Chapter 3

Random Variables and Probability Distributions

3.1 Concept of a Random Variable

Statistics is concerned with making inferences about populations and population characteristics. Experiments are conducted with results that are subject to chance. The testing of a number of electronic components is an example of a **statistical experiment**, a term that is used to describe any process by which several chance observations are generated. It is often important to allocate a numerical description to the outcome. For example, the sample space giving a detailed description of each possible outcome when three electronic components are tested may be written

$$S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\},$$

where N denotes nondefective and D denotes defective. One is naturally concerned with the number of defectives that occur. Thus, each point in the sample space will be *assigned a numerical value* of 0, 1, 2, or 3. These values are, of course, random quantities *determined by the outcome of the experiment.* They may be viewed as values assumed by the *random variable* X, the number of defective items when three electronic components are tested.

Definition 3.1: A **random variable** is a function that associates a real number with each element in the sample space.

We shall use a capital letter, say X, to denote a random variable and its corresponding small letter, x in this case, for one of its values. In the electronic component testing illustration above, we notice that the random variable X assumes the value 2 for all elements in the subset

$$E = \{DDN, DND, NDD\}$$

of the sample space S. That is, each possible value of X represents an event that is a subset of the sample space for the given experiment.

Example 3.1: Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y, where Y is the number of red balls, are

Sample Space	y
RR	2
RB	1
BR	1
BB	0

Example 3.2: A stockroom clerk returns three safety helmets at random to three steel mill employees who had previously checked them. If Smith, Jones, and Brown, in that order, receive one of the three hats, list the sample points for the possible orders of returning the helmets, and find the value m of the random variable M that represents the number of correct matches.

***Solution*:** If S, J, and B stand for Smith's, Jones's, and Brown's helmets, respectively, then the possible arrangements in which the helmets may be returned and the number of correct matches are

Sample Space	m
SJB	3
SBJ	1
BJS	1
JSB	1
JBS	0
BSJ	0

In each of the two preceding examples, the sample space contains a finite number of elements. On the other hand, when a die is thrown until a 5 occurs, we obtain a sample space with an unending sequence of elements,

$$S = \{F, NF, NNF, NNNF, \dots\},$$

where F and N represent, respectively, the occurrence and nonoccurrence of a 5. But even in this experiment, the number of elements can be equated to the number of whole numbers so that there is a first element, a second element, a third element, and so on, and in this sense can be counted.

There are cases where the random variable is categorical in nature. Variables, often called *dummy* variables, are used. A good illustration is the case in which the random variable is binary in nature, as shown in the following example.

Example 3.3: Consider the simple condition in which components are arriving from the production line and they are stipulated to be defective or not defective. Define the random variable X by

$$X = \begin{cases} 1, & \text{if the component is defective,} \\ 0, & \text{if the component is not defective.} \end{cases}$$

Clearly the assignment of 1 or 0 is arbitrary though quite convenient. This will become clear in later chapters. The random variable for which 0 and 1 are chosen to describe the two possible values is called a **Bernoulli random variable**.

Further illustrations of random variables are revealed in the following examples.

Example 3.4: Statisticians use **sampling plans** to either accept or reject batches or lots of material. Suppose one of these sampling plans involves sampling independently 10 items from a lot of 100 items in which 12 are defective.

Let X be the random variable defined as the number of items found defective in the sample of 10. In this case, the random variable takes on the values $0, 1, 2, \ldots, 9, 10$.

Example 3.5: Suppose a sampling plan involves sampling items from a process until a defective is observed. The evaluation of the process will depend on how many consecutive items are observed. In that regard, let X be a random variable defined by the number of items observed before a defective is found. With N a nondefective and D a defective, sample spaces are $S = \{D\}$ given $X = 1$, $S = \{ND\}$ given $X = 2$, $S = \{NND\}$ given $X = 3$, and so on.

Example 3.6: Interest centers around the proportion of people who respond to a certain mail order solicitation. Let X be that proportion. X is a random variable that takes on all values x for which $0 \leq x \leq 1$.

Example 3.7: Let X be the random variable defined by the waiting time, in hours, between successive speeders spotted by a radar unit. The random variable X takes on all values x for which $x \geq 0$.

Definition 3.2: If a sample space contains a finite number of possibilities or an unending sequence with as many elements as there are whole numbers, it is called a **discrete sample space**.

The outcomes of some statistical experiments may be neither finite nor countable. Such is the case, for example, when one conducts an investigation measuring the distances that a certain make of automobile will travel over a prescribed test course on 5 liters of gasoline. Assuming distance to be a variable measured to any degree of accuracy, then clearly we have an infinite number of possible distances in the sample space that cannot be equated to the number of whole numbers. Or, if one were to record the length of time for a chemical reaction to take place, once again the possible time intervals making up our sample space would be infinite in number and uncountable. We see now that all sample spaces need not be discrete.

Definition 3.3: If a sample space contains an infinite number of possibilities equal to the number of points on a line segment, it is called a **continuous sample space**.

A random variable is called a **discrete random variable** if its set of possible outcomes is countable. The random variables in Examples 3.1 to 3.5 are discrete random variables. But a random variable whose set of possible values is an entire interval of numbers is not discrete. When a random variable can take on values on a continuous scale, it is called a **continuous random variable**. Often the possible values of a continuous random variable are precisely the same values that are contained in the continuous sample space. Obviously, the random variables described in Examples 3.6 and 3.7 are continuous random variables.

In most practical problems, continuous random variables represent *measured* data, such as all possible heights, weights, temperatures, distance, or life periods, whereas discrete random variables represent *count* data, such as the number of defectives in a sample of k items or the number of highway fatalities per year in a given state. Note that the random variables Y and M of Examples 3.1 and 3.2 both represent count data, Y the number of red balls and M the number of correct hat matches.

3.2 Discrete Probability Distributions

A discrete random variable assumes each of its values with a certain probability. In the case of tossing a coin three times, the variable X, representing the number of heads, assumes the value 2 with probability 3/8, since 3 of the 8 equally likely sample points result in two heads and one tail. If one assumes equal weights for the simple events in Example 3.2, the probability that no employee gets back the right helmet, that is, the probability that M assumes the value 0, is 1/3. The possible values m of M and their probabilities are

m	0	1	3
$P(M = m)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Note that the values of m exhaust all possible cases and hence the probabilities add to 1.

Frequently, it is convenient to represent all the probabilities of a random variable X by a formula. Such a formula would necessarily be a function of the numerical values x that we shall denote by $f(x)$, $g(x)$, $r(x)$, and so forth. Therefore, we write $f(x) = P(X = x)$; that is, $f(3) = P(X = 3)$. The set of ordered pairs $(x, f(x))$ is called the **probability function**, **probability mass function**, or **probability distribution** of the discrete random variable X.

Definition 3.4: The set of ordered pairs $(x, f(x))$ is a **probability function**, **probability mass function**, or **probability distribution** of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0$,
2. $\sum_x f(x) = 1$,
3. $P(X = x) = f(x)$. probability distribution

Example 3.8: A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Solution: Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. Then x can only take the numbers 0, 1, and

2. Now

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}, \quad f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190},$$

$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}.$$

Thus, the probability distribution of X is

x	0	1	2
$f(x)$	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

Example 3.9: If a car agency sells 50% of its inventory of a certain foreign car equipped with side airbags, find a formula for the probability distribution of the number of cars with side airbags among the next 4 cars sold by the agency.

***Solution*:** Since the probability of selling an automobile with side airbags is 0.5, the $2^4 = 16$ points in the sample space are equally likely to occur. Therefore, the denominator for all probabilities, and also for our function, is 16. To obtain the number of ways of selling 3 cars with side airbags, we need to consider the number of ways of partitioning 4 outcomes into two cells, with 3 cars with side airbags assigned to one cell and the model without side airbags assigned to the other. This can be done in $\binom{4}{3} = 4$ ways. In general, the event of selling x models with side airbags and $4 - x$ models without side airbags can occur in $\binom{4}{x}$ ways, where x can be 0, 1, 2, 3, or 4. Thus, the probability distribution $f(x) = P(X = x)$ is

$$f(x) = \frac{1}{16}\binom{4}{x}, \quad \text{for } x = 0, 1, 2, 3, 4.$$

There are many problems where we may wish to compute the probability that the observed value of a random variable X will be less than or equal to some real number x. Writing $F(x) = P(X \leq x)$ for every real number x, we define $F(x)$ to be the **cumulative distribution function** of the random variable X.

Definition 3.5: The **cumulative distribution function** $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \quad \text{for } -\infty < x < \infty.$$

For the random variable M, the number of correct matches in Example 3.2, we have

$$F(2) = P(M \leq 2) = f(0) + f(1) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6}.$$

The cumulative distribution function of M is

$$F(m) = \begin{cases} 0, & \text{for } m < 0, \\ \frac{1}{3}, & \text{for } 0 \leq m < 1, \\ \frac{5}{6}, & \text{for } 1 \leq m < 3, \\ 1, & \text{for } m \geq 3. \end{cases}$$

One should pay particular notice to the fact that the cumulative distribution function is a monotone nondecreasing function defined not only for the values assumed by the given random variable but for all real numbers.

Example 3.10: Find the cumulative distribution function of the random variable X in Example 3.9. Using $F(x)$, verify that $f(2) = 3/8$.

Solution: Direct calculations of the probability distribution of Example 3.9 give $f(0)= 1/16$, $f(1) = 1/4$, $f(2)= 3/8$, $f(3)= 1/4$, and $f(4)= 1/16$. Therefore,

$$\begin{aligned}
F(0) &= f(0) = \frac{1}{16},\\
F(1) &= f(0) + f(1) = \frac{5}{16},\\
F(2) &= f(0) + f(1) + f(2) = \frac{11}{16},\\
F(3) &= f(0) + f(1) + f(2) + f(3) = \frac{15}{16},\\
F(4) &= f(0) + f(1) + f(2) + f(3) + f(4) = 1.
\end{aligned}$$

Hence,

$$F(x) = \begin{cases} 0, & \text{for } x < 0,\\ \frac{1}{16}, & \text{for } 0 \le x < 1,\\ \frac{5}{16}, & \text{for } 1 \le x < 2,\\ \frac{11}{16}, & \text{for } 2 \le x < 3,\\ \frac{15}{16}, & \text{for } 3 \le x < 4,\\ 1 & \text{for } x \ge 4. \end{cases}$$

Now

$$f(2) = F(2) - F(1) = \frac{11}{16} - \frac{5}{16} = \frac{3}{8}.$$

It is often helpful to look at a probability distribution in graphic form. One might plot the points $(x, f(x))$ of Example 3.9 to obtain Figure 3.1. By joining the points to the x axis either with a dashed or with a solid line, we obtain a probability mass function plot. Figure 3.1 makes it easy to see what values of X are most likely to occur, and it also indicates a perfectly symmetric situation in this case.

Instead of plotting the points $(x, f(x))$, we more frequently construct rectangles, as in Figure 3.2. Here the rectangles are constructed so that their bases of equal width are centered at each value x and their heights are equal to the corresponding probabilities given by $f(x)$. The bases are constructed so as to leave no space between the rectangles. Figure 3.2 is called a **probability histogram**.

Since each base in Figure 3.2 has unit width, $P(X = x)$ is equal to the area of the rectangle centered at x. Even if the bases were not of unit width, we could adjust the heights of the rectangles to give areas that would still equal the probabilities of X assuming any of its values x. This concept of using areas to represent

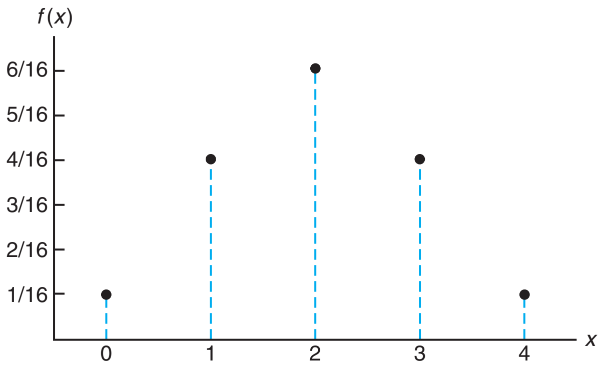


Figure 3.1: Probability mass function plot.

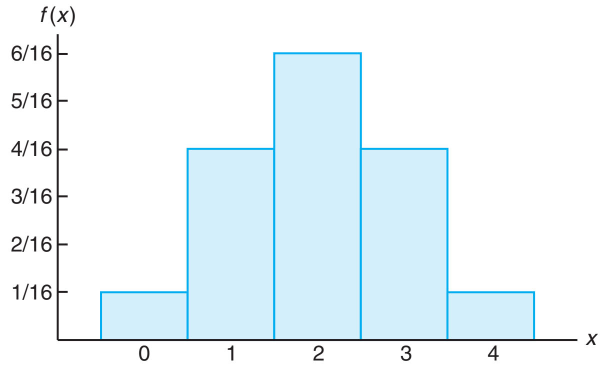


Figure 3.2: Probability histogram.

probabilities is necessary for our consideration of the probability distribution of a continuous random variable.

The graph of the cumulative distribution function of Example 3.9, which appears as a step function in Figure 3.3, is obtained by plotting the points $(x, F(x))$.

Certain probability distributions are applicable to more than one physical situation. The probability distribution of Example 3.9, for example, also applies to the random variable Y, where Y is the number of heads when a coin is tossed 4 times, or to the random variable W, where W is the number of red cards that occur when 4 cards are drawn at random from a deck in succession with each card replaced and the deck shuffled before the next drawing. Special discrete distributions that can be applied to many different experimental situations will be considered in Chapter 5.

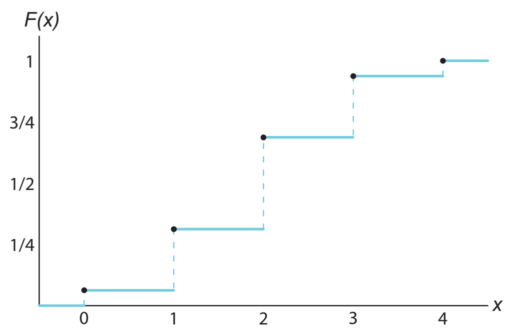


Figure 3.3: Discrete cumulative distribution function.

3.3 Continuous Probability Distributions

A continuous random variable has a probability of 0 of assuming *exactly* any of its values. Consequently, its probability distribution cannot be given in tabular form.

At first this may seem startling, but it becomes more plausible when we consider a particular example. Let us discuss a random variable whose values are the heights of all people over 21 years of age. Between any two values, say 163.5 and 164.5 centimeters, or even 163.99 and 164.01 centimeters, there are an infinite number of heights, one of which is 164 centimeters. The probability of selecting a person at random who is exactly 164 centimeters tall and not one of the infinitely large set of heights so close to 164 centimeters that you cannot humanly measure the difference is remote, and thus we assign a probability of 0 to the event. This is not the case, however, if we talk about the probability of selecting a person who is at least 163 centimeters but not more than 165 centimeters tall. Now we are dealing with an interval rather than a point value of our random variable.

We shall concern ourselves with computing probabilities for various intervals of continuous random variables such as $P(a < X < b)$, $P(W \geq c)$, and so forth. Note that when X is continuous,

$$P(a < X \leq b) = P(a < X < b) + P(X = b) = P(a < X < b).$$

That is, it does not matter whether we include an endpoint of the interval or not. This is not true, though, when X is discrete.

Although the probability distribution of a continuous random variable cannot be presented in tabular form, it can be stated as a formula. Such a formula would necessarily be a function of the numerical values of the continuous random variable X and as such will be represented by the functional notation $f(x)$. In dealing with continuous variables, $f(x)$ is usually called the **probability density function**, or simply the **density function**, of X. Since X is defined over a continuous sample space, it is possible for $f(x)$ to have a finite number of discontinuities. However, most density functions that have practical applications in the analysis of statistical data are continuous and their graphs may take any of several forms, some of which are shown in Figure 3.4. Because areas will be used to represent probabilities and probabilities are positive numerical values, the density function must lie entirely above the x axis.

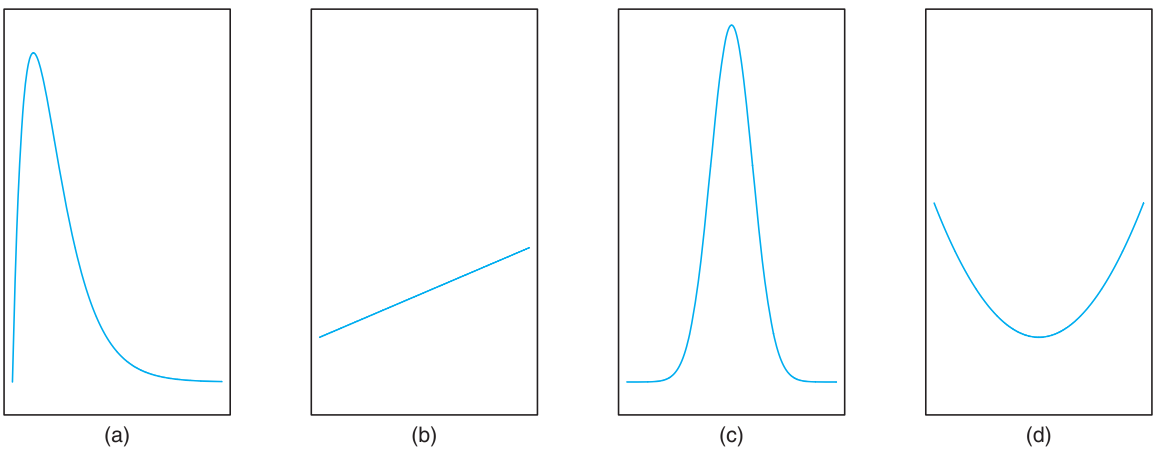


Figure 3.4: Typical density functions.

A probability density function is constructed so that the area under its curve bounded by the x axis is equal to 1 when computed over the range of X for which $f(x)$ is defined. Should this range of X be a finite interval, it is always possible to extend the interval to include the entire set of real numbers by defining $f(x)$ to be zero at all points in the extended portions of the interval. In Figure 3.5, the probability that X assumes a value between a and b is equal to the shaded area under the density function between the ordinates at $x = a$ and $x = b$, and from integral calculus is given by

$$P(a < X < b) = \int_a^b f(x)\, dx.$$

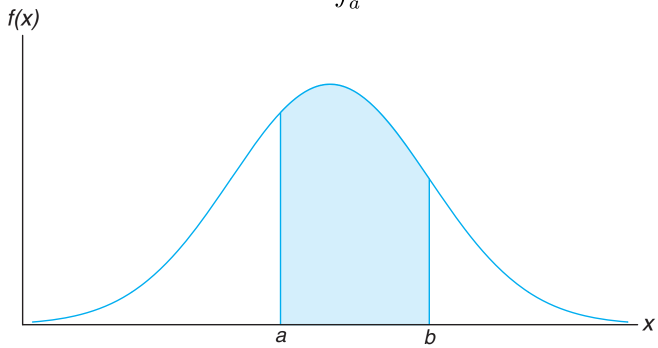


Figure 3.5: $P(a < X < b)$.

Definition 3.6: The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

[for a point, $P(a \leq x \leq a) = 0$]

Example 3.11: Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere}. \end{cases}$$

.

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

***Solution*:** We use Definition 3.6.

(a) Obviously, $f(x) \geq 0$. To verify condition 2 in Definition 3.6, we have

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-1}^{2} \frac{x^2}{3} dx = \frac{x^3}{9}|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1.$$

(b) Using formula 3 in Definition 3.6, we obtain

$$P(0 < X \leq 1) = \int_0^1 \frac{x^2}{3} dx = \left. \frac{x^3}{9} \right|_0^1 = \frac{1}{9}.$$

Definition 3.7: The **cumulative distribution function** $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt, \quad \text{for } -\infty < x < \infty.$$

As an immediate consequence of Definition 3.7, one can write the two results

$$P(a < X < b) = F(b) - F(a) \text{ and } f(x) = \frac{dF(x)}{dx},$$

if the derivative exists.

Example 3.12: For the density function of Example 3.11, find $F(x)$, and use it to evaluate $P(0 < X \leq 1)$.

***Solution*:** For $-1 < x < 2$,

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \int_{-1}^{x} \frac{t^2}{3} dt = \left. \frac{t^3}{9} \right|_{-1}^{x} = \frac{x^3 + 1}{9}.$$

Therefore,

$$F(x) = \begin{cases} 0, & x < -1, \\ \frac{x^3+1}{9}, & -1 \leq x < 2, \\ 1, & x \geq 2. \end{cases}$$

The cumulative distribution function $F(x)$ is expressed in Figure 3.6. Now

$$P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9},$$

which agrees with the result obtained by using the density function in Example 3.11.

Example 3.13: The Department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate b. The DOE has determined that the density function of the winning (low) bid is

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \leq y \leq 2b, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $F(y)$ and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate b.

***Solution*:** For $2b/5 \leq y \leq 2b$,

$$F(y) = \int_{2b/5}^{y} \frac{5}{8b} dy = \left. \frac{5t}{8b} \right|_{2b/5}^{y} = \frac{5y}{8b} - \frac{1}{4}.$$

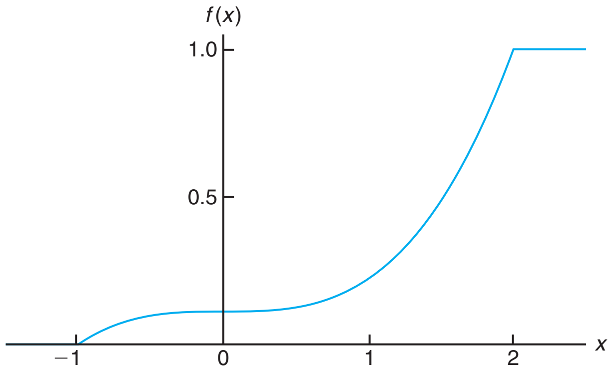


Figure 3.6: Continuous cumulative distribution function.

Thus,

$$F(y) = \begin{cases} 0, & y < \frac{2}{5}b, \\ \frac{5y}{8b} - \frac{1}{4}, & \frac{2}{5}b \le y < 2b, \\ 1, & y \ge 2b. \end{cases}$$

To determine the probability that the winning bid is less than the preliminary bid estimate b, we have

$$P(Y \le b) = F(b) = \frac{5}{8} - \frac{1}{4} = \frac{3}{8}.$$

Exercises

3.1 Classify the following random variables as discrete or continuous:

X: the number of automobile accidents per year in Virginia.

Y: the length of time to play 18 holes of golf.

M: the amount of milk produced yearly by a particular cow.

N: the number of eggs laid each month by a hen.

P: the number of building permits issued each month in a certain city.

Q: the weight of grain produced per acre.

3.2 An overseas shipment of 5 foreign automobiles contains 2 that have slight paint blemishes. If an agency receives 3 of these automobiles at random, list the elements of the sample space S, using the letters B and N for blemished and nonblemished, respectively; then to each sample point assign a value x of the random variable X representing the number of automobiles with paint blemishes purchased by the agency.

3.3 Let W be a random variable giving the number of heads minus the number of tails in three tosses of a coin. List the elements of the sample space S for the three tosses of the coin and to each sample point assign a value w of W.

3.4 A coin is flipped until 3 heads in succession occur. List only those elements of the sample space that require 6 or less tosses. Is this a discrete sample space? Explain.

3.5 Determine the value c so that each of the following functions can serve as a probability distribution of the discrete random variable X:

(a) $f(x) = c(x^2 + 4)$, for $x = 0, 1, 2, 3$;

(b) $f(x) = c\binom{2}{x}\binom{3}{3-x}$, for $x = 0, 1, 2$.

3.6 The shelf life, in days, for bottles of a certain prescribed medicine is a random variable having the density function

$$f(x) = \begin{cases} \frac{20{,}000}{(x+100)^3}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that a bottle of this medicine will have a shell life of

(a) at least 200 days;

(b) anywhere from 80 to 120 days.

3.7 The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

$$f(x) = \begin{cases} x, & 0 < x < 1, \\ 2 - x, & 1 \le x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that over a period of one year, a family runs their vacuum cleaner

(a) less than 120 hours;

(b) between 50 and 100 hours.

3.8 Find the probability distribution of the random variable W in Exercise 3.3, assuming that the coin is biased so that a head is twice as likely to occur as a tail.

3.9 The proportion of people who respond to a certain mail-order solicitation is a continuous random variable X that has the density function

$$f(x) = \begin{cases} \frac{2(x+2)}{5}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Show that $P(0 < X < 1) = 1$.

(b) Find the probability that more than 1/4 but fewer than 1/2 of the people contacted will respond to this type of solicitation.

3.10 Find a formula for the probability distribution of the random variable X representing the outcome when a single die is rolled once.

3.11 A shipment of 7 television sets contains 2 defective sets. A hotel makes a random purchase of 3 of the sets. If x is the number of defective sets purchased by the hotel, find the probability distribution of X. Express the results graphically as a probability histogram.

3.12 An investment firm offers its customers municipal bonds that mature after varying numbers of years. Given that the cumulative distribution function of T, the number of years to maturity for a randomly selected bond, is

$$F(t) = \begin{cases} 0, & t < 1, \\ \frac{1}{4}, & 1 \le t < 3, \\ \frac{1}{2}, & 3 \le t < 5, \\ \frac{3}{4}, & 5 \le t < 7, \\ 1, & t \ge 7, \end{cases}$$

find

(a) $P(T = 5)$;

(b) $P(T > 3)$;

(c) $P(1.4 < T < 6)$;

(d) $P(T \le 5 \mid T \ge 2)$.

3.13 The probability distribution of X, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given by

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Construct the cumulative distribution function of X.

3.14 The waiting time, in hours, between successive speeders spotted by a radar unit is a continuous random variable with cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-8x}, & x \ge 0. \end{cases}$$

Find the probability of waiting less than 12 minutes between successive speeders

(a) using the cumulative distribution function of X;

(b) using the probability density function of X.

3.15 Find the cumulative distribution function of the random variable X representing the number of defectives in Exercise 3.11. Then using $F(x)$, find

(a) $P(X = 1)$;

(b) $P(0 < X \le 2)$.

3.16 Construct a graph of the cumulative distribution function of Exercise 3.15.

3.17 A continuous random variable X that can assume values between $x = 1$ and $x = 3$ has a density function given by $f(x) = 1/2$.

(a) Show that the area under the curve is equal to 1.

(b) Find $P(2 < X < 2.5)$.

(c) Find $P(X \le 1.6)$.

3.18 A continuous random variable X that can assume values between $x = 2$ and $x = 5$ has a density function given by $f(x) = 2(1 + x)/27$. Find

(a) $P(X < 4)$;

(b) $P(3 \leq X < 4)$.

3.19 For the density function of Exercise 3.17, find $F(x)$. Use it to evaluate $P(2 < X < 2.5)$.

3.20 For the density function of Exercise 3.18, find $F(x)$, and use it to evaluate $P(3 \leq X < 4)$.

3.21 Consider the density function

$$f(x) = \begin{cases} k\sqrt{x}, & 0 < x < 1, \\ 0, & \text{elsewhere}. \end{cases}$$

(a) Evaluate k.

(b) Find $F(x)$ and use it to evaluate

$$P(0.3 < X < 0.6).$$

3.22 Three cards are drawn in succession from a deck without replacement. Find the probability distribution for the number of spades.

3.23 Find the cumulative distribution function of the random variable W in Exercise 3.8. Using $F(w)$, find

(a) $P(W > 0)$;

(b) $P(-1 \leq W < 3)$.

3.24 Find the probability distribution for the number of jazz CDs when 4 CDs are selected at random from a collection consisting of 5 jazz CDs, 2 classical CDs, and 3 rock CDs. Express your results by means of a formula.

3.25 From a box containing 4 dimes and 2 nickels, 3 coins are selected at random without replacement. Find the probability distribution for the total T of the 3 coins. Express the probability distribution graphically as a probability histogram.

3.26 From a box containing 4 black balls and 2 green balls, 3 balls are drawn in succession, each ball being replaced in the box before the next draw is made. Find the probability distribution for the number of green balls.

3.27 The time to failure in hours of an important piece of electronic equipment used in a manufactured DVD player has the density function

$$f(x) = \begin{cases} \frac{1}{2000}\exp(-x/2000), & x \geq 0, \\ 0, & x < 0. \end{cases}$$

(a) Find $F(x)$.

(b) Determine the probability that the component (and thus the DVD player) lasts more than 1000 hours before the component needs to be replaced.

(c) Determine the probability that the component fails before 2000 hours.

3.28 A cereal manufacturer is aware that the weight of the product in the box varies slightly from box to box. In fact, considerable historical data have allowed the determination of the density function that describes the probability structure for the weight (in ounces). Letting X be the random variable weight, in ounces, the density function can be described as

$$f(x) = \begin{cases} \frac{2}{5}, & 23.75 \leq x \leq 26.25, \\ 0, & \text{elsewhere}. \end{cases}$$

(a) Verify that this is a valid density function.

(b) Determine the probability that the weight is smaller than 24 ounces.

(c) The company desires that the weight exceeding 26 ounces be an extremely rare occurrence. What is the probability that this rare occurrence does actually occur?

3.29 An important factor in solid missile fuel is the particle size distribution. Significant problems occur if the particle sizes are too large. From production data in the past, it has been determined that the particle size (in micrometers) distribution is characterized by

$$f(x) = \begin{cases} 3x^{-4}, & x > 1, \\ 0, & \text{elsewhere}. \end{cases}$$

(a) Verify that this is a valid density function.

(b) Evaluate $F(x)$.

(c) What is the probability that a random particle from the manufactured fuel exceeds 4 micrometers?

3.30 Measurements of scientific systems are always subject to variation, some more than others. There are many structures for measurement error, and statisticians spend a great deal of time modeling these errors. Suppose the measurement error X of a certain physical quantity is decided by the density function

$$f(x) = \begin{cases} k(3 - x^2), & -1 \leq x \leq 1, \\ 0, & \text{elsewhere}. \end{cases}$$

(a) Determine k that renders $f(x)$ a valid density function.

(b) Find the probability that a random error in measurement is less than 1/2.

(c) For this particular measurement, it is undesirable if the *magnitude* of the error (i.e., $|x|$) exceeds 0.8. What is the probability that this occurs?

3.31 Based on extensive testing, it is determined by the manufacturer of a washing machine that the time Y (in years) before a major repair is required is characterized by the probability density function

$$f(y) = \begin{cases} \frac{1}{4}e^{-y/4}, & y \geq 0, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Critics would certainly consider the product a bargain if it is unlikely to require a major repair before the sixth year. Comment on this by determining $P(Y > 6)$.

(b) What is the probability that a major repair occurs in the first year?

3.32 The proportion of the budget for a certain type of industrial company that is allotted to environmental and pollution control is coming under scrutiny. A data collection project determines that the distribution of these proportions is given by

$$f(y) = \begin{cases} 5(1-y)^4, & 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that the above is a valid density function.

(b) What is the probability that a company chosen at random expends less than 10% of its budget on environmental and pollution controls?

(c) What is the probability that a company selected at random spends more than 50% of its budget on environmental and pollution controls?

3.33 Suppose a certain type of small data processing firm is so specialized that some have difficulty making a profit in their first year of operation. The probability density function that characterizes the proportion Y that make a profit is given by

$$f(y) = \begin{cases} ky^4(1-y)^3, & 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) What is the value of k that renders the above a valid density function?

(b) Find the probability that at most 50% of the firms make a profit in the first year.

(c) Find the probability that at least 80% of the firms make a profit in the first year.

3.34 Magnetron tubes are produced on an automated assembly line. A sampling plan is used periodically to assess quality of the lengths of the tubes. This measurement is subject to uncertainty. It is thought that the probability that a random tube meets length specification is 0.99. A sampling plan is used in which the lengths of 5 random tubes are measured.

(a) Show that the probability function of Y, the number out of 5 that meet length specification, is given by the following discrete probability function:

$$f(y) = \frac{5!}{y!(5-y)!}(0.99)^y(0.01)^{5-y},$$

for $y = 0, 1, 2, 3, 4, 5$.

(b) Suppose random selections are made off the line and 3 are outside specifications. Use $f(y)$ above either to support or to refute the conjecture that the probability is 0.99 that a single tube meets specifications.

3.35 Suppose it is known from large amounts of historical data that X, the number of cars that arrive at a specific intersection during a 20-second time period, is characterized by the following discrete probability function:

$$f(x) = e^{-6}\frac{6^x}{x!}, \text{ for } x = 0, 1, 2, \ldots.$$

(a) Find the probability that in a specific 20-second time period, more than 8 cars arrive at the intersection.

(b) Find the probability that only 2 cars arrive.

3.36 On a laboratory assignment, if the equipment is working, the density function of the observed outcome, X, is

$$f(x) = \begin{cases} 2(1-x), & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Calculate $P(X \leq 1/3)$.

(b) What is the probability that X will exceed 0.5?

(c) Given that $X \geq 0.5$, what is the probability that X will be less than 0.75?

3.4 Joint Probability Distributions

Our study of random variables and their probability distributions in the preceding sections is restricted to one-dimensional sample spaces, in that we recorded outcomes of an experiment as values assumed by a single random variable. There will be situations, however, where we may find it desirable to record the simultaneous outcomes of several random variables. For example, we might measure the amount of precipitate P and volume V of gas released from a controlled chemical experiment, giving rise to a two-dimensional sample space consisting of the outcomes (p, v), or we might be interested in the hardness H and tensile strength T of cold-drawn copper, resulting in the outcomes (h, t). In a study to determine the likelihood of success in college based on high school data, we might use a three-dimensional sample space and record for each individual his or her aptitude test score, high school class rank, and grade-point average at the end of freshman year in college.

If X and Y are two discrete random variables, the probability distribution for their simultaneous occurrence can be represented by a function with values $f(x, y)$ for any pair of values (x, y) within the range of the random variables X and Y. It is customary to refer to this function as the **joint probability distribution** of X and Y.

Hence, in the discrete case,

$$f(x, y) = P(X = x, Y = y);$$

that is, the values $f(x, y)$ give the probability that outcomes x and y occur at the same time. For example, if an 18-wheeler is to have its tires serviced and X represents the number of miles these tires have been driven and Y represents the number of tires that need to be replaced, then $f(30000, 5)$ is the probability that the tires are used over 30,000 miles and the truck needs 5 new tires.

Definition 3.8: The function $f(x, y)$ is a **joint probability distribution** or **probability mass function** of the discrete random variables X and Y if

1. $f(x, y) \geq 0$ for all (x, y),
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

For any region A in the xy plane, $P[(X, Y) \in A] = \sum \sum_A f(x, y)$.

Example 3.14: Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find

(a) the joint probability function $f(x, y)$,

(b) $P[(X, Y) \in A]$, where A is the region $\{(x, y) | x + y \leq 1\}$.

***Solution*:** The possible pairs of values (x, y) are $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$, $(0, 2)$, and $(2, 0)$.

(a) Now, $f(0, 1)$, for example, represents the probability that a red and a green pens are selected. The total number of equally likely ways of selecting any 2 pens from the 8 is $\binom{8}{2} = 28$. The number of ways of selecting 1 red from 2 red pens and 1 green from 3 green pens is $\binom{2}{1}\binom{3}{1} = 6$. Hence, $f(0, 1) = 6/28$ $= 3/14$. Similar calculations yield the probabilities for the other cases, which are presented in Table 3.1. Note that the probabilities sum to 1. In Chapter

5, it will become clear that the joint probability distribution of Table 3.1 can be represented by the formula

$$f(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}},$$

for $x = 0, 1, 2$; $y = 0, 1, 2$; and $0 \le x + y \le 2$.

(b) The probability that (X, Y) fall in the region A is

$$P[(X,Y) \in A] = P(X + Y \le 1) = f(0,0) + f(0,1) + f(1,0)$$
$$= \frac{3}{28} + \frac{3}{14} + \frac{9}{28} = \frac{9}{14}.$$

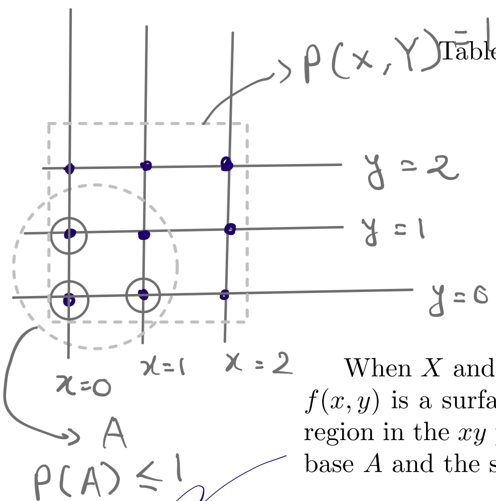


Table 3.1: Joint Probability Distribution for Example 3.14

	$f(x,y)$	x = 0	1	2	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

When X and Y are continuous random variables, the **joint density function** $f(x, y)$ is a surface lying above the xy plane, and $P[(X, Y) \in A]$, where A is any region in the xy plane, is equal to the volume of the right cylinder bounded by the base A and the surface.

Definition 3.9: The function $f(x, y)$ is a **joint density function** of the continuous random variables X and Y if

1. $f(x, y) \ge 0$, for all (x, y),
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$,
3. $P[(X, Y) \in A] = \int\int_A f(x, y)\, dx\, dy$, for any region A in the xy plane.

Example 3.15: A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify condition 2 of Definition 3.9.

(b) Find $P[(X, Y) \in A]$, where $A = \{(x, y) \mid 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$.

Solution: (a) The integration of $f(x, y)$ over the whole region is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ dx\ dy = \int_0^1\int_0^1 \frac{2}{5}(2x+3y)\ dx\ dy$$
$$= \int_0^1 \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Bigg|_{x=0}^{x=1} dy$$
$$= \int_0^1 \left(\frac{2}{5} + \frac{6y}{5}\right) dy = \left(\frac{2y}{5} + \frac{3y^2}{5}\right)\Bigg|_0^1 = \frac{2}{5} + \frac{3}{5} = 1.$$

(b) To calculate the probability, we use

$$P[(X,Y) \in A] = P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right)$$
$$= \int_{1/4}^{1/2}\int_0^{1/2} \frac{2}{5}(2x+3y)\ dx\ dy$$
$$= \int_{1/4}^{1/2} \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Bigg|_{x=0}^{x=1/2} dy = \int_{1/4}^{1/2}\left(\frac{1}{10} + \frac{3y}{5}\right) dy$$
$$= \left(\frac{y}{10} + \frac{3y^2}{10}\right)\Bigg|_{1/4}^{1/2}$$
$$= \frac{1}{10}\left[\left(\frac{1}{2} + \frac{3}{4}\right) - \left(\frac{1}{4} + \frac{3}{16}\right)\right] = \frac{13}{160}.$$

Given the joint probability distribution $f(x, y)$ of the discrete random variables X and Y, the probability distribution $g(x)$ of X alone is obtained by summing $f(x, y)$ over the values of Y. Similarly, the probability distribution $h(y)$ of Y alone is obtained by summing $f(x, y)$ over the values of X. We define $g(x)$ and $h(y)$ to be the **marginal distributions** of X and Y, respectively. When X and Y are continuous random variables, summations are replaced by integrals. We can now make the following general definition.

Definition 3.10: The **marginal distributions** of X alone and of Y alone are

$$g(x) = \sum_y f(x,y) \quad \text{and} \quad h(y) = \sum_x f(x,y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\ dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x,y)\ dx$$

for the continuous case.

The term *marginal* is used here because, in the discrete case, the values of $g(x)$ and $h(y)$ are just the marginal totals of the respective columns and rows when the values of $f(x, y)$ are displayed in a rectangular table.

Example 3.16: Show that the column and row totals of Table 3.1 give the marginal distribution of X alone and of Y alone.

Solution: For the random variable X, we see that

$$g(0) = f(0,0) + f(0,1) + f(0,2) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14},$$
$$g(1) = f(1,0) + f(1,1) + f(1,2) = \frac{9}{28} + \frac{3}{14} + 0 = \frac{15}{28},$$

and

$$g(2) = f(2,0) + f(2,1) + f(2,2) = \frac{3}{28} + 0 + 0 = \frac{3}{28},$$

which are just the column totals of Table 3.1. In a similar manner we could show that the values of $h(y)$ are given by the row totals. In tabular form, these marginal distributions may be written as follows:

x	0	1	2
$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$

y	0	1	2
$h(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

Example 3.17: Find $g(x)$ and $h(y)$ for the joint density function of Example 3.15.

Solution: By definition,

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\ dy = \int_0^1 \frac{2}{5}(2x+3y)\ dy = \left(\frac{4xy}{5} + \frac{6y^2}{10}\right)\Bigg|_{y=0}^{y=1} = \frac{4x+3}{5},$$

for $0 \le x \le 1$, and $g(x) = 0$ elsewhere. Similarly,

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\ dx = \int_0^1 \frac{2}{5}(2x+3y)\ dx = \frac{2(1+3y)}{5},$$

for $0 \le y \le 1$, and $h(y) = 0$ elsewhere.

The fact that the marginal distributions $g(x)$ and $h(y)$ are indeed the probability distributions of the individual variables X and Y alone can be verified by showing that the conditions of Definition 3.4 or Definition 3.6 are satisfied. For example, in the continuous case

$$\int_{-\infty}^{\infty} g(x)\ dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ dy\ dx = 1,$$

and

$$P(a < X < b) = P(a < X < b, -\infty < Y < \infty)$$
$$= \int_a^b \int_{-\infty}^{\infty} f(x,y)\ dy\ dx = \int_a^b g(x)\ dx.$$

In Section 3.1, we stated that the value x of the random variable X represents an event that is a subset of the sample space. If we use the definition of conditional probability as stated in Chapter 2,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ provided } P(A) > 0,$$

where A and B are now the events defined by $X = x$ and $Y = y$, respectively, then

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f(x, y)}{g(x)}, \text{ provided } g(x) > 0,$$

where X and Y are discrete random variables.

It is not difficult to show that the function $f(x, y)/g(x)$, which is strictly a function of y with x fixed, satisfies all the conditions of a probability distribution. This is also true when $f(x, y)$ and $g(x)$ are the joint density and marginal distribution, respectively, of continuous random variables. As a result, it is extremely important that we make use of the special type of distribution of the form $f(x, y)/g(x)$ in order to be able to effectively compute conditional probabilities. This type of distribution is called a **conditional probability distribution**; the formal definition follows.

Definition 3.11: Let X and Y be two random variables, discrete or continuous. The **conditional distribution** of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x, y)}{g(x)}, \text{ provided } g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x|y) = \frac{f(x, y)}{h(y)}, \text{ provided } h(y) > 0.$$

If we wish to find the probability that the discrete random variable X falls between a and b when it is known that the discrete variable $Y = y$, we evaluate

$$P(a < X < b \mid Y = y) = \sum_{a<x<b} f(x|y),$$

where the summation extends over all values of X between a and b. When X and Y are continuous, we evaluate

$$P(a < X < b \mid Y = y) = \int_a^b f(x|y)\, dx.$$

Example 3.18: Referring to Example 3.14, find the conditional distribution of X, given that $Y = 1$, and use it to determine $P(X = 0 \mid Y = 1)$.

Solution: We need to find $f(x|y)$, where $y = 1$. First, we find that

$$h(1) = \sum_{x=0}^{2} f(x, 1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}.$$

Now

$$f(x|1) = \frac{f(x, 1)}{h(1)} = \left(\frac{7}{3}\right) f(x, 1), \quad x = 0, 1, 2.$$

Therefore,

$$f(0|1) = \left(\frac{7}{3}\right) f(0,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2}, \ f(1|1) = \left(\frac{7}{3}\right) f(1,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2},$$

$$f(2|1) = \left(\frac{7}{3}\right) f(2,1) = \left(\frac{7}{3}\right)(0) = 0,$$

and the conditional distribution of X, given that $Y = 1$, is

x	0	1	2
$f(x\|1)$	$\frac{1}{2}$	$\frac{1}{2}$	0

Finally,

$$P(X = 0 \mid Y = 1) = f(0|1) = \frac{1}{2}.$$

Therefore, if it is known that 1 of the 2 pen refills selected is red, we have a probability equal to 1/2 that the other refill is not blue.

Example 3.19: The joint density for the random variables (X, Y), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces, is

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the marginal densities $g(x)$, $h(y)$, and the conditional density $f(y|x)$.

(b) Find the probability that the spectrum shifts more than half of the total observations, given that the temperature is increased by 0.25 unit.

***Solution*:** (a) By definition,

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_x^1 10xy^2\, dy$$

$$= \frac{10}{3} xy^3 \bigg|_{y=x}^{y=1} = \frac{10}{3} x(1 - x^3), \ 0 < x < 1,$$

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^y 10xy^2\, dx = 5x^2y^2\big|_{x=0}^{x=y} = 5y^4, \ 0 < y < 1.$$

Now

$$f(y|x) = \frac{f(x, y)}{g(x)} = \frac{10xy^2}{\frac{10}{3}x(1 - x^3)} = \frac{3y^2}{1 - x^3}, \ 0 < x < y < 1.$$

(b) Therefore,

$$P\left(Y > \frac{1}{2} \;\middle|\; X = 0.25\right) = \int_{1/2}^1 f(y \mid x = 0.25)\, dy = \int_{1/2}^1 \frac{3y^2}{1 - 0.25^3}\, dy = \frac{8}{9}.$$

Example 3.20: Given the joint density function

$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2, \ 0 < y < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

find $g(x)$, $h(y)$, $f(x|y)$, and evaluate $P(\frac{1}{4} < X < \frac{1}{2} \mid Y = \frac{1}{3})$.

***Solution*:** By definition of the marginal density. for $0 < x < 2$,

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\ dy = \int_0^1 \frac{x(1 + 3y^2)}{4} dy$$
$$= \left(\frac{xy}{4} + \frac{xy^3}{4}\right)\Bigg|_{y=0}^{y=1} = \frac{x}{2},$$

and for $0 < y < 1$,

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\ dx = \int_0^2 \frac{x(1 + 3y^2)}{4} dx$$
$$= \left(\frac{x^2}{8} + \frac{3x^2y^2}{8}\right)\Bigg|_{x=0}^{x=2} = \frac{1 + 3y^2}{2}.$$

Therefore, using the conditional density definition, for $0 < x < 2$,

$$f(x|y) = \frac{f(x, y)}{h(y)} = \frac{x(1 + 3y^2)/4}{(1 + 3y^2)/2} = \frac{x}{2},$$

and

$$P\left(\frac{1}{4} < X < \frac{1}{2} \;\Bigg|\; Y = \frac{1}{3}\right) = \int_{1/4}^{1/2} \frac{x}{2} dx = \frac{3}{64}.$$

Statistical Independence

If $f(x|y)$ does not depend on y, as is the case for Example 3.20, then $f(x|y) = g(x)$ and $f(x, y) = g(x)h(y)$. The proof follows by substituting

$$f(x, y) = f(x|y)h(y)$$

into the marginal distribution of X. That is,

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\ dy = \int_{-\infty}^{\infty} f(x|y)h(y)\ dy.$$

If $f(x|y)$ does not depend on y, we may write

$$g(x) = f(x|y) \int_{-\infty}^{\infty} h(y)\ dy.$$

Now

$$\int_{-\infty}^{\infty} h(y)\ dy = 1,$$

since $h(y)$ is the probability density function of Y. Therefore,

$$g(x) = f(x|y) \quad \text{and then} \quad f(x, y) = g(x)h(y).$$

It should make sense to the reader that if $f(x|y)$ does not depend on y, then of course the outcome of the random variable Y has no impact on the outcome of the random variable X. In other words, we say that X and Y are independent random variables. We now offer the following formal definition of statistical independence.

Definition 3.12: Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x, y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be **statistically independent** if and only if

$$f(x, y) = g(x)h(y)$$

for all (x, y) within their range.

The continuous random variables of Example 3.20 are statistically independent, since the product of the two marginal distributions gives the joint density function. This is obviously not the case, however, for the continuous variables of Example 3.19. Checking for statistical independence of discrete random variables requires a more thorough investigation, since it is possible to have the product of the marginal distributions equal to the joint probability distribution for some but not all combinations of (x, y). If you can find any point (x, y) for which $f(x, y)$ is defined such that $f(x, y) \neq g(x)h(y)$, the discrete variables X and Y are not statistically independent.

Example 3.21: Show that the random variables of Example 3.14 are not statistically independent.

***Proof*:** Let us consider the point $(0, 1)$. From Table 3.1 we find the three probabilities $f(0, 1)$, $g(0)$, and $h(1)$ to be

$$\begin{aligned} f(0,1) &= \frac{3}{14}, \\ g(0) &= \sum_{y=0}^{2} f(0, y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}, \\ h(1) &= \sum_{x=0}^{2} f(x, 1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}. \end{aligned}$$

Clearly,

$$f(0, 1) \neq g(0)h(1),$$

and therefore X and Y are not statistically independent. ∎

All the preceding definitions concerning two random variables can be generalized to the case of n random variables. Let $f(x_1, x_2, \ldots, x_n)$ be the joint probability function of the random variables $X_1, X_2, \ldots, X_n$. The marginal distribution of X_1, for example, is

$$g(x_1) = \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \ldots, x_n)$$

for the discrete case, and

$$g(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_2\, dx_3 \cdots dx_n$$

for the continuous case. We can now obtain **joint marginal distributions** such as $g(x_1, x_2)$, where

$$g(x_1, x_2) = \begin{cases} \sum\limits_{x_3} \cdots \sum\limits_{x_n} f(x_1, x_2, \ldots, x_n) & \text{(discrete case)}, \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_3\, dx_4 \cdots dx_n & \text{(continuous case)}. \end{cases}$$

We could consider numerous conditional distributions. For example, the **joint conditional distribution** of X_1, X_2, and X_3, given that $X_4 = x_4, X_5 = x_5, \ldots, X_n = x_n$, is written

$$f(x_1, x_2, x_3 \mid x_4, x_5, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n)}{g(x_4, x_5, \ldots, x_n)},$$

where $g(x_4, x_5, \ldots, x_n)$ is the joint marginal distribution of the random variables $X_4, X_5, \ldots, X_n$.

A generalization of Definition 3.12 leads to the following definition for the mutual statistical independence of the variables $X_1, X_2, \ldots, X_n$.

Definition 3.13: Let $X_1, X_2, \ldots, X_n$ be n random variables, discrete or continuous, with joint probability distribution $f(x_1, x_2, \ldots, x_n)$ and marginal distribution $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$, respectively. The random variables $X_1, X_2, \ldots, X_n$ are said to be mutually **statistically independent** if and only if

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n)$$

for all $(x_1, x_2, \ldots, x_n)$ within their range.

Example 3.22: Suppose that the shelf life, in years, of a certain perishable food product packaged in cardboard containers is a random variable whose probability density function is given by

$$f(x) = \begin{cases} e^{-x}, & x > 0, \\ 0, & \text{elsewhere}. \end{cases}$$

Let X_1, X_2, and X_3 represent the shelf lives for three of these containers selected independently and find $P(X_1 < 2, 1 < X_2 < 3, X_3 > 2)$.

Solution: Since the containers were selected independently, we can assume that the random variables X_1, X_2, and X_3 are statistically independent, having the joint probability density

$$f(x_1, x_2, x_3) = f(x_1) f(x_2) f(x_3) = e^{-x_1} e^{-x_2} e^{-x_3} = e^{-x_1 - x_2 - x_3},$$

for $x_1 > 0$, $x_2 > 0$, $x_3 > 0$, and $f(x_1, x_2, x_3) = 0$ elsewhere. Hence

$$\begin{aligned} P(X_1 < 2, 1 < X_2 < 3, X_3 > 2) &= \int_2^{\infty} \int_1^3 \int_0^2 e^{-x_1 - x_2 - x_3}\, dx_1\, dx_2\, dx_3 \\ &= (1 - e^{-2})(e^{-1} - e^{-3}) e^{-2} = 0.0372. \end{aligned}$$

What Are Important Characteristics of Probability Distributions and Where Do They Come From?

This is an important point in the text to provide the reader with a transition into the next three chapters. We have given illustrations in both examples and exercises of practical scientific and engineering situations in which probability distributions and their properties are used to solve important problems. These probability distributions, either discrete or continuous, were introduced through phrases like "it is known that" or "suppose that" or even in some cases "historical evidence suggests that." These are situations in which the nature of the distribution and even a good estimate of the probability structure can be determined through historical data, data from long-term studies, or even large amounts of planned data. The reader should remember the discussion of the use of histograms in Chapter 1 and from that recall how frequency distributions are estimated from the histograms. However, not all probability functions and probability density functions are derived from large amounts of historical data. There are a substantial number of situations in which the nature of the scientific scenario suggests a distribution type. Indeed, many of these are reflected in exercises in both Chapter 2 and this chapter. When independent repeated observations are binary in nature (e.g., defective or not, survive or not, allergic or not) with value 0 or 1, the distribution covering this situation is called the **binomial distribution** and the probability function is known and will be demonstrated in its generality in Chapter 5. Exercise 3.34 in Section 3.3 and Review Exercise 3.80 are examples, and there are others that the reader should recognize. The scenario of a continuous distribution in time to failure, as in Review Exercise 3.69 or Exercise 3.27 on page 93, often suggests a distribution type called the **exponential distribution**. These types of illustrations are merely two of many so-called standard distributions that are used extensively in real-world problems because the scientific scenario that gives rise to each of them is recognizable and occurs often in practice. Chapters 5 and 6 cover many of these types along with some underlying theory concerning their use.

A second part of this transition to material in future chapters deals with the notion of **population parameters** or **distributional parameters**. Recall in Chapter 1 we discussed the need to use data to provide information about these parameters. We went to some length in discussing the notions of a **mean** and **variance** and provided a vision for the concepts in the context of a population. Indeed, the population mean and variance are easily found from the probability function for the discrete case or probability density function for the continuous case. These parameters and their importance in the solution of many types of real-world problems will provide much of the material in Chapters 8 through 17.

Exercises

3.37 Determine the values of c so that the following functions represent joint probability distributions of the random variables X and Y:

(a) $f(x, y) = cxy$, for $x = 1, 2, 3$; $y = 1, 2, 3$;

(b) $f(x, y) = c|x - y|$, for $x = -2, 0, 2$; $y = -2, 3$.

3.38 If the joint probability distribution of X and Y is given by

$$f(x, y) = \frac{x + y}{30}, \quad \text{for } x = 0, 1, 2, 3;\ y = 0, 1, 2,$$

find

(a) $P(X \leq 2, Y = 1)$;
(b) $P(X > 2, Y \leq 1)$;
(c) $P(X > Y)$;
(d) $P(X + Y = 4)$.

3.39 From a sack of fruit containing 3 oranges, 2 apples, and 3 bananas, a random sample of 4 pieces of fruit is selected. If X is the number of oranges and Y is the number of apples in the sample, find

(a) the joint probability distribution of X and Y;
(b) $P[(X, Y) \in A]$, where A is the region that is given by $\{(x, y) \mid x + y \leq 2\}$.

3.40 A fast-food restaurant operates both a drive-through facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-through and walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x, y) = \begin{cases} \frac{2}{3}(x + 2y), & 0 \leq x \leq 1,\ 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the marginal density of X.
(b) Find the marginal density of Y.
(c) Find the probability that the drive-through facility is busy less than one-half of the time.

3.41 A candy company distributes boxes of chocolates with a mixture of creams, toffees, and cordials. Suppose that the weight of each box is 1 kilogram, but the individual weights of the creams, toffees, and cordials vary from box to box. For a randomly selected box, let X and Y represent the weights of the creams and the toffees, respectively, and suppose that the joint density function of these variables is

$$f(x, y) = \begin{cases} 24xy, & 0 \leq x \leq 1,\ 0 \leq y \leq 1,\ x + y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the probability that in a given box the cordials account for more than 1/2 of the weight.
(b) Find the marginal density for the weight of the creams.
(c) Find the probability that the weight of the toffees in a box is less than 1/8 of a kilogram if it is known that creams constitute 3/4 of the weight.

3.42 Let X and Y denote the lengths of life, in years, of two components in an electronic system. If the joint density function of these variables is

$$f(x, y) = \begin{cases} e^{-(x+y)}, & x > 0,\ y > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

find $P(0 < X < 1 \mid Y = 2)$.

3.43 Let X denote the reaction time, in seconds, to a certain stimulus and Y denote the temperature (°F) at which a certain reaction starts to take place. Suppose that two random variables X and Y have the joint density

$$f(x, y) = \begin{cases} 4xy, & 0 < x < 1,\ 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find

(a) $P(0 \leq X \leq \frac{1}{2} \text{ and } \frac{1}{4} \leq Y \leq \frac{1}{2})$;
(b) $P(X < Y)$.

3.44 Each rear tire on an experimental airplane is supposed to be filled to a pressure of 40 pounds per square inch (psi). Let X denote the actual air pressure for the right tire and Y denote the actual air pressure for the left tire. Suppose that X and Y are random variables with the joint density function

$$f(x, y) = \begin{cases} k(x^2 + y^2), & 30 \leq x < 50,\ 30 \leq y < 50, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find k.
(b) Find $P(30 \leq X \leq 40 \text{ and } 40 \leq Y < 50)$.
(c) Find the probability that both tires are underfilled.

3.45 Let X denote the diameter of an armored electric cable and Y denote the diameter of the ceramic mold that makes the cable. Both X and Y are scaled so that they range between 0 and 1. Suppose that X and Y have the joint density

$$f(x, y) = \begin{cases} \frac{1}{y}, & 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find $P(X + Y > 1/2)$.

3.46 Referring to Exercise 3.38, find

(a) the marginal distribution of X;
(b) the marginal distribution of Y.

3.47 The amount of kerosene, in thousands of liters, in a tank at the beginning of any day is a random amount Y from which a random amount X is sold during that day. Suppose that the tank is not resupplied during the day so that $x \leq y$, and assume that the joint density function of these variables is

$$f(x, y) = \begin{cases} 2, & 0 < x \leq y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Determine if X and Y are independent.

(b) Find $P(1/4 < X < 1/2 \mid Y = 3/4)$.

3.48 Referring to Exercise 3.39, find
(a) $f(y|2)$ for all values of y;
(b) $P(Y = 0 \mid X = 2)$.

3.49 Let X denote the number of times a certain numerical control machine will malfunction: 1, 2, or 3 times on any given day. Let Y denote the number of times a technician is called on an emergency call. Their joint probability distribution is given as

	$f(x,y)$	x = 1	2	3
	1	0.05	0.05	0.10
y	3	0.05	0.10	0.35
	5	0.00	0.20	0.10

(a) Evaluate the marginal distribution of X.
(b) Evaluate the marginal distribution of Y.
(c) Find $P(Y = 3 \mid X = 2)$.

3.50 Suppose that X and Y have the following joint probability distribution:

	$f(x,y)$	x = 2	4
	1	0.10	0.15
y	3	0.20	0.30
	5	0.10	0.15

(a) Find the marginal distribution of X.
(b) Find the marginal distribution of Y.

3.51 Three cards are drawn without replacement from the 12 face cards (jacks, queens, and kings) of an ordinary deck of 52 playing cards. Let X be the number of kings selected and Y the number of jacks. Find
(a) the joint probability distribution of X and Y;
(b) $P[(X, Y) \in A]$, where A is the region given by $\{(x, y) \mid x + y \geq 2\}$.

3.52 A coin is tossed twice. Let Z denote the number of heads on the first toss and W the total number of heads on the 2 tosses. If the coin is unbalanced and a head has a 40% chance of occurring, find
(a) the joint probability distribution of W and Z;
(b) the marginal distribution of W;
(c) the marginal distribution of Z;
(d) the probability that at least 1 head occurs.

3.53 Given the joint density function

$$f(x,y) = \begin{cases} \frac{6-x-y}{8}, & 0 < x < 2,\ 2 < y < 4, \\ 0, & \text{elsewhere}, \end{cases}$$

find $P(1 < Y < 3 \mid X = 1)$.

3.54 Determine whether the two random variables of Exercise 3.49 are dependent or independent.

3.55 Determine whether the two random variables of Exercise 3.50 are dependent or independent.

3.56 The joint density function of the random variables X and Y is

$$f(x,y) = \begin{cases} 6x, & 0 < x < 1,\ 0 < y < 1 - x, \\ 0, & \text{elsewhere}. \end{cases}$$

(a) Show that X and Y are not independent.
(b) Find $P(X > 0.3 \mid Y = 0.5)$.

3.57 Let X, Y, and Z have the joint probability density function

$$f(x,y,z) = \begin{cases} kxy^2z, & 0 < x, y < 1,\ 0 < z < 2, \\ 0, & \text{elsewhere}. \end{cases}$$

(a) Find k.
(b) Find $P(X < \frac{1}{4}, Y > \frac{1}{2}, 1 < Z < 2)$.

3.58 Determine whether the two random variables of Exercise 3.43 are dependent or independent.

3.59 Determine whether the two random variables of Exercise 3.44 are dependent or independent.

3.60 The joint probability density function of the random variables X, Y, and Z is

$$f(x,y,z) = \begin{cases} \frac{4xyz^2}{9}, & 0 < x, y < 1,\ 0 < z < 3, \\ 0, & \text{elsewhere}. \end{cases}$$

Find
(a) the joint marginal density function of Y and Z;
(b) the marginal density of Y;
(c) $P(\frac{1}{4} < X < \frac{1}{2},\ Y > \frac{1}{3},\ 1 < Z < 2)$;
(d) $P(0 < X < \frac{1}{2} \mid Y = \frac{1}{4},\ Z = 2)$.

Review Exercises

3.61 A tobacco company produces blends of tobacco, with each blend containing various proportions of Turkish, domestic, and other tobaccos. The proportions of Turkish and domestic in a blend are random variables with joint density function (X = Turkish and Y = domestic)

$$f(x, y) = \begin{cases} 24xy, & 0 \le x, y \le 1, \ x + y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the probability that in a given box the Turkish tobacco accounts for over half the blend.

(b) Find the marginal density function for the proportion of the domestic tobacco.

(c) Find the probability that the proportion of Turkish tobacco is less than 1/8 if it is known that the blend contains 3/4 domestic tobacco.

3.62 An insurance company offers its policyholders a number of different premium payment options. For a randomly selected policyholder, let X be the number of months between successive payments. The cumulative distribution function of X is

$$F(x) = \begin{cases} 0, & \text{if } x < 1, \\ 0.4, & \text{if } 1 \le x < 3, \\ 0.6, & \text{if } 3 \le x < 5, \\ 0.8, & \text{if } 5 \le x < 7, \\ 1.0, & \text{if } x \ge 7. \end{cases}$$

(a) What is the probability mass function of X?

(b) Compute $P(4 < X \le 7)$.

3.63 Two electronic components of a missile system work in harmony for the success of the total system. Let X and Y denote the life in hours of the two components. The joint density of X and Y is

$$f(x, y) = \begin{cases} ye^{-y(1+x)}, & x, y \ge 0, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Give the marginal density functions for both random variables.

(b) What is the probability that the lives of both components will exceed 2 hours?

3.64 A service facility operates with two service lines. On a randomly selected day, let X be the proportion of time that the first line is in use whereas Y is the proportion of time that the second line is in use. Suppose that the joint probability density function for (X, Y) is

$$f(x, y) = \begin{cases} \frac{3}{2}(x^2 + y^2), & 0 \le x, y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Compute the probability that neither line is busy more than half the time.

(b) Find the probability that the first line is busy more than 75% of the time.

3.65 Let the number of phone calls received by a switchboard during a 5-minute interval be a random variable X with probability function

$$f(x) = \frac{e^{-2}2^x}{x!}, \quad \text{for } x = 0, 1, 2, \ldots.$$

(a) Determine the probability that X equals 0, 1, 2, 3, 4, 5, and 6.

(b) Graph the probability mass function for these values of x.

(c) Determine the cumulative distribution function for these values of X.

3.66 Consider the random variables X and Y with joint density function

$$f(x, y) = \begin{cases} x + y, & 0 \le x, y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the marginal distributions of X and Y.

(b) Find $P(X > 0.5, Y > 0.5)$.

3.67 An industrial process manufactures items that can be classified as either defective or not defective. The probability that an item is defective is 0.1. An experiment is conducted in which 5 items are drawn randomly from the process. Let the random variable X be the number of defectives in this sample of 5. What is the probability mass function of X?

3.68 Consider the following joint probability density function of the random variables X and Y:

$$f(x, y) = \begin{cases} \frac{3x-y}{9}, & 1 < x < 3, \ 1 < y < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the marginal density functions of X and Y.

(b) Are X and Y independent?

(c) Find $P(X > 2)$.

3.69 The life span in hours of an electrical component is a random variable with cumulative distribution function

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{50}}, & x > 0, \\ 0, & \text{eleswhere.} \end{cases}$$

(a) Determine its probability density function.

(b) Determine the probability that the life span of such a component will exceed 70 hours.

3.70 Pairs of pants are being produced by a particular outlet facility. The pants are checked by a group of 10 workers. The workers inspect pairs of pants taken randomly from the production line. Each inspector is assigned a number from 1 through 10. A buyer selects a pair of pants for purchase. Let the random variable X be the inspector number.

(a) Give a reasonable probability mass function for X.

(b) Plot the cumulative distribution function for X.

3.71 The shelf life of a product is a random variable that is related to consumer acceptance. It turns out that the shelf life Y in days of a certain type of bakery product has a density function

$$f(y) = \begin{cases} \frac{1}{2}e^{-y/2}, & 0 \le y < \infty, \\ 0, & \text{elsewhere.} \end{cases}$$

What fraction of the loaves of this product stocked today would you expect to be sellable 3 days from now?

3.72 Passenger congestion is a service problem in airports. Trains are installed within the airport to reduce the congestion. With the use of the train, the time X in minutes that it takes to travel from the main terminal to a particular concourse has density function

$$f(x) = \begin{cases} \frac{1}{10}, & 0 \le x \le 10, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Show that the above is a valid probability density function.

(b) Find the probability that the time it takes a passenger to travel from the main terminal to the concourse will not exceed 7 minutes.

3.73 Impurities in a batch of final product of a chemical process often reflect a serious problem. From considerable plant data gathered, it is known that the proportion Y of impurities in a batch has a density function given by

$$f(y) = \begin{cases} 10(1-y)^9, & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that the above is a valid density function.

(b) A batch is considered not sellable and then not acceptable if the percentage of impurities exceeds 60%. With the current quality of the process, what is the percentage of batches that are not acceptable?

3.74 The time Z in minutes between calls to an electrical supply system has the probability density function

$$f(z) = \begin{cases} \frac{1}{10}e^{-z/10}, & 0 < z < \infty, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) What is the probability that there are no calls within a 20-minute time interval?

(b) What is the probability that the first call comes within 10 minutes of opening?

3.75 A chemical system that results from a chemical reaction has two important components among others in a blend. The joint distribution describing the proportions X_1 and X_2 of these two components is given by

$$f(x_1, x_2) = \begin{cases} 2, & 0 < x_1 < x_2 < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Give the marginal distribution of X_1.

(b) Give the marginal distribution of X_2.

(c) What is the probability that component proportions produce the results $X_1 < 0.2$ and $X_2 > 0.5$?

(d) Give the conditional distribution $f_{X_1|X_2}(x_1|x_2)$.

3.76 Consider the situation of Review Exercise 3.75. But suppose the joint distribution of the two proportions is given by

$$f(x_1, x_2) = \begin{cases} 6x_2, & 0 < x_2 < x_1 < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Give the marginal distribution $f_{X_1}(x_1)$ of the proportion X_1 and verify that it is a valid density function.

(b) What is the probability that proportion X_2 is less than 0.5, given that X_1 is 0.7?

3.77 Consider the random variables X and Y that represent the number of vehicles that arrive at two separate street corners during a certain 2-minute period. These street corners are fairly close together so it is important that traffic engineers deal with them jointly if necessary. The joint distribution of X and Y is known to be

$$f(x, y) = \frac{9}{16} \cdot \frac{1}{4^{(x+y)}},$$

for $x = 0, 1, 2, \ldots$ and $y = 0, 1, 2, \ldots$.

(a) Are the two random variables X and Y independent? Explain why or why not.

(b) What is the probability that during the time period in question less than 4 vehicles arrive at the two street corners?

3.78 The behavior of series of components plays a huge role in scientific and engineering reliability problems. The reliability of the entire system is certainly no better than that of the weakest component in the series. In a series system, the components operate independently of each other. In a particular system containing three components, the probabilities of meeting specifications for components 1, 2, and 3, respectively, are 0.95, 0.99, and 0.92. What is the probability that the entire system works?

3.79 Another type of system that is employed in engineering work is a group of parallel components or a parallel system. In this more conservative approach, the probability that the system operates is larger than the probability that any component operates. The system fails only when all components fail. Consider a situation in which there are 4 independent components in a parallel system with probability of operation given by

Component 1: 0.95; Component 2: 0.94;
Component 3: 0.90; Component 4: 0.97.

What is the probability that the system does not fail?

3.80 Consider a system of components in which there are 5 independent components, each of which possesses an operational probability of 0.92. The system does have a redundancy built in such that it does not fail if 3 out of the 5 components are operational. What is the probability that the total system is operational?

3.81 **Project**: Take 5 class periods to observe the shoe color of individuals in class. Assume the shoe color categories are red, white, black, brown, and other. Complete a frequency table for each color category.

(a) Estimate and interpret the meaning of the probability distribution.

(b) What is the estimated probability that in the next class period a randomly selected student will be wearing a red or a white pair of shoes?

3.5 Potential Misconceptions and Hazards; Relationship to Material in Other Chapters

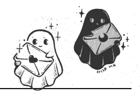
In future chapters it will become apparent that probability distributions represent the structure through which probabilities that are computed aid in the evaluation and understanding of a process. For example, in Review Exercise 3.65, the probability distribution that quantifies the probability of a heavy load during certain time periods can be very useful in planning for any changes in the system. Review Exercise 3.69 describes a scenario in which the life span of an electronic component is studied. Knowledge of the probability structure for the component will contribute significantly to an understanding of the reliability of a large system of which the component is a part. In addition, an understanding of the general nature of probability distributions will enhance understanding of the concept of a ***P*-value**, which was introduced briefly in Chapter 1 and will play a major role beginning in Chapter 10 and extending throughout the balance of the text.

Chapters 4, 5, and 6 depend heavily on the material in this chapter. In Chapter 4, we discuss the meaning of important **parameters** in probability distributions. These important parameters quantify notions of **central tendency** and **variability** in a system. In fact, knowledge of these quantities themselves, quite apart from the complete distribution, can provide insight into the nature of the system. Chapters 5 and 6 will deal with engineering, biological, or general scientific scenarios that identify special types of distributions. For example, the structure of the probability function in Review Exercise 3.65 will easily be identified under certain assumptions discussed in Chapter 5. The same holds for the scenario of Review Exercise 3.69. This is a special type of **time to failure** problem for which the probability density function will be discussed in Chapter 6.

As far as potential hazards with the use of material in this chapter, the warning to the reader is not to read more into the material than is evident. The general nature of the probability distribution for a specific scientific phenomenon is not obvious from what is learned in this chapter. The purpose of this chapter is for readers to learn how to manipulate a probability distribution, not to learn how to identify a specific type. Chapters 5 and 6 go a long way toward identification according to the general nature of the scientific system.

Chapter 4

Mathematical Expectation



4.1 Mean of a Random Variable

In Chapter 1, we discussed the sample mean, which is the arithmetic mean of the data. Now consider the following. If two coins are tossed 16 times and X is the number of heads that occur per toss, then the values of X are 0, 1, and 2. Suppose that the experiment yields no heads, one head, and two heads a total of 4, 7, and 5 times, respectively. The average number of heads per toss of the two coins is then

$$\frac{(0)(4) + (1)(7) + (2)(5)}{16} = 1.06.$$

This is an average value of the data and yet it is not a possible outcome of $\{0, 1, 2\}$. Hence, an average is not necessarily a possible outcome for the experiment. For instance, a salesman's average monthly income is not likely to be equal to any of his monthly paychecks.

Let us now restructure our computation for the average number of heads so as to have the following equivalent form:

$$(0)\left(\frac{4}{16}\right) + (1)\left(\frac{7}{16}\right) + (2)\left(\frac{5}{16}\right) = 1.06.$$

The numbers $4/16$, $7/16$, and $5/16$ are the fractions of the total tosses resulting in 0, 1, and 2 heads, respectively. These fractions are also the relative frequencies for the different values of X in our experiment. In fact, then, we can calculate the mean, or average, of a set of data by knowing the distinct values that occur and their relative frequencies, without any knowledge of the total number of observations in our set of data. Therefore, if $4/16$, or $1/4$, of the tosses result in no heads, $7/16$ of the tosses result in one head, and $5/16$ of the tosses result in two heads, the mean number of heads per toss would be 1.06 no matter whether the total number of tosses were 16, 1000, or even 10,000.

This method of relative frequencies is used to calculate the average number of heads per toss of two coins that we might expect in the long run. We shall refer to this average value as the **mean of the random variable** X or the **mean of the probability distribution of** X and write it as μ_x or simply as μ when it is clear to which random variable we refer. It is also common among statisticians to refer to this mean as the mathematical expectation, or the expected value of the random variable X, and denote it as $E(X)$.

Assuming that 1 fair coin was tossed twice, we find that the sample space for our experiment is

$$S = \{HH, HT, TH, TT\}.$$

Since the 4 sample points are all equally likely, it follows that

$$P(X = 0) = P(TT) = \frac{1}{4}, \quad P(X = 1) = P(TH) + P(HT) = \frac{1}{2},$$

and

$$P(X = 2) = P(HH) = \frac{1}{4},$$

where a typical element, say TH, indicates that the first toss resulted in a tail followed by a head on the second toss. Now, these probabilities are just the relative frequencies for the given events in the long run. Therefore,

$$\mu = E(X) = (0)\left(\frac{1}{4}\right) + (1)\left(\frac{1}{2}\right) + (2)\left(\frac{1}{4}\right) = 1.$$

This result means that a person who tosses 2 coins over and over again will, on the average, get 1 head per toss.

The method described above for calculating the expected number of heads per toss of 2 coins suggests that the mean, or expected value, of any discrete random variable may be obtained by multiplying each of the values $x_1, x_2, \ldots, x_n$ of the random variable X by its corresponding probability $f(x_1), f(x_2), \ldots, f(x_n)$ and summing the products. This is true, however, only if the random variable is discrete. In the case of continuous random variables, the definition of an expected value is essentially the same with summations replaced by integrations.

Definition 4.1: Let X be a random variable with probability distribution $f(x)$. The **mean**, or **expected value**, of X is

$$\mu = E(X) = \sum_x x f(x)$$

if X is discrete, and

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

if X is continuous.

The reader should note that the way to calculate the expected value, or mean, shown here is different from the way to calculate the sample mean described in Chapter 1, where the sample mean is obtained by using data. In mathematical expectation, the expected value is calculated by using the probability distribution.

However, the mean is usually understood as a "center" value of the underlying distribution if we use the expected value, as in Definition 4.1.

Example 4.1: A lot containing 7 components is sampled by a quality inspector; the lot contains 4 good components and 3 defective components. A sample of 3 is taken by the inspector. Find the expected value of the number of good components in this sample.

***Solution*:** Let X represent the number of good components in the sample. The probability distribution of X is

$$f(x) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, \qquad x = 0, 1, 2, 3.$$

Simple calculations yield $f(0) = 1/35$, $f(1) = 12/35$, $f(2) = 18/35$, and $f(3) = 4/35$. Therefore,

$$\mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7} = 1.7.$$

Thus, if a sample of size 3 is selected at random over and over again from a lot of 4 good components and 3 defective components, it will contain, on average, 1.7 good components.

Example 4.2: A salesperson for a medical device company has two appointments on a given day. At the first appointment, he believes that he has a 70% chance to make the deal, from which he can earn \$1000 commission if successful. On the other hand, he thinks he only has a 40% chance to make the deal at the second appointment, from which, if successful, he can make \$1500. What is his expected commission based on his own probability belief? Assume that the appointment results are independent of each other.

***Solution*:** First, we know that the salesperson, for the two appointments, can have 4 possible commission totals: \$0, \$1000, \$1500, and \$2500. We then need to calculate their associated probabilities. By independence, we obtain

$$f(\$0) = (1-0.7)(1-0.4) = 0.18, \quad f(\$2500) = (0.7)(0.4) = 0.28,$$
$$f(\$1000) = (0.7)(1-0.4) = 0.42, \text{ and } f(\$1500) = (1-0.7)(0.4) = 0.12.$$

Therefore, the expected commission for the salesperson is

$$\begin{aligned} E(X) &= (\$0)(0.18) + (\$1000)(0.42) + (\$1500)(0.12) + (\$2500)(0.28) \\ &= \$1300. \end{aligned}$$

Examples 4.1 and 4.2 are designed to allow the reader to gain some insight into what we mean by the expected value of a random variable. In both cases the random variables are discrete. We follow with an example involving a continuous random variable, where an engineer is interested in the *mean life* of a certain type of electronic device. This is an illustration of a *time to failure* problem that occurs often in practice. The expected value of the life of a device is an important parameter for its evaluation.

Example 4.3: Let X be the random variable that denotes the life in hours of a certain electronic device. The probability density function is

$$f(x) = \begin{cases} \frac{20{,}000}{x^3}, & x > 100, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected life of this type of device.

Solution: Using Definition 4.1, we have

average

$$\mu = E(X) = \int_{100}^{\infty} x \frac{20,000}{x^3}\, dx = \int_{100}^{\infty} \frac{20,000}{x^2}\, dx = 200.$$

Therefore, we can expect this type of device to last, *on average*, 200 hours.

Now let us consider a new random variable $g(X)$, which depends on X; that is, each value of $g(X)$ is determined by the value of X. For instance, $g(X)$ might be X^2 or $3X - 1$, and whenever X assumes the value 2, $g(X)$ assumes the value $g(2)$. In particular, if X is a discrete random variable with probability distribution $f(x)$, for $x = -1, 0, 1, 2$, and $g(X) = X^2$, then

$$\begin{aligned} P[g(X) = 0] &= P(X = 0) = f(0), \\ P[g(X) = 1] &= P(X = -1) + P(X = 1) = f(-1) + f(1), \\ P[g(X) = 4] &= P(X = 2) = f(2), \end{aligned}$$

and so the probability distribution of $g(X)$ may be written

$g(x)$	0	1	4
$P[g(X) = g(x)]$	$f(0)$	$f(-1) + f(1)$	$f(2)$

By the definition of the expected value of a random variable, we obtain

$$\begin{aligned} \mu_{g(X)} &= E[g(x)] = 0f(0) + 1[f(-1) + f(1)] + 4f(2) \\ &= (-1)^2 f(-1) + (0)^2 f(0) + (1)^2 f(1) + (2)^2 f(2) = \sum_x g(x)f(x). \end{aligned}$$

This result is generalized in Theorem 4.1 for both discrete and continuous random variables.

Theorem 4.1: Let X be a random variable with probability distribution $f(x)$. The expected value of the random variable $g(X)$ is

$$\mu_{g(X)} = E[g(X)] = \sum_x g(x)f(x)$$

if X is discrete, and

$$\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\, dx$$

if X is continuous.

Example 4.4: Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

x	4	5	6	7	8	9
$P(X = x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Let $g(X) = 2X-1$ represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Solution: By Theorem 4.1, the attendant can expect to receive

$$\begin{aligned} E[g(X)] = E(2X-1) &= \sum_{x=4}^{9}(2x-1)f(x) \\ &= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right) \\ &\quad + (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right) = \$12.67. \end{aligned}$$

Example 4.5: Let X be a random variable with density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of $g(X) = 4X + 3$.

Solution: By Theorem 4.1, we have

$$E(4X+3) = \int_{-1}^{2} \frac{(4x+3)x^2}{3}\, dx = \frac{1}{3}\int_{-1}^{2}(4x^3 + 3x^2)\, dx = 8.$$

We shall now extend our concept of mathematical expectation to the case of two random variables X and Y with joint probability distribution $f(x, y)$.

Definition 4.2: Let X and Y be random variables with joint probability distribution $f(x, y)$. The mean, or expected value, of the random variable $g(X, Y)$ is

$$\mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x,y)f(x,y)$$

if X and Y are discrete, and

$$\mu_{g(X,Y)} = E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\, dx\, dy$$

if X and Y are continuous.

Generalization of Definition 4.2 for the calculation of mathematical expectations of functions of several random variables is straightforward.

Example 4.6: Let X and Y be the random variables with joint probability distribution indicated in Table 3.1 on page 96. Find the expected value of $g(X,Y) = XY$. The table is reprinted here for convenience.

	$f(x,y)$	$x=0$	$x=1$	$x=2$	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Solution: By Definition 4.2, we write

$$\begin{aligned}E(XY) &= \sum_{x=0}^{2}\sum_{y=0}^{2} xyf(x,y)\\ &= (0)(0)f(0,0) + (0)(1)f(0,1)\\ &\quad + (1)(0)f(1,0) + (1)(1)f(1,1) + (2)(0)f(2,0)\\ &= f(1,1) = \frac{3}{14}.\end{aligned}$$

Example 4.7: Find $E(Y/X)$ for the density function

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1,\\ 0, & \text{elsewhere.}\end{cases}$$

Solution: We have

$$E\left(\frac{Y}{X}\right) = \int_0^1\int_0^2 \frac{y(1+3y^2)}{4}\,dxdy = \int_0^1 \frac{y+3y^3}{2}\,dy = \frac{5}{8}.$$

Note that if $g(X,Y) = X$ in Definition 4.2, we have

$$E(X) = \begin{cases} \sum\limits_x\sum\limits_y xf(x,y) = \sum\limits_x xg(x) & \text{(discrete case)},\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\ dy\ dx = \int_{-\infty}^{\infty} xg(x)\ dx & \text{(continuous case)},\end{cases}$$

where $g(x)$ is the marginal distribution of X. Therefore, in calculating $E(X)$ over a two-dimensional space, one may use either the joint probability distribution of X and Y or the marginal distribution of X. Similarly, we define

$$E(Y) = \begin{cases} \sum\limits_y\sum\limits_x yf(x,y) = \sum\limits_y yh(y) & \text{(discrete case)},\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\ dxdy = \int_{-\infty}^{\infty} yh(y)\ dy & \text{(continuous case)},\end{cases}$$

where $h(y)$ is the marginal distribution of the random variable Y.

Exercises

4.1 The probability distribution of X, the number of imperfections per 10 meters of a synthetic fabric in continuous rolls of uniform width, is given in Exercise 3.13 on page 92 as

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Find the average number of imperfections per 10 meters of this fabric.

4.2 The probability distribution of the discrete random variable X is

$$f(x) = \binom{3}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{3-x}, \quad x = 0, 1, 2, 3.$$

Find the mean of X.

4.3 Find the mean of the random variable T representing the total of the three coins in Exercise 3.25 on page 93.

4.4 A coin is biased such that a head is three times as likely to occur as a tail. Find the expected number of tails when this coin is tossed twice.

4.5 In a gambling game, a woman is paid \$3 if she draws a jack or a queen and \$5 if she draws a king or an ace from an ordinary deck of 52 playing cards. If she draws any other card, she loses. How much should she pay to play if the game is fair?

4.6 An attendant at a car wash is paid according to the number of cars that pass through. Suppose the probabilities are 1/12, 1/12, 1/4, 1/4, 1/6, and 1/6, respectively, that the attendant receives \$7, \$9, \$11, \$13, \$15, or \$17 between 4:00 P.M. and 5:00 P.M. on any sunny Friday. Find the attendant's expected earnings for this particular period.

4.7 By investing in a particular stock, a person can make a profit in one year of \$4000 with probability 0.3 or take a loss of \$1000 with probability 0.7. What is this person's expected gain?

4.8 Suppose that an antique jewelry dealer is interested in purchasing a gold necklace for which the probabilities are 0.22, 0.36, 0.28, and 0.14, respectively, that she will be able to sell it for a profit of \$250, sell it for a profit of \$150, break even, or sell it for a loss of \$150. What is her expected profit?

4.9 A private pilot wishes to insure his airplane for \$200,000. The insurance company estimates that a total loss will occur with probability 0.002, a 50% loss with probability 0.01, and a 25% loss with probability 0.1. Ignoring all other partial losses, what premium should the insurance company charge each year to realize an average profit of \$500?

4.10 Two tire-quality experts examine stacks of tires and assign a quality rating to each tire on a 3-point scale. Let X denote the rating given by expert A and Y denote the rating given by B. The following table gives the joint distribution for X and Y.

$f(x,y)$		$y=1$	$y=2$	$y=3$
x	1	0.10	0.05	0.02
	2	0.10	0.35	0.05
	3	0.03	0.10	0.20

Find μ_X and μ_Y.

4.11 The density function of coded measurements of the pitch diameter of threads of a fitting is

$$f(x) = \begin{cases} \frac{4}{\pi(1+x^2)}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of X.

4.12 If a dealer's profit, in units of \$5000, on a new automobile can be looked upon as a random variable X having the density function

$$f(x) = \begin{cases} 2(1-x), & 0 < x < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

find the average profit per automobile.

4.13 The density function of the continuous random variable X, the total number of hours, in units of 100 hours, that a family runs a vacuum cleaner over a period of one year, is given in Exercise 3.7 on page 92 as

$$f(x) = \begin{cases} x, & 0 < x < 1, \\ 2-x, & 1 \le x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the average number of hours per year that families run their vacuum cleaners.

4.14 Find the proportion X of individuals who can be expected to respond to a certain mail-order solicitation if X has the density function

$$f(x) = \begin{cases} \frac{2(x+2)}{5}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

4.15 Assume that two random variables (X, Y) are uniformly distributed on a circle with radius a. Then the joint probability density function is

$$f(x,y) = \begin{cases} \frac{1}{\pi a^2}, & x^2 + y^2 \le a^2, \\ 0, & \text{otherwise.} \end{cases}$$

Find μ_X, the expected value of X.

4.16 Suppose that you are inspecting a lot of 1000 light bulbs, among which 20 are defectives. You choose two light bulbs randomly from the lot without replacement. Let

$$X_1 = \begin{cases} 1, & \text{if the 1st light bulb is defective,} \\ 0, & \text{otherwise,} \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{if the 2nd light bulb is defective,} \\ 0, & \text{otherwise.} \end{cases}$$

Find the probability that at least one light bulb chosen is defective. [*Hint*: Compute $P(X_1 + X_2 = 1)$.]

4.17 Let X be a random variable with the following probability distribution:

x	-3	6	9
$f(x)$	$1/6$	$1/2$	$1/3$

Find $\mu_{g(X)}$, where $g(X) = (2X + 1)^2$.

4.18 Find the expected value of the random variable $g(X) = X^2$, where X has the probability distribution of Exercise 4.2.

4.19 A large industrial firm purchases several new word processors at the end of each year, the exact number depending on the frequency of repairs in the previous year. Suppose that the number of word processors, X, purchased each year has the following probability distribution:

x	0	1	2	3
$f(x)$	$1/10$	$3/10$	$2/5$	$1/5$

If the cost of the desired model is \$1200 per unit and at the end of the year a refund of $50X^2$ dollars will be issued, how much can this firm expect to spend on new word processors during this year?

4.20 A continuous random variable X has the density function

$$f(x) = \begin{cases} e^{-x}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of $g(X) = e^{2X/3}$.

4.21 What is the dealer's average profit per automobile if the profit on each automobile is given by $g(X) = X^2$, where X is a random variable having the density function of Exercise 4.12?

4.22 The hospitalization period, in days, for patients following treatment for a certain type of kidney disorder is a random variable $Y = X + 4$, where X has the density function

$$f(x) = \begin{cases} \frac{32}{(x+4)^3}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the average number of days that a person is hospitalized following treatment for this disorder.

4.23 Suppose that X and Y have the following joint probability function:

$f(x,y)$		$x=2$	$x=4$
y	1	0.10	0.15
	3	0.20	0.30
	5	0.10	0.15

(a) Find the expected value of $g(X,Y) = XY^2$.

(b) Find μ_X and μ_Y.

4.24 Referring to the random variables whose joint probability distribution is given in Exercise 3.39 on page 105,

(a) find $E(X^2Y - 2XY)$;

(b) find $\mu_X - \mu_Y$.

4.25 Referring to the random variables whose joint probability distribution is given in Exercise 3.51 on page 106, find the mean for the total number of jacks and kings when 3 cards are drawn without replacement from the 12 face cards of an ordinary deck of 52 playing cards.

4.26 Let X and Y be random variables with joint density function

$$f(x,y) = \begin{cases} 4xy, & 0 < x,\ y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of $Z = \sqrt{X^2 + Y^2}$.

4.27 In Exercise 3.27 on page 93, a density function is given for the time to failure of an important component of a DVD player. Find the mean number of hours to failure of the component and thus the DVD player.

4.28 Consider the information in Exercise 3.28 on page 93. The problem deals with the weight in ounces of the product in a cereal box, with

$$f(x) = \begin{cases} \frac{2}{5}, & 23.75 \le x \le 26.25, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Plot the density function.

(b) Compute the expected value, or mean weight, in ounces.

(c) Are you surprised at your answer in (b)? Explain why or why not.

4.29 Exercise 3.29 on page 93 dealt with an important particle size distribution characterized by

$$f(x) = \begin{cases} 3x^{-4}, & x > 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Plot the density function.

(b) Give the mean particle size.

4.30 In Exercise 3.31 on page 94, the distribution of times before a major repair of a washing machine was given as

$$f(y) = \begin{cases} \frac{1}{4}e^{-y/4}, & y \geq 0, \\ 0, & \text{elsewhere.} \end{cases}$$

What is the population mean of the times to repair?

4.31 Consider Exercise 3.32 on page 94.

(a) What is the mean proportion of the budget allocated to environmental and pollution control?

(b) What is the probability that a company selected at random will have allocated to environmental and pollution control a proportion that exceeds the population mean given in (a)?

4.32 In Exercise 3.13 on page 92, the distribution of the number of imperfections per 10 meters of synthetic fabric is given by

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

(a) Plot the probability function.

(b) Find the expected number of imperfections, $E(X) = \mu$.

(c) Find $E(X^2)$.

4.2 Variance and Covariance of Random Variables

The mean, or expected value, of a random variable X is of special importance in statistics because it describes where the probability distribution is centered. By itself, however, the mean does not give an adequate description of the shape of the distribution. We also need to characterize the variability in the distribution. In Figure 4.1, we have the histograms of two discrete probability distributions that have the same mean, $\mu = 2$, but differ considerably in variability, or the dispersion of their observations about the mean.

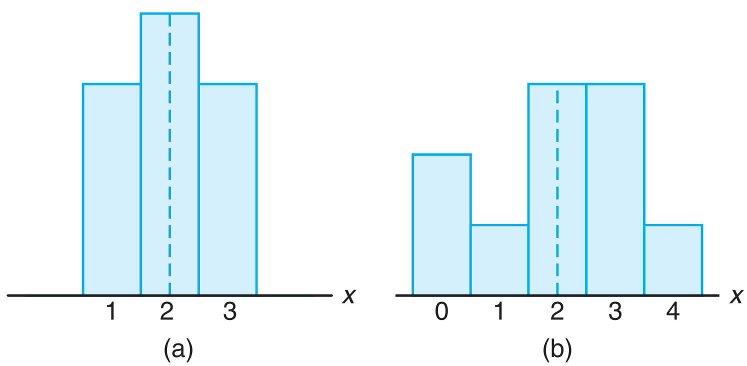


Figure 4.1: Distributions with equal means and unequal dispersions.

The most important measure of variability of a random variable X is obtained by applying Theorem 4.1 with $g(X) = (X - \mu)^2$. The quantity is referred to as the **variance of the random variable** X or the **variance of the probability**

distribution of X and is denoted by $\text{Var}(X)$ or the symbol σ_X^2, or simply by σ^2 when it is clear to which random variable we refer.

Definition 4.3: Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \qquad \text{if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx, \qquad \text{if } X \text{ is continuous.}$$

The positive square root of the variance, σ, is called the **standard deviation** of X.

The quantity $x-\mu$ in Definition 4.3 is called the **deviation of an observation** from its mean. Since the deviations are squared and then averaged, σ^2 will be much smaller for a set of x values that are close to μ than it will be for a set of values that vary considerably from μ.

Example 4.8: Let the random variable X represent the number of automobiles that are used for official business purposes on any given workday. The probability distribution for company A [Figure 4.1(a)] is

x	1	2	3
$f(x)$	0.3	0.4	0.3

and that for company B [Figure 4.1(b)] is

x	0	1	2	3	4
$f(x)$	0.2	0.1	0.3	0.3	0.1

Show that the variance of the probability distribution for company B is greater than that for company A.

***Solution*:** For company A, we find that

$$\mu_A = E(X) = (1)(0.3) + (2)(0.4) + (3)(0.3) = 2.0,$$

and then

$$\sigma_A^2 = \sum_{x=1}^{3}(x-2)^2 = (1-2)^2(0.3) + (2-2)^2(0.4) + (3-2)^2(0.3) = 0.6.$$

For company B, we have

$$\mu_B = E(X) = (0)(0.2) + (1)(0.1) + (2)(0.3) + (3)(0.3) + (4)(0.1) = 2.0,$$

and then

$$\begin{aligned}\sigma_B^2 &= \sum_{x=0}^{4}(x-2)^2 f(x)\\ &= (0-2)^2(0.2) + (1-2)^2(0.1) + (2-2)^2(0.3)\\ &\quad + (3-2)^2(0.3) + (4-2)^2(0.1) = 1.6.\end{aligned}$$

Clearly, the variance of the number of automobiles that are used for official business purposes is greater for company B than for company A. ∎

An alternative and preferred formula for finding σ^2, which often simplifies the calculations, is stated in the following theorem.

Theorem 4.2: The variance of a random variable X is

$$\sigma^2 = E(X^2) - \mu^2.$$

***Proof*:** For the discrete case, we can write

$$\begin{aligned}\sigma^2 &= \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x)\\ &= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x).\end{aligned}$$

Since $\mu = \sum_x x f(x)$ by definition, and $\sum_x f(x) = 1$ for any discrete probability distribution, it follows that

$$\sigma^2 = \sum_x x^2 f(x) - \mu^2 = E(X^2) - \mu^2.$$

For the continuous case the proof is step by step the same, with summations replaced by integrations. ∎

Example 4.9: Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X.

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

Using Theorem 4.2, calculate σ^2.

***Solution*:** First, we compute

$$\mu = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61.$$

Now,

$$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87.$$

Therefore,

$$\sigma^2 = 0.87 - (0.61)^2 = 0.4979.$$

∎

Example 4.10: The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a continuous random variable X having the probability density

$$f(x) = \begin{cases} 2(x-1), & 1 < x < 2,\\ 0, & \text{elsewhere.}\end{cases}$$

Find the mean and variance of X.

***Solution*:** Calculating $E(X)$ and $E(X^2)$, we have

$$\mu = E(X) = 2\int_1^2 x(x-1)\,dx = \frac{5}{3}$$

and

$$E(X^2) = 2\int_1^2 x^2(x-1)\,dx = \frac{17}{6}.$$

Therefore,

$$\sigma^2 = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$$

At this point, the variance or standard deviation has meaning only when we compare two or more distributions that have the same units of measurement. Therefore, we could compare the variances of the distributions of contents, measured in liters, of bottles of orange juice from two companies, and the larger value would indicate the company whose product was more variable or less uniform. It would not be meaningful to compare the variance of a distribution of heights to the variance of a distribution of aptitude scores. In Section 4.4, we show how the standard deviation can be used to describe a single distribution of observations.

We shall now extend our concept of the variance of a random variable X to include random variables related to X. For the random variable $g(X)$, the variance is denoted by $\sigma^2_{g(X)}$ and is calculated by means of the following theorem.

Theorem 4.3: Let X be a random variable with probability distribution $f(x)$. The variance of the random variable $g(X)$ is

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \sum_x [g(x) - \mu_{g(X)}]^2 f(x)$$

if X is discrete, and

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x)\,dx$$

if X is continuous.

***Proof*:** Since $g(X)$ is itself a random variable with mean $\mu_{g(X)}$ as defined in Theorem 4.1, it follows from Definition 4.3 that

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]\}.$$

Now, applying Theorem 4.1 again to the random variable $[g(X)-\mu_{g(X)}]^2$ completes the proof.

Example 4.11: Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

Solution: First, we find the mean of the random variable $2X+3$. According to Theorem 4.1,

$$\mu_{2X+3} = E(2X+3) = \sum_{x=0}^{3}(2x+3)f(x) = 6.$$

Now, using Theorem 4.3, we have

$$\begin{aligned}\sigma^2_{2X+3} &= E\{[(2X+3) - \mu_{2x+3}]^2\} = E[(2X+3-6)^2] \\ &= E(4X^2 - 12X + 9) = \sum_{x=0}^{3}(4x^2 - 12x + 9)f(x) = 4.\end{aligned}$$

Example 4.12: Let X be a random variable having the density function given in Example 4.5 on page 115. Find the variance of the random variable $g(X) = 4X + 3$.

Solution: In Example 4.5, we found that $\mu_{4X+3} = 8$. Now, using Theorem 4.3,

$$\begin{aligned}\sigma^2_{4X+3} &= E\{[(4X+3) - 8]^2\} = E[(4X-5)^2] \\ &= \int_{-1}^{2}(4x-5)^2\frac{x^2}{3}\,dx = \frac{1}{3}\int_{-1}^{2}(16x^4 - 40x^3 + 25x^2)\,dx = \frac{51}{5}.\end{aligned}$$

If $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$, where $\mu_X = E(X)$ and $\mu_Y = E(Y)$, Definition 4.2 yields an expected value called the **covariance** of X and Y, which we denote by σ_{XY} or $\mathrm{Cov}(X,Y)$.

Definition 4.4: Let X and Y be random variables with joint probability distribution $f(x,y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x\sum_y(x-\mu_X)(y-\mu_y)f(x,y)$$

if X and Y are discrete, and

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_y)f(x,y)\,dx\,dy$$

if X and Y are continuous.

The covariance between two random variables is a measure of the nature of the association between the two. If large values of X often result in large values of Y or small values of X result in small values of Y, positive $X-\mu_X$ will often result in positive $Y-\mu_Y$ and negative $X-\mu_X$ will often result in negative $Y-\mu_Y$. Thus, the product $(X-\mu_X)(Y-\mu_Y)$ will tend to be positive. On the other hand, if large X values often result in small Y values, the product $(X-\mu_X)(Y-\mu_Y)$ will tend to be negative. The *sign* of the covariance indicates whether the relationship between two dependent random variables is positive or negative. When X and Y are statistically independent, it can be shown that the covariance is zero (see Corollary 4.5). The converse, however, is not generally true. Two variables may have zero covariance and still not be statistically independent. Note that the covariance only describes the *linear* relationship between two random variables. Therefore, if a covariance between X and Y is zero, X and Y may have a nonlinear relationship, which means that they are not necessarily independent.

The alternative and preferred formula for σ_{XY} is stated by Theorem 4.4.

Theorem 4.4: The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y.$$

***Proof*:** For the discrete case, we can write

$$\begin{aligned}\sigma_{XY} &= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)f(x,y) \\ &= \sum_x \sum_y xyf(x,y) - \mu_X \sum_x \sum_y yf(x,y) \\ &\quad - \mu_Y \sum_x \sum_y xf(x,y) + \mu_X\mu_Y \sum_x \sum_y f(x,y).\end{aligned}$$

Since

$$\mu_X = \sum_x xf(x,y), \quad \mu_Y = \sum_y yf(x,y), \text{ and } \sum_x \sum_y f(x,y) = 1$$

for any joint discrete distribution, it follows that

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y = E(XY) - \mu_X\mu_Y.$$

For the continuous case, the proof is identical with summations replaced by integrals.

Example 4.13: Example 3.14 on page 95 describes a situation involving the number of blue refills X and the number of red refills Y. Two refills for a ballpoint pen are selected at random from a certain box, and the following is the joint probability distribution:

		x			
	$f(x,y)$	0	1	2	$h(y)$
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
	$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Find the covariance of X and Y.

***Solution*:** From Example 4.6, we see that $E(XY) = 3/14$. Now

$$\mu_X = \sum_{x=0}^{2} xg(x) = (0)\left(\frac{5}{14}\right) + (1)\left(\frac{15}{28}\right) + (2)\left(\frac{3}{28}\right) = \frac{3}{4},$$

and

$$\mu_Y = \sum_{y=0}^{2} yh(y) = (0)\left(\frac{15}{28}\right) + (1)\left(\frac{3}{7}\right) + (2)\left(\frac{1}{28}\right) = \frac{1}{2}.$$

Therefore,

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{3}{14} - \left(\frac{3}{4}\right)\left(\frac{1}{2}\right) = -\frac{9}{56}.$$

Example 4.14: The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x, y) = \begin{cases} 8xy, & 0 \le y \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the covariance of X and Y.

Solution: We first compute the marginal density functions. They are

$$g(x) = \begin{cases} 4x^3, & 0 \le x \le 1, \\ 0, & \text{elsewhere,} \end{cases}$$

and

$$h(y) = \begin{cases} 4y(1 - y^2), & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

From these marginal density functions, we compute

$$\mu_X = E(X) = \int_0^1 4x^4 \, dx = \frac{4}{5} \text{ and } \mu_Y = \int_0^1 4y^2(1 - y^2) \, dy = \frac{8}{15}.$$

From the joint density function given above, we have

$$E(XY) = \int_0^1 \int_y^1 8x^2 y^2 \, dx \, dy = \frac{4}{9}.$$

Then

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{4}{9} - \left(\frac{4}{5}\right)\left(\frac{8}{15}\right) = \frac{4}{225}.$$

Although the covariance between two random variables does provide information regarding the nature of the relationship, the magnitude of σ_{XY} *does not indicate anything regarding the strength of the relationship*, since σ_{XY} is not scale-free. Its magnitude will depend on the units used to measure both X and Y. There is a scale-free version of the covariance called the **correlation coefficient** that is used widely in statistics.

Definition 4.5: Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y, respectively. The correlation coefficient of X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

It should be clear to the reader that ρ_{XY} is free of the units of X and Y. The correlation coefficient satisfies the inequality $-1 \le \rho_{XY} \le 1$. It assumes a value of zero when $\sigma_{XY} = 0$. Where there is an exact linear dependency, say $Y \equiv a + bX$,

$\rho_{XY} = 1$ if $b > 0$ and $\rho_{XY} = -1$ if $b < 0$. (See Exercise 4.48.) The correlation coefficient is the subject of more discussion in Chapter 12, where we deal with linear regression.

Example 4.15: Find the correlation coefficient between X and Y in Example 4.13.

Solution: Since

$$E(X^2) = (0^2)\left(\frac{5}{14}\right) + (1^2)\left(\frac{15}{28}\right) + (2^2)\left(\frac{3}{28}\right) = \frac{27}{28}$$

and

$$E(Y^2) = (0^2)\left(\frac{15}{28}\right) + (1^2)\left(\frac{3}{7}\right) + (2^2)\left(\frac{1}{28}\right) = \frac{4}{7},$$

we obtain

$$\sigma_X^2 = \frac{27}{28} - \left(\frac{3}{4}\right)^2 = \frac{45}{112} \text{ and } \sigma_Y^2 = \frac{4}{7} - \left(\frac{1}{2}\right)^2 = \frac{9}{28}.$$

Therefore, the correlation coefficient between X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-9/56}{\sqrt{(45/112)(9/28)}} = -\frac{1}{\sqrt{5}}.$$

Example 4.16: Find the correlation coefficient of X and Y in Example 4.14.

Solution: Because

$$E(X^2) = \int_0^1 4x^5 \, dx = \frac{2}{3} \text{ and } E(Y^2) = \int_0^1 4y^3(1-y^2) \, dy = 1 - \frac{2}{3} = \frac{1}{3},$$

we conclude that

$$\sigma_X^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75} \text{ and } \sigma_Y^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}.$$

Hence,

$$\rho_{XY} = \frac{4/225}{\sqrt{(2/75)(11/225)}} = \frac{4}{\sqrt{66}}.$$

Note that although the covariance in Example 4.15 is larger in magnitude (disregarding the sign) than that in Example 4.16, the relationship of the magnitudes of the correlation coefficients in these two examples is just the reverse. This is evidence that we cannot look at the magnitude of the covariance to decide on how strong the relationship is.

Exercises

4.33 Use Definition 4.3 on page 120 to find the variance of the random variable X of Exercise 4.7 on page 117.

4.34 Let X be a random variable with the following probability distribution:

x	-2	3	5
$f(x)$	0.3	0.2	0.5

Find the standard deviation of X.

4.35 The random variable X, representing the number of errors per 100 lines of software code, has the following probability distribution:

x	2	3	4	5	6
$f(x)$	0.01	0.25	0.4	0.3	0.04

Using Theorem 4.2 on page 121, find the variance of X.

4.36 Suppose that the probabilities are 0.4, 0.3, 0.2, and 0.1, respectively, that 0, 1, 2, or 3 power failures will strike a certain subdivision in any given year. Find the mean and variance of the random variable X representing the number of power failures striking this subdivision.

4.37 A dealer's profit, in units of \$5000, on a new automobile is a random variable X having the density function given in Exercise 4.12 on page 117. Find the variance of X.

4.38 The proportion of people who respond to a certain mail-order solicitation is a random variable X having the density function given in Exercise 4.14 on page 117. Find the variance of X.

4.39 The total number of hours, in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a random variable X having the density function given in Exercise 4.13 on page 117. Find the variance of X.

4.40 Referring to Exercise 4.14 on page 117, find $\sigma^2_{g(X)}$ for the function $g(X) = 3X^2 + 4$.

4.41 Find the standard deviation of the random variable $g(X) = (2X + 1)^2$ in Exercise 4.17 on page 118.

4.42 Using the results of Exercise 4.21 on page 118, find the variance of $g(X) = X^2$, where X is a random variable having the density function given in Exercise 4.12 on page 117.

4.43 The length of time, in minutes, for an airplane to obtain clearance for takeoff at a certain airport is a random variable $Y = 3X - 2$, where X has the density function

$$f(x) = \begin{cases} \frac{1}{4}e^{-x/4}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the mean and variance of the random variable Y.

4.44 Find the covariance of the random variables X and Y of Exercise 3.39 on page 105.

4.45 Find the covariance of the random variables X and Y of Exercise 3.49 on page 106.

4.46 Find the covariance of the random variables X and Y of Exercise 3.44 on page 105.

4.47 For the random variables X and Y whose joint density function is given in Exercise 3.40 on page 105, find the covariance.

4.48 Given a random variable X, with standard deviation σ_X, and a random variable $Y = a + bX$, show that if $b < 0$, the correlation coefficient $\rho_{XY} = -1$, and if $b > 0$, $\rho_{XY} = 1$.

4.49 Consider the situation in Exercise 4.32 on page 119. The distribution of the number of imperfections per 10 meters of synthetic failure is given by

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Find the variance and standard deviation of the number of imperfections.

4.50 For a laboratory assignment, if the equipment is working, the density function of the observed outcome X is

$$f(x) = \begin{cases} 2(1-x), & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Find the variance and standard deviation of X.

4.51 For the random variables X and Y in Exercise 3.39 on page 105, determine the correlation coefficient between X and Y.

4.52 Random variables X and Y follow a joint distribution

$$f(x, y) = \begin{cases} 2, & 0 < x \le y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Determine the correlation coefficient between X and Y.

4.3 Means and Variances of Linear Combinations of Random Variables

We now develop some useful properties that will simplify the calculations of means and variances of random variables that appear in later chapters. These properties will permit us to deal with expectations in terms of other parameters that are either known or easily computed. All the results that we present here are valid for both discrete and continuous random variables. Proofs are given only for the continuous case. We begin with a theorem and two corollaries that should be, intuitively, reasonable to the reader.

Theorem 4.5: If a and b are constants, then

$$E(aX + b) = aE(X) + b.$$

***Proof*:** By the definition of expected value,

$$E(aX + b) = \int_{-\infty}^{\infty} (ax + b)f(x)\ dx = a\int_{-\infty}^{\infty} xf(x)\ dx + b\int_{-\infty}^{\infty} f(x)\ dx.$$

The first integral on the right is $E(X)$ and the second integral equals 1. Therefore, we have

$$E(aX + b) = aE(X) + b.$$

Corollary 4.1: Setting $a = 0$, we see that $E(b) = b$.

Corollary 4.2: Setting $b = 0$, we see that $E(aX) = aE(X)$.

Example 4.17: Applying Theorem 4.5 to the discrete random variable $f(X) = 2X - 1$, rework Example 4.4 on page 115.

***Solution*:** According to Theorem 4.5, we can write

$$E(2X - 1) = 2E(X) - 1.$$

Now

$$\begin{aligned}\mu = E(X) &= \sum_{x=4}^{9} xf(x)\\ &= (4)\left(\frac{1}{12}\right) + (5)\left(\frac{1}{12}\right) + (6)\left(\frac{1}{4}\right) + (7)\left(\frac{1}{4}\right) + (8)\left(\frac{1}{6}\right) + (9)\left(\frac{1}{6}\right) = \frac{41}{6}.\end{aligned}$$

Therefore,

$$\mu_{2X-1} = (2)\left(\frac{41}{6}\right) - 1 = \$12.67,$$

as before.

Example 4.18: Applying Theorem 4.5 to the continuous random variable $g(X) = 4X + 3$, rework Example 4.5 on page 115.

***Solution*:** For Example 4.5, we may use Theorem 4.5 to write

$$E(4X + 3) = 4E(X) + 3.$$

Now

$$E(X) = \int_{-1}^{2} x\left(\frac{x^2}{3}\right) dx = \int_{-1}^{2} \frac{x^3}{3}\, dx = \frac{5}{4}.$$

Therefore,

$$E(4X + 3) = (4)\left(\frac{5}{4}\right) + 3 = 8,$$

as before.

Theorem 4.6: The expected value of the sum or difference of two or more functions of a random variable X is the sum or difference of the expected values of the functions. That is,

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)].$$

***Proof*:** By definition,

$$\begin{aligned} E[g(X) \pm h(X)] &= \int_{-\infty}^{\infty} [g(x) \pm h(x)]f(x)\, dx \\ &= \int_{-\infty}^{\infty} g(x)f(x)\, dx \pm \int_{-\infty}^{\infty} h(x)f(x)\, dx \\ &= E[g(X)] \pm E[h(X)]. \end{aligned}$$

Example 4.19: Let X be a random variable with probability distribution as follows:

x	0	1	2	3
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

Find the expected value of $Y = (X - 1)^2$.

***Solution*:** Applying Theorem 4.6 to the function $Y = (X - 1)^2$, we can write

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

From Corollary 4.1, $E(1) = 1$, and by direct computation,

$$E(X) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (2)(0) + (3)\left(\frac{1}{6}\right) = 1 \text{ and}$$
$$E(X^2) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{1}{2}\right) + (4)(0) + (9)\left(\frac{1}{6}\right) = 2.$$

Hence,

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

Example 4.20: The weekly demand for a certain drink, in thousands of liters, at a chain of convenience stores is a continuous random variable $g(X) = X^2 + X - 2$, where X has the density function

$$f(x) = \begin{cases} 2(x-1), & 1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of the weekly demand for the drink.

***Solution*:** By Theorem 4.6, we write

$$E(X^2 + X - 2) = E(X^2) + E(X) - E(2).$$

From Corollary 4.1, $E(2) = 2$, and by direct integration,

$$E(X) = \int_1^2 2x(x-1)\ dx = \frac{5}{3} \text{ and } E(X^2) = \int_1^2 2x^2(x-1)\ dx = \frac{17}{6}.$$

Now

$$E(X^2 + X - 2) = \frac{17}{6} + \frac{5}{3} - 2 = \frac{5}{2},$$

so the average weekly demand for the drink from this chain of efficiency stores is 2500 liters.

Suppose that we have two random variables X and Y with joint probability distribution $f(x, y)$. Two additional properties that will be very useful in succeeding chapters involve the expected values of the sum, difference, and product of these two random variables. First, however, let us prove a theorem on the expected value of the sum or difference of functions of the given variables. This, of course, is merely an extension of Theorem 4.6.

Theorem 4.7: The expected value of the sum or difference of two or more functions of the random variables X and Y is the sum or difference of the expected values of the functions. That is,

$$E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)].$$

***Proof*:** By Definition 4.2,

$$\begin{aligned} E[g(X,Y) \pm h(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [g(x,y) \pm h(x,y)] f(x,y)\ dx\ dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f(x,y)\ dx\ dy \pm \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\ dx\ dy \\ &= E[g(X,Y)] \pm E[h(X,Y)]. \end{aligned}$$

Corollary 4.3: Setting $g(X,Y) = g(X)$ and $h(X,Y) = h(Y)$, we see that

$$E[g(X) \pm h(Y)] = E[g(X)] \pm E[h(Y)].$$

Corollary 4.4: Setting $g(X,Y) = X$ and $h(X,Y) = Y$, we see that

$$E[X \pm Y] = E[X] \pm E[Y].$$

If X represents the daily production of some item from machine A and Y the daily production of the same kind of item from machine B, then $X+Y$ represents the total number of items produced daily by both machines. Corollary 4.4 states that the average daily production for both machines is equal to the sum of the average daily production of each machine.

Theorem 4.8: Let X and Y be two independent random variables. Then

$$E(XY) = E(X)E(Y).$$

***Proof*:** By Definition 4.2,

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)\;dx\;dy.$$

Since X and Y are independent, we may write

$$f(x,y) = g(x)h(y),$$

where $g(x)$ and $h(y)$ are the marginal distributions of X and Y, respectively. Hence,

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyg(x)h(y)\;dx\;dy = \int_{-\infty}^{\infty} xg(x)\;dx \int_{-\infty}^{\infty} yh(y)\;dy$$
$$= E(X)E(Y).$$

Theorem 4.8 can be illustrated for discrete variables by considering the experiment of tossing a green die and a red die. Let the random variable X represent the outcome on the green die and the random variable Y represent the outcome on the red die. Then XY represents the product of the numbers that occur on the pair of dice. In the long run, the average of the products of the numbers is equal to the product of the average number that occurs on the green die and the average number that occurs on the red die.

Corollary 4.5: Let X and Y be two independent random variables. Then $\sigma_{XY} = 0$.

***Proof*:** The proof can be carried out by using Theorems 4.4 and 4.8.

Example 4.21: It is known that the ratio of gallium to arsenide does not affect the functioning of gallium-arsenide wafers, which are the main components of microchips. Let X denote the ratio of gallium to arsenide and Y denote the functional wafers retrieved during a 1-hour period. X and Y are independent random variables with the joint density function

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $E(XY) = E(X)E(Y)$, as Theorem 4.8 suggests.

***Solution*:** By definition,

$$E(XY) = \int_0^1 \int_0^2 \frac{x^2y(1+3y^2)}{4}\, dxdy = \frac{5}{6},\ E(X) = \frac{4}{3},\ \text{and } E(Y) = \frac{5}{8}.$$

Hence,

$$E(X)E(Y) = \left(\frac{4}{3}\right)\left(\frac{5}{8}\right) = \frac{5}{6} = E(XY).$$

We conclude this section by proving one theorem and presenting several corollaries that are useful for calculating variances or standard deviations.

Theorem 4.9: If X and Y are random variables with joint probability distribution $f(x,y)$ and a, b, and c are constants, then

$$\sigma^2_{aX+bY+c} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}.$$

***Proof*:** By definition, $\sigma^2_{aX+bY+c} = E\{[(aX+bY+c) - \mu_{aX+bY+c}]^2\}$. Now

$$\mu_{aX+bY+c} = E(aX+bY+c) = aE(X) + bE(Y) + c = a\mu_X + b\mu_Y + c,$$

by using Corollary 4.4 followed by Corollary 4.2. Therefore,

$$\begin{aligned}\sigma^2_{aX+bY+c} &= E\{[a(X-\mu_X) + b(Y-\mu_Y)]^2\}\\ &= a^2E[(X-\mu_X)^2] + b^2E[(Y-\mu_Y)^2] + 2abE[(X-\mu_X)(Y-\mu_Y)]\\ &= a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}.\end{aligned}$$

Using Theorem 4.9, we have the following corollaries.

Corollary 4.6: Setting $b = 0$, we see that

$$\sigma^2_{aX+c} = a^2\sigma^2_X = a^2\sigma^2.$$

Corollary 4.7: Setting $a = 1$ and $b = 0$, we see that

$$\sigma^2_{X+c} = \sigma^2_X = \sigma^2.$$

Corollary 4.8: Setting $b = 0$ and $c = 0$, we see that

$$\sigma^2_{aX} = a^2\sigma^2_X = a^2\sigma^2.$$

Corollaries 4.6 and 4.7 state that the variance is unchanged if a constant is added to or subtracted from a random variable. The addition or subtraction of a constant simply shifts the values of X to the right or to the left but does not change their variability. However, if a random variable is multiplied or divided by a constant, then Corollaries 4.6 and 4.8 state that the variance is multiplied or divided by the square of the constant.

Corollary 4.9: If X and Y are independent random variables, then

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y.$$

The result stated in Corollary 4.9 is obtained from Theorem 4.9 by invoking Corollary 4.5.

Corollary 4.10: If X and Y are independent random variables, then

$$\sigma^2_{aX-bY} = a^2\sigma^2_X + b^2\sigma^2_Y.$$

Corollary 4.10 follows when b in Corollary 4.9 is replaced by $-b$. Generalizing to a linear combination of n independent random variables, we have Corollary 4.11.

Corollary 4.11: If $X_1, X_2, \dots, X_n$ are independent random variables, then

$$\sigma^2_{a_1X_1+a_2X_2+\cdots+a_nX_n} = a_1^2\sigma^2_{X_1} + a_2^2\sigma^2_{X_2} + \cdots + a_n^2\sigma^2_{X_n}.$$

Example 4.22: If X and Y are random variables with variances $\sigma^2_X = 2$ and $\sigma^2_Y = 4$ and covariance $\sigma_{XY} = -2$, find the variance of the random variable $Z = 3X - 4Y + 8$.

***Solution*:**

$$\begin{aligned}\sigma^2_Z &= \sigma^2_{3X-4Y+8} = \sigma^2_{3X-4Y} && \text{(by Corollary 4.6)}\\ &= 9\sigma^2_X + 16\sigma^2_Y - 24\sigma_{XY} && \text{(by Theorem 4.9)}\\ &= (9)(2) + (16)(4) - (24)(-2) = 130.\end{aligned}$$

Example 4.23: Let X and Y denote the amounts of two different types of impurities in a batch of a certain chemical product. Suppose that X and Y are independent random variables with variances $\sigma^2_X = 2$ and $\sigma^2_Y = 3$. Find the variance of the random variable $Z = 3X - 2Y + 5$.

***Solution*:**

$$\begin{aligned}\sigma^2_Z &= \sigma^2_{3X-2Y+5} = \sigma^2_{3X-2Y} && \text{(by Corollary 4.6)}\\ &= 9\sigma^2_x + 4\sigma^2_y && \text{(by Corollary 4.10)}\\ &= (9)(2) + (4)(3) = 30.\end{aligned}$$

What If the Function Is Nonlinear?

In that which has preceded this section, we have dealt with properties of linear functions of random variables for very important reasons. Chapters 8 through 15 will discuss and illustrate practical real-world problems in which the analyst is constructing a **linear model** to describe a data set and thus to describe or explain the behavior of a certain scientific phenomenon. Thus, it is natural that expected values and variances of linear combinations of random variables are encountered. However, there are situations in which properties of **nonlinear** functions of random variables become important. Certainly there are many scientific phenomena that are nonlinear, and certainly statistical modeling using nonlinear functions is very important. In fact, in Chapter 12, we deal with the modeling of what have become standard nonlinear models. Indeed, even a simple function of random variables, such as $Z = X/Y$, occurs quite frequently in practice, and yet unlike in the case of the expected value of linear combinations of random variables, there is no simple general rule. For example,

$$E(Z) = E(X/Y) \neq E(X)/E(Y),$$

except in very special circumstances.

The material provided by Theorems 4.5 through 4.9 and the various corollaries is extremely useful in that there are no restrictions on the form of the density or probability functions, apart from the property of independence when it is required as in the corollaries following Theorems 4.9. To illustrate, consider Example 4.23; the variance of $Z = 3X - 2Y + 5$ does not require restrictions on the distributions of the amounts X and Y of the two types of impurities. Only independence between X and Y is required. Now, we do have at our disposal the capacity to find $\mu_{g(X)}$ and $\sigma^2_{g(X)}$ for any function $g(\cdot)$ from first principles established in Theorems 4.1 and 4.3, where it is assumed that the corresponding distribution $f(x)$ is **known**. Exercises 4.40, 4.41, and 4.42, among others, illustrate the use of these theorems. Thus, if the function $g(x)$ is nonlinear and the density function (or probability function in the discrete case) is known, $\mu_{g(X)}$ and $\sigma^2_{g(X)}$ can be evaluated exactly. But, similar to the rules given for linear combinations, are there rules for nonlinear functions that can be used when the form of the distribution of the pertinent random variables is not known?

In general, suppose X is a random variable and $Y = g(x)$. The general solution for $E(Y)$ or $\text{Var}(Y)$ can be difficult to find and depends on the complexity of the function $g(\cdot)$. However, there are approximations available that depend on a linear approximation of the function $g(x)$. For example, suppose we denote $E(X)$ as μ and $\text{Var}(X) = \sigma^2_X$. Then a Taylor series approximation of $g(x)$ around $X = \mu_X$ gives

$$g(x) = g(\mu_X) + \left.\frac{\partial g(x)}{\partial x}\right|_{x=\mu_X} (x - \mu_X) + \left.\frac{\partial^2 g(x)}{\partial x^2}\right|_{x=\mu_X} \frac{(x - \mu_X)^2}{2} + \cdots .$$

As a result, if we truncate after the linear term and take the expected value of both sides, we obtain $E[g(X)] \approx g(\mu_X)$, which is certainly intuitive and in some cases gives a reasonable approximation. However, if we include the second-order term of the Taylor series, then we have a second-order adjustment for this *first-order approximation* as follows:

Approximation of $E[g(X)]$

$$E[g(X)] \approx g(\mu_X) + \left.\frac{\partial^2 g(x)}{\partial x^2}\right|_{x=\mu_X} \frac{\sigma^2_X}{2}.$$

Example 4.24: Given the random variable X with mean μ_X and variance σ^2_X, give the second-order approximation to $E(e^X)$.

Solution: Since $\frac{\partial e^x}{\partial x} = e^x$ and $\frac{\partial^2 e^x}{\partial x^2} = e^x$, we obtain $E(e^X) \approx e^{\mu_X}(1 + \sigma^2_X/2)$.

Similarly, we can develop an approximation for $\text{Var}[g(x)]$ by taking the variance of both sides of the first-order Taylor series expansion of $g(x)$.

Approximation of $\text{Var}[g(X)]$

$$\text{Var}[g(X)] \approx \left[\frac{\partial g(x)}{\partial x}\right]^2_{x=\mu_X} \sigma^2_X.$$

Example 4.25: Given the random variable X as in Example 4.24, give an approximate formula for $\text{Var}[g(x)]$.