, \ldots, X_n$ are independent random variables with moment generating functions $M_{X_1}(t), M_{X_2}(t), \ldots, M_{X_n}(t)$, respectively, and if $Y = X_1 + X_2 + \cdots + X_n$, then the moment generating function of Y is

(3) $M_Y(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \cdots \cdot M_{X_n}(t)$ (S5-10)

Property (1) follows from $M_{X+a}(t) = E[e^{t(X+a)}] = e^{at}E(e^{tX}) = e^{at}M_X(t)$. Property (2) follows from $M_{aX}(t) = E[e^{t(aX)}] = E[e^{(at)X}] = M_X(at)$. Consider property (3) for the case where the X's are continuous random variables:

$$\begin{aligned} M_Y(t) &= E(e^{tY}) = E[e^{t(X_1+X_2+\cdots+X_n)}] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{t(x_1+x_2+\cdots+x_n)} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \ldots dx_n \end{aligned}$$

Since the X's are independent,

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \cdots \cdot f_{X_n}(x_n)$$

and one may write

$$\begin{aligned} M_Y(t) &= \int_{-\infty}^{\infty} e^{tx_1} f_{X_1}(x_1)\,dx_1 \int_{-\infty}^{\infty} e^{tx_2} f_{X_2}(x_2)\,dx_2 \cdots \int_{-\infty}^{\infty} e^{tx_n} f_{X_n}(x_n)\,dx_n \\ &= M_{X_1}(t) \cdot M_{X_2}(t) \cdot \cdots \cdot M_{X_n}(t) \end{aligned}$$

For the case when the X's are discrete, we would use the same approach replacing integrals with summations.

Equation S5-10 is particularly useful. In many situations we need to find the distribution of the sum of two or more independent random variables, and often this result makes the problem very easy. This is illustrated in the following example.

EXAMPLE S5-7 Suppose that X_1 and X_2 are two independent Poisson random variables with parameters λ_1 and λ_2, respectively. Find the probability distribution of $Y = X_1 + X_2$.

The moment generating function of a Poisson random variable with parameter λ is

$$M_X(t) = e^{\lambda(e^t - 1)}$$

so the moment generating functions of X_1 and X_2 are $M_{X_1}(t) = e^{\lambda_1(e^t-1)}$ and $M_{X_2}(t) = e^{\lambda_2(e^t-1)}$, respectively. Using Equation S5-10, we find that the moment generating function of $Y = X_1 + X_2$ is

$$M_Y(t) = M_{X_1}(t)M_{X_2}(t) = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

which is recognized as the moment generating function of a Poisson random variable with parameter $\lambda_1 + \lambda_2$. Therefore, we have shown that the sum of two independent Poisson random variables with parameters λ_1 and λ_2 is a Poisson random variable with parameters equal to the sum of the two parameters $\lambda_1 + \lambda_2$.

EXERCISES FOR SECTION 5-9

S5-13. A random variable X has the discrete uniform distribution

$$f(x) = \frac{1}{m}, \qquad x = 1, 2, \ldots, m$$

(a) Show that the moment generating function is

$$M_X(t) = \frac{e^t(1 - e^{tm})}{m(1 - e^t)}$$

(b) Use $M_X(t)$ to find the mean and variance of X.

S5-14. A random variable X has the Poisson distribution

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, \ldots$$

(a) Show that the moment generating function is

$$M_X(t) = e^{\lambda(e^t - 1)}$$

(b) Use $M_X(t)$ to find the mean and variance of the Poisson random variable.

S5-15. The geometric random variable X has probability distribution

$$f(x) = (1 - p)^{x-1}p, \quad x = 1, 2, \ldots$$

(a) Show that the moment generating function is

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t}$$

(b) Use $M_X(t)$ to find the mean and variance of X.

S5-16. The chi-squared random variable with k degrees of freedom has moment generating function $M_X(t) = (1-2t)^{-k/2}$. Suppose that X_1 and X_2 are independent chi-squared random variables with k_1 and k_2 degrees of freedom, respectively. What is the distribution of $Y = X_1 + X_2$?

S5-17. A continuous random variable X has the following probability distribution:

$$f(x) = 4xe^{-2x}, \qquad x > 0$$

(a) Find the moment generating function for X.
(b) Find the mean and variance of X.

S5-18. The continuous uniform random variable X has density function

$$f(x) = \frac{1}{\beta-\alpha}, \qquad \alpha \le x \le \beta$$

(a) Show that the moment generating function is

$$M_X(t) = \frac{e^{t\beta} - e^{t\alpha}}{t(\beta-\alpha)}$$

(b) Use $M_X(t)$ to find the mean and variance of X.

S5-19. A random variable X has the exponential distribution

$$f(x) = \lambda e^{-\lambda x}, \qquad x > 0$$

(a) Show that the moment generating function of X is

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}$$

(b) Find the mean and variance of X.

S5-20. A random variable X has the gamma distribution

$$f(x) = \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x}, \qquad x > 0$$

(a) Show that the moment generating function of X is

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-r}$$

(b) Find the mean and variance of X.

S5-21. Let $X_1, X_2, \ldots, X_r$ be independent exponential random variables with parameter λ.
(a) Find the moment generating function of $Y = X_1 + X_2 + \cdots + X_r$.
(b) What is the distribution of the random variable Y?
[*Hint:* Use the results of Exercise S5-20].

S5-22. Suppose that X_i has a normal distribution with mean μ_i and variance $\sigma_i^2, i = 1, 2$. Let X_1 and X_2 be independent.
(a) Find the moment generating function of $Y = X_1 + X_2$.
(b) What is the distribution of the random variable Y?

S5-23. Show that the moment generating function of the chi-squared random variable with k degrees of freedom is $M_X(t) = (1-2t)^{-k/2}$. Show that the mean and variance of this random variable are k and $2k$, respectively.

S5-24. Continuation of Exercise S5-20.
(a) Show that by expanding e^{tX} in a power series and taking expectations term by term we may write the moment generating function as

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= 1 + \mu_1' t + \mu_2' \frac{t^2}{2!} + \cdots \\ &\quad + \mu_r' \frac{t^r}{r!} + \cdots \end{aligned}$$

Thus, the coefficient of $t^r/r!$ in this expansion is μ_r', the rth origin moment.
(b) Continuation of Exercise S5-20. Write the power series expansion for $M_X(t)$, the gamma random variable.
(c) Continuation of Exercise S5-20. Find μ_1' and μ_2' using the results of parts (a) and (b). Does this approach give the same answers that you found for the mean and variance of the gamma random variable in Exercise S5-20?

5-10 CHEBYSHEV'S INEQUALITY (CD ONLY)

In Chapter 3 we showed that if X is a normal random variable with mean μ and standard deviation σ, $P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 0.95$. This result relates the probability of a normal random variable to the magnitude of the standard deviation. An interesting, similar result that applies to any discrete or continuous random variable was developed by the mathematician Chebyshev in 1867.

Chebyshev's Inequality

For any random variable X with mean μ and variance σ^2,

$$P(|X - \mu| \geq c\sigma) \leq 1/c^2$$

for $c > 0$.

This result is interpreted as follows. The probability that a random variable differs from its mean by at least c standard deviations is less than or equal to $1/c^2$. Note that the rule is useful only for $c > 1$.

For example, using $c = 2$ implies that the probability that *any* random variable differs from its mean by at least two standard deviations is no greater than 1/4. We know that for a normal random variable, this probability is less than 0.05. Also, using $c = 3$ implies that the probability that any random variable differs from its mean by at least three standard deviations is no greater than 1/9. **Chebyshev's inequality** provides a relationship between the standard deviation and the dispersion of the probability distribution of any random variable. The proof is left as an exercise.

Table S5-1 compares probabilities computed by Chebyshev's rule to probabilities computed for a normal random variable.

EXAMPLE S5-8

The process of drilling holes in printed circuit boards produces diameters with a standard deviation of 0.01 millimeter. How many diameters must be measured so that the probability is at least 8/9 that the average of the measured diameters is within 0.005 of the process mean diameter μ?

Let $X_1, X_2, \ldots, X_n$ be the random variables that denote the diameters of n holes. The average measured diameter is $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$. Assume that the X's are independent random variables. From Equation 5-40, $E(\overline{X}) = \mu$ and $V(\overline{X}) = 0.01^2/n$. Consequently, the standard deviation of $\overline{X}$ is $(0.01^2/n)^{1/2}$. By applying Chebyshev's inequality to $\overline{X}$,

$$P(|\overline{X} - \mu| \geq c(0.01^2/n)^{1/2}) \leq 1/c^2$$

Let $c = 3$. Then,

$$P(|\overline{X} - \mu| \geq 3(0.01^2/n)^{1/2}) \leq 1/9$$

Therefore,

$$P(|\overline{X} - \mu| < 3(0.01^2/n)^{1/2}) \geq 8/9$$

Table S5-1 Percentage of Distribution Greater than c Standard Deviations from the Mean

c	Chebyshev's Rule for any Probability Distribution	Normal Distribution
1.5	less than 44.4%	13.4%
2	less than 25.0%	4.6%
3	less than 11.1%	0.27%
4	less than 6.3%	0.01%

Thus, the probability that $\overline{X}$ is within $3(0.01^2/n)^{1/2}$ of μ is at least 8/9. Finally, n is chosen such that $3(0.01^2/n)^{1/2} = 0.005$. That is,

$$n = 3^2[0.01^2/0.005^2] = 36$$

EXERCISES FOR SECTION 5-10

S5-25. The photoresist thickness in semiconductor manufacturing has a mean of 10 micrometers and a standard deviation of 1 micrometer. Bound the probability that the thickness is less than 6 or greater than 14 micrometers.

S5-26. Suppose X has a continuous uniform distribution with range $0 < x < 10$. Use Chebyshev's rule to bound the probability that X differs from its mean by more than two standard deviations and compare to the actual probability.

S5-27. Suppose X has an exponential distribution with mean 20. Use Chebyshev's rule to bound the probability that X differs from its mean by more than two standard deviations and by more than three standard deviations and compare to the actual probabilities.

S5-28. Suppose X has a Poisson distribution with mean $\lambda = 4$. Use Chebyshev's rule to bound the probability that X differs from its mean by more than two standard deviations and by more than three standard deviations and compare to the actual probabilities.

S5-29. Consider the process of drilling holes in printed circuits boards. Assume that the standard deviation of the diameters is 0.01 and that the diameters are independent. Suppose that the average of 500 diameters is used to estimate the process mean.

(a) The probability is at least 15/16 that the measured average is within some bound of the process mean. What is the bound?

(b) If it is assumed that the diameters are normally distributed, determine the bound such that the probability is 15/16 that the measured average is closer to the process mean than the bound.

S5-30. Prove Chebyshev's rule from the following steps. Define the random variable Y as follows:

$$Y = \begin{cases} 1 & \text{if } |X - \mu| \geq c\sigma \\ 0 & \text{otherwise} \end{cases}$$

(a) Determine $E(Y)$

(b) Show that $(X - \mu)^2 \geq (X - \mu)^2\, Y \geq c^2\sigma^2 Y$

(c) Using part (b), show that $E[(X - \mu)^2] \geq c^2\sigma^2\, E[Y]$

(d) Using part (c), complete the derivation of Chebyshev's inequality.