Chapter

5

DISCRETE RANDOM VARIABLES AND THEIR PROBABILITY DISTRIBUTIONS

Now that you know a little about probability, do you feel lucky? If you've got $20 to spend on lunch today, are you willing to spend it all on twenty $1 lotto tickets to increase your chances of winning? What if you know that, depending on what state you are in, it could mean buying as many as 18 million $1 tickets to cover all the possible combinations to have a definite chance to win? (See Case Study 5–2.) That is a lot of lunch! How do we go about determining the outcomes and probabilities in a lotto game?

5.1 Random Variables
5.2 Probability Distribution of a Discrete Random Variable
5.3 Mean of a Discrete Random Variable
Case Study 5–1 Aces High Instant Lottery Game
5.4 Standard Deviation of a Discrete Random Variable
5.5 Factorials and Combinations
Case Study 5–2 Playing Lotto
5.6 The Binomial Probability Distribution
5.7 The Hypergeometric Probability Distribution
5.8 The Poisson Probability Distribution
Case Study 5–3 Ask Mr. Statistics

Chapter 4 discussed the concepts and rules of probability. This chapter extends the concept of probability to explain probability distributions. As we saw in Chapter 4, any given statistical experiment has more than one outcome. It is impossible to predict which of the many possible outcomes will occur if an experiment is performed. Consequently, decisions are made under uncertain conditions. For example, a lottery player does not know in advance whether or not he is going to win that lottery. If he knows that he is not going to win, he will definitely not play. It is the uncertainty about winning (some positive probability of winning) that makes him play. This chapter shows that if the outcomes and their probabilities for a statistical experiment are known, we can find out what will happen on average, if that experiment is performed many times. For the lottery example, we can find out what a lottery player can expect to win (or lose), on average, if he continues playing this lottery again and again.

In this chapter, random variables and types of random variables are explained. Then, the concept of a probability distribution and its mean and standard deviation are discussed. Finally, three special probability distributions for a discrete random variable—the binomial probability distribution, the hypergeometric probability distribution, and the Poisson probability distribution—are developed.

5.1 RANDOM VARIABLES

Suppose Table 5.1 gives the frequency and relative frequency distributions of the number of vehicles owned by all 2000 families living in a small town.

Table 5.1 Frequency and Relative Frequency Distributions of the Number of Vehicles Owned by Families

Number of Vehicles Owned	Frequency	Relative Frequency
0	30	30/2000 = .015
1	470	470/2000 = .235
2	850	850/2000 = .425
3	490	490/2000 = .245
4	160	160/2000 = .080
	$N = 2000$	Sum = 1.000

Suppose one family is randomly selected from this population. The process of randomly selecting a family is called a *random* or *chance experiment*. Let x denote the number of vehicles owned by the selected family. Then x can assume any of the five possible values (0, 1, 2, 3, and 4) listed in the first column of Table 5.1. The value assumed by x depends on which family is selected. Thus, this value depends on the outcome of a random experiment. Consequently, x is called a **random variable** or a **chance variable**. In general, a random variable is denoted by x or y.

Definition

Random Variable A *random variable* is a variable whose value is determined by the outcome of a random experiment.

As will be explained next, a random variable can be discrete or continuous.

5.1.1 Discrete Random Variable

A **discrete random variable** assumes values that can be counted. In other words, the consecutive values of a discrete random variable are separated by a certain gap.

Definition

Discrete Random Variable A *random variable* that assumes countable values is called a *discrete random variable.*

In Table 5.1, *the number of vehicles owned by a family* is an example of a discrete random variable because the values of the random variable x are countable: 0, 1, 2, 3, and 4. Here are some other examples of discrete random variables:

1. The number of cars sold at a dealership during a given month
2. The number of houses in a certain block
3. The number of fish caught on a fishing trip
4. The number of complaints received at the office of an airline on a given day
5. The number of customers who visit a bank during any given hour
6. The number of heads obtained in three tosses of a coin

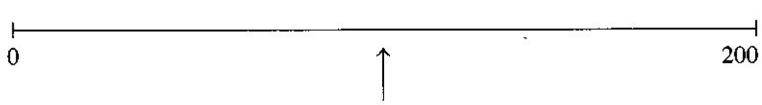
5.1.2 Continuous Random Variable

A random variable whose values are not countable is called a **continuous random variable.** A continuous random variable can assume any value over an interval or intervals.

Definition

Continuous Random Variable A random variable that can assume any value contained in one or more intervals is called a *continuous random variable.*

Because the number of values contained in any interval is infinite, the possible number of values that a continuous random variable can assume is also infinite. Moreover, we cannot count these values. Consider the life of a battery. We can measure it as precisely as we want. For instance, the life of this battery may be 40 hours, or 40.25 hours, or 40.247 hours. Assume that the maximum life of a battery is 200 hours. Let x denote the life of a randomly selected battery of this kind. Then, x can assume any value in the interval 0 to 200. Consequently, x is a continuous random variable. As shown in the diagram, every point on the line representing the interval 0 to 200 gives a possible value of x.



Every point on this line represents a possible value of x that denotes the life of a battery. There are an infinite number of points on this line. The values represented by points on this line are uncountable.

The following are some examples of continuous random variables:

1. The height of a person
2. The time taken to complete an examination
3. The amount of milk in a gallon (Note that we do not expect a gallon to contain exactly one gallon of milk but either slightly more or slightly less than a gallon.)

4. The weight of a fish
5. The price of a house

This chapter is limited to a discussion of discrete random variables and their probability distributions. Continuous random variables will be discussed in Chapter 6.

EXERCISES

CONCEPTS AND PROCEDURES

5.1 Explain the meaning of a random variable, a discrete random variable, and a continuous random variable. Give one example each of a discrete random variable and a continuous random variable.

5.2 Classify each of the following random variables as discrete or continuous.
- **a.** The time left on a parking meter
- **b.** The number of bats broken by a major league baseball team in a season
- **c.** The number of fish caught on a fishing trip
- **d.** The total pounds of fish caught on a fishing trip
- **e.** The number of gumballs in a vending machine
- **f.** The time spent by a physician examining a patient

5.3 Indicate which of the following random variables are discrete and which are continuous.
- **a.** The number of new accounts opened at a bank during a certain month
- **b.** The time taken to run a marathon
- **c.** The price of a concert ticket
- **d.** The number of rotten eggs in a randomly selected box
- **e.** The points scored in a football game
- **f.** The weight of a randomly selected package

APPLICATIONS

5.4 A household can watch news on any of the three networks—ABC, CBS, or NBC. On a certain day, five households randomly and independently decide which channel to watch. Let x be the number of households among these five that decide to watch news on ABC. Is x a discrete or a continuous random variable? Explain.

5.5 One of the four gas stations located at an intersection of two major roads is a Texaco station. Suppose the next six cars that stop at any of these four gas stations make their selections randomly and independently. Let x be the number of cars in these six that stop at the Texaco station. Is x a discrete or a continuous random variable? Explain.

5.2 PROBABILITY DISTRIBUTION OF A DISCRETE RANDOM VARIABLE

Let x be a discrete random variable. The **probability distribution** of x describes how the probabilities are distributed over all the possible values of x.

Definition

Probability Distribution of a Discrete Random Variable The *probability distribution of a discrete random variable* lists all the possible values that the random variable can assume and their corresponding probabilities.

1. The probability assigned to each value of a random variable x lies in the range 0 to 1; that is, $0 \leq P(x) \leq 1$ for each x.
2. The sum of the probabilities assigned to all possible values of x is equal to 1.0; that is, $\Sigma P(x) = 1$. (Remember, if the probabilities are rounded, the sum may not be exactly 1.0.)

Two Characteristics of a Probability Distribution The probability distribution of a discrete random variable possesses the following two characteristics.

1. $0 \leq P(x) \leq 1$ for each value of x
2. $\Sigma P(x) = 1$

These two characteristics are also called the *two conditions* that a probability distribution must satisfy. Notice that in Table 5.3 each probability listed in the column labeled $P(x)$ is between 0 and 1. Also, $\Sigma P(x) = 1.0$. Because both conditions are satisfied, Table 5.3 represents the probability distribution of x.

From Table 5.3, we can read the probability for any value of x. For example, the probability that a randomly selected family from this town owns two vehicles is .425. This probability is written as

$$P(x = 2) = .425 \quad \text{or} \quad P(2) = .425$$

The probability that the selected family owns more than two vehicles is given by the sum of the probabilities of owning three and four vehicles. This probability is .245 + .080 = .325, which can be written as

$$P(x > 2) = P(x = 3) + P(x = 4) = P(3) + P(4) = .245 + .080 = .325$$

The probability distribution of a discrete random variable can be presented in the form of a *mathematical formula*, a *table*, or a *graph*. Table 5.3 presented the probability distribution in tabular form. Figure 5.1 shows the graphical presentation of the probability distribution of Table 5.3. In this figure, each value of x is marked on the horizontal axis. The probability for each value of x is exhibited by the height of the corresponding bar. Such a graph is called a **line** or **bar graph**. This section does not discuss the presentation of a probability distribution in a mathematical formula.

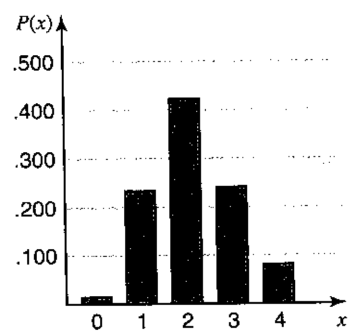


Figure 5.1 Graphical presentation of the probability distribution of Table 5.3.

EXAMPLE 5–2

Verifying the conditions of a probability distribution.

Each of the following tables lists certain values of x and their probabilities. Determine whether or not each table represents a valid probability distribution.

(a)

x	$P(x)$
0	.08
1	.11
2	.39
3	.27

(b)

x	$P(x)$
2	.25
3	.34
4	.28
5	.13

(c)

x	$P(x)$
7	.70
8	.50
9	−.20

Solution

(a) Because each probability listed in this table is in the range 0 to 1, it satisfies the first condition of a probability distribution. However, the sum of all probabilities is not equal to 1.0 because $\Sigma P(x) = .08 + .11 + .39 + .27 = .85$. Therefore, the second condition is not satisfied. Consequently, this table does not represent a valid probability distribution.

(b) Each probability listed in this table is in the range 0 to 1. Also, $\Sigma P(x) = .25 + .34 + .28 + .13 = 1.0$. Consequently, this table represents a valid probability distribution.

(c) Although the sum of all probabilities listed in this table is equal to 1.0, one of the probabilities is negative. This violates the first condition of a probability distribution. Therefore, this table does not represent a valid probability distribution.

EXAMPLE 5–3

The following table lists the probability distribution of the number of breakdowns per week for a machine based on past data.

Breakdowns per week	0	1	2	3
Probability	.15	.20	.35	.30

(a) Present this probability distribution graphically.

(b) Find the probability that the number of breakdowns for this machine during a given week is

i. exactly 2 **ii.** 0 to 2

iii. more than 1 **iv.** at most 1

Solution Let x denote the number of breakdowns for this machine during a given week. Table 5.4 lists the probability distribution of x.

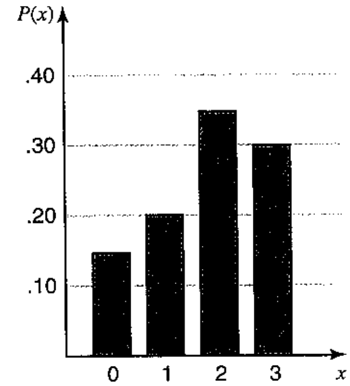
Table 5.4 Probability Distribution of the Number of Breakdowns

x	$P(x)$
0	.15
1	.20
2	.35
3	.30
	$\Sigma P(x) = 1.00$

Graphing a probability distribution.

(a) Figure 5.2 shows the bar graph of the probability distribution of Table 5.4.

Figure 5.2 Graphical presentation of the probability distribution of Table 5.4.



Finding the probabilities of events for a discrete random variable.

(b) Using Table 5.4, we can calculate the required probabilities as follows.

i. The probability of exactly two breakdowns is

$$P(\text{exactly 2 breakdowns}) = P(x = 2) = \mathbf{.35}$$

ii. The probability of zero to two breakdowns is given by the sum of the probabilities of 0, 1, and 2 breakdowns.

$$\begin{aligned} P(\text{0 to 2 breakdowns}) &= P(0 \le x \le 2) \\ &= P(x = 0) + P(x = 1) + P(x = 2) \\ &= .15 + .20 + .35 = \mathbf{.70} \end{aligned}$$

iii. The probability of more than one breakdown is obtained by adding the probabilities of 2 and 3 breakdowns.

$$\begin{aligned} P(\text{more than 1 breakdown}) &= P(x > 1) \\ &= P(x = 2) + P(x = 3) \\ &= .35 + .30 = \mathbf{.65} \end{aligned}$$

iv. The probability of at most one breakdown is given by the sum of the probabilities of 0 and 1 breakdown.

$$\begin{aligned} P(\text{at most 1 breakdown}) &= P(x \le 1) \\ &= P(x = 0) + P(x = 1) \\ &= .15 + .20 = \mathbf{.35} \end{aligned}$$

EXAMPLE 5–4

Constructing a probability distribution.

According to a survey, 60% of all students at a large university suffer from math anxiety. Two students are randomly selected from this university. Let x denote the number of students in this sample who suffer from math anxiety. Develop the probability distribution of x.

Solution Let us define the following two events:

N = the student selected does not suffer from math anxiety

M = the student selected suffers from math anxiety

As we can observe from the tree diagram of Figure 5.3, there are four possible outcomes for this experiment: *NN* (neither of the students suffers from math anxiety), *NM* (the first student does not suffer from math anxiety and the second does), *MN* (the first student suffers from math anxiety and the second does not), and *MM* (both students suffer from math anxiety). The probabilities of these four outcomes are listed in the tree diagram. Because 60% of the students suffer from math anxiety and 40% do not, the probability is .60 that any student selected suffers from math anxiety and .40 that he or she does not.

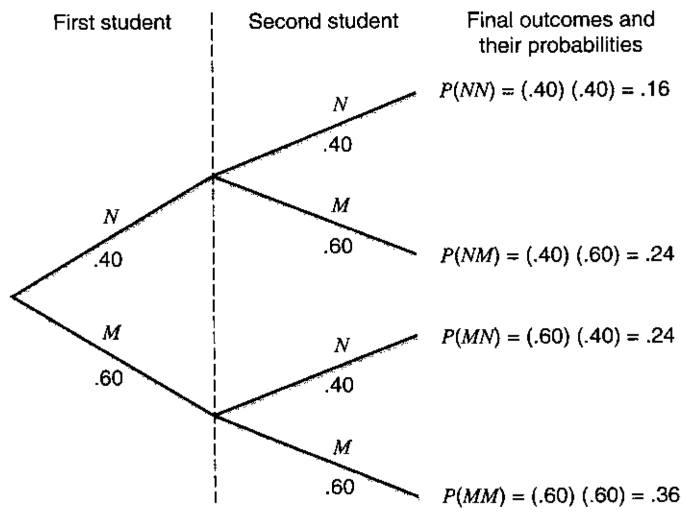


Figure 5.3 Tree diagram.

In a sample of two students, the number who suffer from math anxiety can be 0 (*NN*), 1 (*NM* or *MN*), or 2 (*MM*). Thus, x can assume any of three possible values: 0, 1, or 2. The probabilities of these three outcomes are calculated as follows:

$$P(x = 0) = P(NN) = .16$$

$$P(x = 1) = P(NM \text{ or } MN) = P(NM) + P(MN) = .24 + .24 = .48$$

$$P(x = 2) = P(MM) = .36$$

Using these probabilities, we can write the probability distribution of x as in Table 5.5.

Table 5.5 Probability Distribution of the Number of Students with Math Anxiety in a Sample of Two Students

x	$P(x)$
0	.16
1	.48
2	.36
	$\Sigma P(x) = 1.00$

5.17 According to a survey of adults conducted in April–May 2002 by the Pew Research Center for the People and the Press, 64% of adults owned cell phones (*USA TODAY*, August 2, 2002). Assume that this result holds true for the current population of all adults. Suppose that two adults are selected at random. Let x denote the number of adults in this sample who own a cell phone. Construct the probability distribution table of x. Draw a tree diagram for this problem.

5.18 According to a survey, 30% of adults are against using animals for research. Assume that this result holds true for the current population of all adults. Let x be the number of adults who are against using animals for research in a random sample of two adults. Obtain the probability distribution of x. Draw a tree diagram for this problem.

5.19 According to survey results from the Pew Internet and American Life Project, 78% of teens surveyed said that teachers are "totally clueless" about using the Internet for teaching and learning (*USA TODAY*, August 15, 2002). Assume that this result holds true for the current population of teens. Suppose that two teens are selected at random. Let x be the number of teens in this sample who hold this opinion. Write the probability distribution of x. Draw a tree diagram for this problem.

***5.20** In a group of 12 persons, three are left-handed. Suppose that two persons are randomly selected from this group. Let x denote the number of left-handed persons in this sample. Write the probability distribution of x. You may draw a tree diagram and use it to write the probability distribution. (*Hint:* Note that the draws are made without replacement from a small population. Hence, the probabilities of outcomes do not remain constant for each draw.)

***5.21** In a group of 20 athletes, 6 have used performance-enhancing drugs that are illegal. Suppose that two athletes are randomly selected from this group. Let x denote the number of athletes in this sample who have used these illegal drugs. Write the probability distribution of x. You may draw a tree diagram and use that to write the probability distribution. (*Hint:* Note that the draws are made without replacement from a small population. Hence, the probabilities of outcomes do not remain constant for each draw.)

5.3 MEAN OF A DISCRETE RANDOM VARIABLE

The **mean of a discrete random variable**, denoted by μ, is actually the mean of its probability distribution. The mean of a discrete random variable x is also called its *expected value* and is denoted by $E(x)$. The mean (or expected value) of a discrete random variable is the value that we expect to observe per repetition, on average, if we perform an experiment a large number of times. For example, we may expect a car salesperson to sell, on average, 2.4 cars per week. This does not mean that every week this salesperson will sell exactly 2.4 cars. (Obviously one cannot sell exactly 2.4 cars.) This simply means that if we observe for many weeks, this salesperson will sell a different number of cars during different weeks; however, the average for all these weeks will be 2.4 cars.

To calculate the mean of a discrete random variable x, we multiply each value of x by the corresponding probability and sum the resulting products. This sum gives the mean (or expected value) of the discrete random variable x.

Mean of a Discrete Random Variable The *mean of a discrete random variable x* is the value that is expected to occur per repetition, on average, if an experiment is repeated a large number of times. It is denoted by μ and calculated as

$$\mu = \Sigma x P(x)$$

The mean of a discrete random variable x is also called its expected value and is denoted by $E(x)$; that is,

$$E(x) = \Sigma x P(x)$$

Example 5–5 illustrates the calculation of the mean of a discrete random variable.

EXAMPLE 5–5

Calculating and interpreting the mean of a discrete random variable.

Recall Example 5–3 of Section 5.2. The probability distribution Table 5.4 from that example is reproduced below. In this table, x represents the number of breakdowns for a machine during a given week, and $P(x)$ is the probability of the corresponding value of x.

x	$P(x)$
0	.15
1	.20
2	.35
3	.30

Find the mean number of breakdowns per week for this machine.

Solution To find the mean number of breakdowns per week for this machine, we multiply each value of x by its probability and add these products. This sum gives the mean of the probability distribution of x. The products $xP(x)$ are listed in the third column of Table 5.6. The sum of these products gives $\Sigma xP(x)$, which is the mean of x.

Table 5.6 Calculating the Mean for the Probability Distribution of Breakdowns

x	$P(x)$	$xP(x)$
0	.15	$0(.15) = .00$
1	.20	$1(.20) = .20$
2	.35	$2(.35) = .70$
3	.30	$3(.30) = .90$
		$\Sigma xP(x) = 1.80$

The mean is

$$\mu = \Sigma xP(x) = \mathbf{1.80}$$

Thus, on average, this machine is expected to break down 1.80 times per week over a period of time. In other words, if this machine is used for many weeks, then for certain weeks we will observe no breakdowns; for some other weeks we will observe one breakdown per week; and for still other weeks we will observe two or three breakdowns per week. The mean number of breakdowns is expected to be 1.80 per week for the entire period.

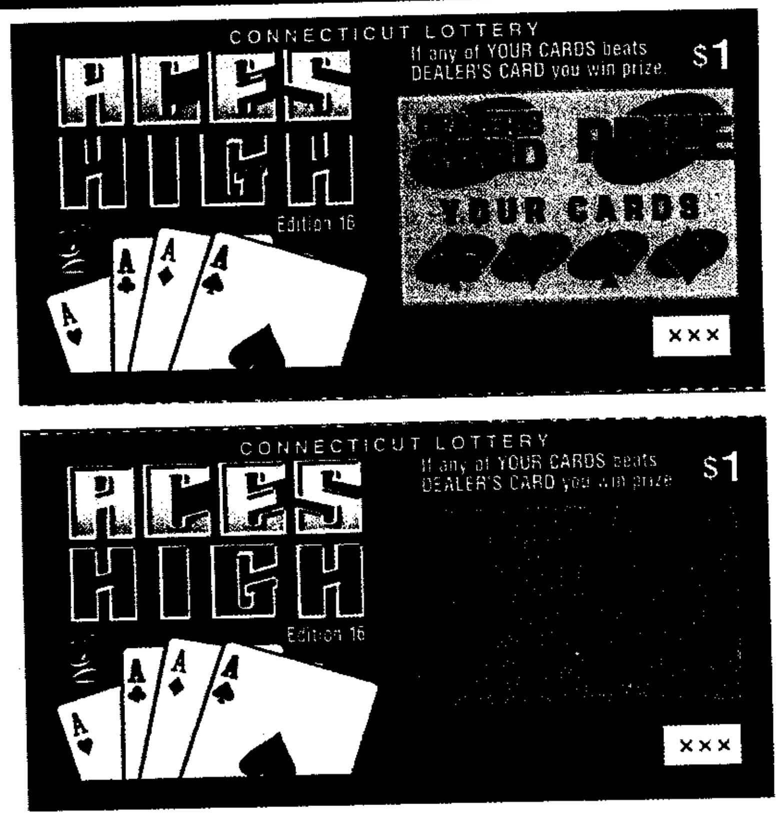
Note that $\mu = 1.80$ is also the expected value of x. It can also be written as

$$E(x) = 1.80$$

Case Study 5–1 illustrates the calculation of the mean amount that an instant lottery ticket is expected to win.

CASE STUDY 5–1

ACES HIGH INSTANT LOTTERY GAME



Currently (2002) the state of Connecticut has in circulation an instant lottery game called Aces High, Edition 16. The cost of each ticket for this game is \$1. A player can instantly win \$1000, \$100, \$40, \$25, \$10, \$4, \$2, or a free ticket (which is equivalent to winning \$1). Each ticket has six "scratchable" spots, one of which contains dealer's card, four spots contain player's cards, and one spot shows the prize amount. A player will win the prize shown in the prize spot if any of the player's cards beats the dealer's card.

Based on the information from the prize structure dated September 13, 2002, Version A, the first table on the next page lists the number of tickets with different prizes in a total of 20,160,000 tickets printed. As is obvious from this table, out of a total of 20,160,000 tickets, 15,875,580 are nonwinning tickets (the ones with a prize of \$0 in this table). Of the remaining tickets, 2,016,000 have a prize of free ticket, 1,451,520 have a prize of \$2 each, and so on.

5.4 STANDARD DEVIATION OF A DISCRETE RANDOM VARIABLE

The **standard deviation of a discrete random variable**, denoted by σ, measures the spread of its probability distribution. A higher value for the standard deviation of a discrete random variable indicates that x can assume values over a larger range about the mean. In contrast, The net gain to a player for each of the instant winning tickets is equal to the amount of the prize minus \$1, which is the cost of the ticket. Thus, the net gain for each of the nonwinning tickets is −\$1, which is the cost of the ticket. Let

x = the net amount a player wins by playing this lottery

The second table below shows the probability distribution of x, and all the calculations required to compute the mean of x for this probability distribution. The probability of an outcome (net winnings) is calculated by dividing the number of tickets with that outcome by the total number of tickets.

Prize (dollars)	Number of Tickets
0	15,875,580
Free ticket	2,016,000
2	1,451,520
4	544,320
10	161,280
25	80,640
40	17,304
100	13,188
1000	168
	Total = 20,160,000

x (dollars)	$P(x)$	$xP(x)$
−1	15,875,580/20,160,000 = .78747917	−.78747917
0	2,016,000/20,160,000 = .10000000	.00000000
1	1,451,520/20,160,000 = .07200000	.07200000
3	544,320/20,160,000 = .02700000	.08100000
9	161,280/20,160,000 = .00800000	.07200000
24	80,640/20,160,000 = .00400000	.09600000
39	17,304/20,160,000 = .00085833	.03347487
99	13,188/20,160,000 = .00065417	.06476283
999	168/20,160,000 = .00000833	.00832167
		$\Sigma xP(x)$ = −.35991980

Hence, the mean or expected value of x is

$$\mu = \Sigma xP(x) = -\$.35991980 \approx -\$.3599$$

This mean gives the expected value of the random variable x, that is,

$$E(x) = \Sigma xP(x) = -\$.3599$$

Thus, the mean of net winnings for this lottery is −\$.3599. In other words, all players taken together will lose an average of \$.3599 (or approximately 36 cents) per ticket. This can also be interpreted as follows. Only 100 − 35.99 = 64.01% of the total money spent by all players on buying lottery tickets for this game will be returned to them in the form of prizes, and 35.99% will not be returned. (The money that will not be returned to players will cover the costs of operating the lottery, the commission paid to agents, and revenue to the state of Connecticut.)

Source: Connecticut Lottery Corporation. Lottery ticket reproduced with permission.

A smaller value for the standard deviation indicates that most of the values that x can assume are clustered closely about the mean. The basic formula to compute the standard deviation of a discrete random variable is

$$\sigma = \sqrt{\Sigma[(x - \mu)^2 \cdot P(x)]}$$

However, it is more convenient to use the following shortcut formula to compute the standard deviation of a discrete random variable.

Standard Deviation of a Discrete Random Variable The *standard deviation of a discrete random variable x* measures the spread of its probability distribution and is computed as

$$\sigma = \sqrt{\Sigma x^2 P(x) - \mu^2}$$

Note that the variance σ^2 of a discrete random variable is obtained by squaring its standard deviation.

Example 5–6 illustrates how to use the shortcut formula to compute the standard deviation of a discrete random variable.

EXAMPLE 5–6

Calculating the standard deviation of a discrete random variable.

Baier's Electronics manufactures computer parts that are supplied to many computer companies. Despite the fact that two quality control inspectors at Baier's Electronics check every part for defects before it is shipped to another company, a few defective parts do pass through these inspections undetected. Let x denote the number of defective computer parts in a shipment of 400. The following table gives the probability distribution of x.

x	0	1	2	3	4	5
$P(x)$	.02	.20	.30	.30	.10	.08

Compute the standard deviation of x.

Solution Table 5.7 shows all the calculations required for the computation of the standard deviation of x.

Table 5.7 Computations to Find the Standard Deviation

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
0	.02	.00	0	.00
1	.20	.20	1	.20
2	.30	.60	4	1.20
3	.30	.90	9	2.70
4	.10	.40	16	1.60
5	.08	.40	25	2.00
		$\Sigma xP(x) = 2.50$		$\Sigma x^2P(x) = 7.70$

We perform the following steps to compute the standard deviation of x.

Step 1. Compute the mean of the discrete random variable.

The sum of the products $xP(x)$, recorded in the third column of Table 5.7, gives the mean of x.

$$\mu = \Sigma xP(x) = 2.50 \text{ defective computer parts in 400}$$

Step 2. Compute the value of $\Sigma x^2 P(x)$.

First we square each value of x and record it in the fourth column of Table 5.7. Then we multiply these values of x^2 by the corresponding values of $P(x)$. The resulting values of

$x^2P(x)$ are recorded in the fifth column of Table 5.7. The sum of this column is

$$\Sigma x^2 P(x) = 7.70$$

Step 3. Substitute the values of μ and $\Sigma x^2P(x)$ in the formula for the standard deviation of x and simplify.

By performing this step, we obtain

$$\sigma = \sqrt{\Sigma x^2 P(x) - \mu^2} = \sqrt{7.70 - (2.50)^2} = \sqrt{1.45}$$
$$= \mathbf{1.204} \text{ defective computer parts}$$

Thus, a given shipment of 400 computer parts is expected to contain an average of 2.50 defective parts with a standard deviation of 1.204.

◀ ***Remember*** Because the standard deviation of a discrete random variable is obtained by taking the positive square root, its value is never negative.

EXAMPLE 5–7

Loraine Corporation is planning to market a new makeup product. According to the analysis made by the financial department of the company, it will earn an annual profit of \$4.5 million if this product has high sales and an annual profit of \$1.2 million if the sales are mediocre, and it will lose \$2.3 million a year if the sales are low. The probabilities of these three scenarios are .32, .51, and .17, respectively.

(a) Let x be the profits (in millions of dollars) earned per annum by the company from this product. Write the probability distribution of x.

(b) Calculate the mean and standard deviation of x.

Solution

Writing the probability distribution of a discrete random variable.

(a) The following table lists the probability distribution of x. Note that because x denotes profits earned by the company, the loss is written as a *negative profit* in the table.

x	$P(x)$
4.5	.32
1.2	.51
−2.3	.17

Calculating the mean and standard deviation of a discrete random variable.

(b) Table 5.8 shows all the calculations needed for the computation of the mean and standard deviation of x.

Table 5.8 Computations to Find the Mean and Standard Deviation

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
4.5	.32	1.440	20.25	6.4800
1.2	.51	.612	1.44	.7344
−2.3	.17	−.391	5.29	.8993
		$\Sigma xP(x) = 1.661$		$\Sigma x^2P(x) = 8.1137$

The mean of x is

$$\mu = \Sigma x P(x) = \textbf{\$1.661 million}$$

The standard deviation of x is

$$\sigma = \sqrt{\Sigma x^2 P(x) - \mu^2} = \sqrt{8.1137 - (1.661)^2} = \textbf{\$2.314 million}$$

Thus, it is expected that Loraine Corporation will earn an average of \$1.661 million in profits per year from the new product with a standard deviation of \$2.314 million.

▶ Interpretation of the Standard Deviation

The standard deviation of a discrete random variable can be interpreted or used the same way as the standard deviation of a data set in Section 3.4 of Chapter 3. In that section, we learned that according to Chebyshev's theorem, at least $[1 - (1/k^2)] \times 100\%$ of the total area under a curve lies within k standard deviations of the mean, where k is any number greater than 1. Thus, if $k = 2$, then at least 75% of the area under a curve lies between $\mu - 2\sigma$ and $\mu + 2\sigma$. In Example 5–6,

$$\mu = 2.50 \quad \text{and} \quad \sigma = 1.204$$

Hence,

$$\mu - 2\sigma = 2.50 - 2(1.204) = .092$$

$$\mu + 2\sigma = 2.50 + 2(1.204) = 4.908$$

Using Chebyshev's theorem, we can state that at least 75% of the shipments (each containing 400 computer parts) are expected to contain .092 to 4.908 defective computer parts each.

EXERCISES

CONCEPTS AND PROCEDURES

5.22 Briefly explain the concept of the mean and standard deviation of a discrete random variable.

5.23 Find the mean and standard deviation for each of the following probability distributions.

a.

x	$P(x)$
0	.16
1	.27
2	.39
3	.18

b.

x	$P(x)$
6	.40
7	.26
8	.21
9	.13

5.24 Find the mean and standard deviation for each of the following probability distributions.

a.

x	$P(x)$
3	.09
4	.21
5	.34
6	.23
7	.13

b.

x	$P(x)$
0	.43
1	.31
2	.17
3	.09

5.33 Refer to the probability distribution you developed in Exercise 5.16 for the number of lemons in two selected cars. Calculate the mean and standard deviation of x for that probability distribution.

5.34 Refer to the probability distribution you developed in Exercise 5.17 for the number of adults in a sample of two who own a cell phone. Compute the mean and standard deviation of x for that probability distribution.

5.35 A contractor has submitted bids on three state jobs: an office building, a theater, and a parking garage. State rules do not allow a contractor to be offered more than one of these jobs. If this contractor is awarded any of these jobs, the profits earned from these contracts are: $10 million from the office building, $5 million from the theater, and $2 million from the parking garage. His profit is zero if he gets no contract. The contractor estimates that the probabilities of getting the office building contract, the theater contract, the parking garage contract, or nothing are .15, .30, .45, and .10, respectively. Let x be the random variable that represents the contractor's profits in millions of dollars. Write the probability distribution of x. Find the mean and standard deviation of x. Give a brief interpretation of the values of the mean and standard deviation.

5.36 An instant lottery ticket costs $2. Out of a total of 10,000 tickets printed for this lottery, 1000 tickets contain a prize of $5 each, 100 tickets have a prize of $10 each, 5 tickets have a prize of $1000 each, and 1 ticket has a prize of $5000. Let x be the random variable that denotes the net amount a player wins by playing this lottery. Write the probability distribution of x. Determine the mean and standard deviation of x. How will you interpret the values of the mean and standard deviation of x?

***5.37** Refer to the probability distribution you developed in Exercise 5.20 for the number of left-handed persons in a sample of two persons. Calculate the mean and standard deviation of x for that distribution.

***5.38** Refer to the probability distribution you developed in Exercise 5.21 for the number of athletes in a random sample of two who have used illegal performance-enhancing drugs. Calculate the mean and standard deviation of x for that distribution.

5.5 FACTORIALS AND COMBINATIONS

This section introduces factorials and combinations, which will be used in the binomial formula discussed in Section 5.6.

5.5.1 Factorials

The symbol ! (read as *factorial*) is used to denote **factorials**. The value of the factorial of a number is obtained by multiplying all the integers from that number to 1. For example, 7! is read as "seven factorial" and is evaluated by multiplying all the integers from 7 to 1.

Definition

Factorials The symbol $n!$, read as "n factorial," represents the product of all the integers from n to 1. In other words,

$$n! = n(n-1)(n-2)(n-3)\cdots 3\cdot 2\cdot 1$$

By definition,

$$0! = 1$$

EXAMPLE 5–8

Evaluating a factorial.

Evaluate 7!.

Solution To evaluate 7!, we multiply all the integers from 7 to 1.

$$7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = \mathbf{5040}$$

Thus, the value of 7! is 5040.

EXAMPLE 5–9

Evaluating a factorial.

Evaluate 10!.

Solution The value of 10! is given by the product of all the integers from 10 to 1. Thus,

$$10! = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = \mathbf{3{,}628{,}800}$$

EXAMPLE 5–10

Evaluating a factorial of the difference between two numbers.

Evaluate (12 − 4)!.

Solution The value of (12 − 4)! is

$$(12 - 4)! = 8! = 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = \mathbf{40{,}320}$$

EXAMPLE 5–11

Evaluating a factorial of zero.

Evaluate (5 − 5)!.

Solution The value of (5 − 5)! is 1.

$$(5 - 5)! = 0! = \mathbf{1}$$

Note that 0! is always equal to 1.

We can read the value of $n!$ for $n = 1$ through $n = 25$ from Table II of Appendix C. Example 5–12 illustrates how to read that table.

EXAMPLE 5–12

Using the table of factorials.

Find the value of 15! by using Table II of Appendix C.

Solution To find the value of 15! from Table II, we locate 15 in the column labeled n. Then we read the value in the column for $n!$ entered next to 15. Thus,

$$15! = \mathbf{1{,}307{,}674{,}368{,}000}$$

5.5.2 Combinations

Quite often we face the problem of selecting a few elements from a large number of distinct elements. For example, a student may be required to attempt any two questions out of four in an examination. As another example, the faculty in a department may need to select 3 professors from 20 to form a committee. Or a lottery player may have to pick 6 numbers from 49. The question arises: In how many ways can we make the selections in each of these examples? For instance, how many possible selections exist for the student who is to choose any two questions out of four? The answer is six. Let the four questions be denoted by the numbers 1, 2, 3, and 4. Then the six selections are

(1 and 2) (1 and 3) (1 and 4) (2 and 3) (2 and 4) (3 and 4)

The student can choose questions 1 and 2, or 1 and 3, or 1 and 4, and so on.

Each of the possible selections in this list is called a **combination**. All six combinations are distinct; that is, each combination contains a different set of questions. It is important to remember that the order in which the selections are made is not significant in the case of combinations. Thus, whether we write (1 and 2) or (2 and 1), both these arrangements represent only one combination.

Definition

Combinations Notation *Combinations* give the number of ways x elements can be selected from n elements. The notation used to denote the total number of combinations is

$$_nC_x$$

which is read as "the number of combinations of n elements selected x at a time."

Suppose there are a total of n elements from which we want to select x elements. Then,

n denotes the total number of elements

$_nC_x$ = the number of combinations of n elements selected x at a time

x denotes the number of elements selected per selection

Number of Combinations The *number of combinations* for selecting x from n distinct elements is given by the formula

$$_nC_x = \frac{n!}{x!(n-x)!}$$

where $n!$, $x!$, and $(n-x)!$ are read as "n factorial," "x factorial," and "n minus x factorial," respectively.

In the combinations formula,

$$n! = n(n-1)(n-2)(n-3)\cdots 3\cdot 2\cdot 1$$
$$x! = x(x-1)(x-2)\cdots 3\cdot 2\cdot 1$$
$$(n-x)! = (n-x)(n-x-1)(n-x-2)\cdots 3\cdot 2\cdot 1$$

Note that in combinations, n is always greater than or equal to x. If n is less than x, then we cannot select x distinct elements from n.

EXAMPLE 5–13

Finding the number of combinations using the formula.

An ice cream parlor has six flavors of ice cream. Kristen wants to buy two flavors of ice cream. If she randomly selects two flavors out of six, how many possible combinations are there?

Solution For this example,

$$n = \text{total number of ice cream flavors} = 6$$

$$x = \text{number of ice cream flavors to be selected} = 2$$

Therefore, the number of ways in which Kristen can select two flavors of ice cream out of six is

$${}_6C_2 = \frac{6!}{2!(6-2)!} = \frac{6!}{2!\,4!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \mathbf{15}$$

Thus, there are 15 ways for Kristen to select two ice cream flavors out of six.

EXAMPLE 5–14

Finding the number of combinations and listing them.

Three members of a jury will be randomly selected from five people. How many different combinations are possible?

Solution There are a total of five persons, and we are to select three of them. Hence,

$$n = 5 \quad \text{and} \quad x = 3$$

Applying the combinations formula, we get

$${}_5C_3 = \frac{5!}{3!(5-3)!} = \frac{5!}{3!\,2!} = \frac{120}{6 \cdot 2} = \mathbf{10}$$

If we assume that the five persons are A, B, C, D, and E, then the 10 possible combinations for the selection of three members of the jury are

ABC ABD ABE ACD ACE ADE BCD BCE BDE CDE

Case Study 5–2 on the next page describes the number of ways a lottery player can select six numbers in a lotto game.

5.5.3 Using the Table of Combinations

Table III in Appendix C lists the number of combinations of n elements selected x at a time. The following example on page 216 illustrates how to read that table to find combinations.

APPLICATIONS

5.41 An English department at a university has 16 faculty members. Two of the faculty members will be randomly selected to represent the department on a committee. In how many ways can the department select 2 faculty members from 16? Use the appropriate formula.

5.42 An ice cream shop offers 25 flavors of ice cream. How many ways are there to select 2 different flavors from these 25 flavors? Use the appropriate formula.

5.43 A veterinarian assigned to a racetrack has received a tip that 1 or more of the 12 horses in the third race have been doped. She has time to test only three horses. How many ways are there to randomly select 3 horses from these 12 horses? Use the appropriate formula. Verify your answer by using Table III of Appendix C.

5.44 An environmental agency will randomly select 4 houses from a block containing 25 houses for a radon check. How many total selections are possible? Use the appropriate formula. Verify your answer by using Table III of Appendix C.

5.45 An investor will randomly select 6 stocks from 20 for an investment. How many total combinations are possible? Use the appropriate formula. Verify your answer by using Table III of Appendix C.

5.46 A company employs a total of 16 workers. The management has asked these employees to select two workers who will negotiate a new contract with management. The employees have decided to select the two workers randomly. How many total selections are possible? Use the appropriate formula. Verify your answer by using Table III of Appendix C.

5.47 In how many ways can a sample (without replacement) of 9 items be selected from a population of 20 items?

5.48 In how many ways can a sample (without replacement) of 5 items be selected from a population of 15 items?

5.6 THE BINOMIAL PROBABILITY DISTRIBUTION

The **binomial probability distribution** is one of the most widely used discrete probability distributions. It is applied to find the probability that an outcome will occur x times in n performances of an experiment. For example, given that the probability is .05 that a VCR manufactured at a firm is defective, we may be interested in finding the probability that in a random sample of three VCRs manufactured at this firm, exactly one will be defective. As a second example, we may be interested in finding the probability that a baseball player with a batting average of .250 will have no hits in 10 trips to the plate.

To apply the binomial probability distribution, the random variable x must be a discrete dichotomous random variable. In other words, the variable must be a discrete random variable and each repetition of the experiment must result in one of two possible outcomes. The binomial distribution is applied to experiments that satisfy the four conditions of a *binomial experiment*. (These conditions are described in Section 5.6.1.) Each repetition of a binomial experiment is called a **trial** or a **Bernoulli trial** (after Jacob Bernoulli). For example, if an experiment is defined as one toss of a coin and this experiment is repeated 10 times, then each repetition (toss) is called a trial. Consequently, there are 10 total trials for this experiment.

5.6.1 The Binomial Experiment

An experiment that satisfies the following four conditions is called a **binomial experiment**.

1. There are n identical trials. In other words, the given experiment is repeated n times. All these repetitions are performed under identical conditions.
2. Each trial has two and only two outcomes. These outcomes are usually called a *success* and a *failure*.

3. The probability of success is denoted by p and that of failure by q, and $p + q = 1$. The probabilities p and q remain constant for each trial.
4. The trials are independent. In other words, the outcome of one trial does not affect the outcome of another trial.

Conditions of a Binomial Experiment A binomial experiment must satisfy the following four conditions.

1. There are n identical trials.
2. Each trial has only two possible outcomes.
3. The probabilities of the two outcomes remain constant.
4. The trials are independent.

Note that one of the two outcomes of a trial is called a *success* and the other a *failure.* Notice that a success does not mean that the corresponding outcome is considered favorable or desirable. Similarly, a failure does not necessarily refer to an unfavorable or undesirable outcome. Success and failure are simply the names used to denote the two possible outcomes of a trial. The outcome to which the question refers is usually called a success; the outcome to which it does not refer is called a failure.

EXAMPLE 5–16

Verifying the conditions of a binomial experiment.

Consider the experiment consisting of 10 tosses of a coin. Determine whether or not it is a binomial experiment.

Solution The experiment consisting of 10 tosses of a coin satisfies all four conditions of a binomial experiment.

1. There are a total of 10 trials (tosses), and they are all identical. All 10 tosses are performed under identical conditions.
2. Each trial (toss) has only two possible outcomes: a head and a tail. Let a head be called a success and a tail be called a failure.
3. The probability of obtaining a head (a success) is 1/2 and that of a tail (a failure) is 1/2 for any toss. That is,

$$p = P(H) = 1/2 \quad \text{and} \quad q = P(T) = 1/2$$

The sum of these two probabilities is 1.0. Also, these probabilities remain the same for each toss.

4. The trials (tosses) are independent. The result of any preceding toss has no bearing on the result of any succeeding toss.

Consequently, the experiment consisting of 10 tosses is a binomial experiment.

EXAMPLE 5–17

Verifying the conditions of a binomial experiment.

Five percent of all VCRs manufactured by a large electronics company are defective. Three VCRs are randomly selected from the production line of this company. The selected VCRs are inspected to determine whether each of them is defective or good. Is this experiment a binomial experiment?

Solution

1. This example consists of three identical trials. A trial represents the selection of a VCR.
2. Each trial has two outcomes: a VCR is defective or a VCR is good. Let a defective VCR be called a success and a good VCR be called a failure.
3. Five percent of all VCRs are defective. So, the probability p that a VCR is defective is .05. As a result, the probability q that a VCR is good is .95. These two probabilities add up to 1.
4. Each trial (VCR) is independent. In other words, if one VCR is defective, it does not affect the outcome of another VCR being defective or good. This is so because the size of the population is very large compared to the sample size.

Because all four conditions of a binomial experiment are satisfied, this is an example of a binomial experiment.

5.6.2 The Binomial Probability Distribution and Binomial Formula

The random variable x that represents the number of successes in n trials for a binomial experiment is called a *binomial random variable*. The probability distribution of x in such experiments is called the **binomial probability distribution** or simply the *binomial distribution*. Thus, the binomial probability distribution is applied to find the probability of x successes in n trials for a binomial experiment. The number of successes x in such an experiment is a discrete random variable. Consider Example 5–17. Let x be the number of defective VCRs in a sample of three. Because we can obtain any number of defective VCRs from zero to three in a sample of three, x can assume any of the values 0, 1, 2, and 3. Since the values of x are countable, it is a discrete random variable.

Binomial Formula For a binomial experiment, the probability of exactly x successes in n trials is given by the binomial formula

$$P(x) = {}_nC_x p^x q^{n-x}$$

where

$$\begin{aligned} n &= \text{total number of trials} \\ p &= \text{probability of success} \\ q &= 1 - p = \text{probability of failure} \\ x &= \text{number of successes in } n \text{ trials} \\ n - x &= \text{number of failures in } n \text{ trials} \end{aligned}$$

In the binomial formula, n is the total number of trials and x is the total number of successes. The difference between the total number of trials and the total number of successes, $n - x$, gives the total number of failures in n trials. The value of ${}_nC_x$ gives the number of ways to obtain x successes in n trials. As mentioned earlier, p and q are the probabilities of success and failure, respectively. Again, although it does not matter which of the two outcomes is called a success and which one a failure, usually the outcome to which the question refers is called a success.

To solve a binomial problem, we determine the values of n, x, $n - x$, p, and q and then substitute these values in the binomial formula. To find the value of ${}_nC_x$, we can use either

$_nC_x = \frac{n!}{x!(n-x)!}$

the combinations formula from Section 5.5.2 or the table of combinations (Table III of Appendix C).

To find the probability of x successes in n trials for a binomial experiment, the only values needed are those of n and p. These are called the *parameters of the binomial probability distribution* or simply the **binomial parameters**. The value of q is obtained by subtracting the value of p from 1.0. Thus, $q = 1 - p$.

Next we solve a binomial problem, first without using the binomial formula and then by using the binomial formula.

EXAMPLE 5–18

Calculating the probability using a tree diagram and the binomial formula.

Five percent of all VCRs manufactured by a large electronics company are defective. A quality control inspector randomly selects three VCRs from the production line. What is the probability that exactly one of these three VCRs is defective?

Solution Let

$$D = \text{a selected VCR is defective}$$

$$G = \text{a selected VCR is good}$$

As the tree diagram in Figure 5.4 shows, there are a total of eight outcomes, and three of them contain exactly one defective VCR. These three outcomes are

$$DGG, \quad GDG, \quad \text{and} \quad GGD$$

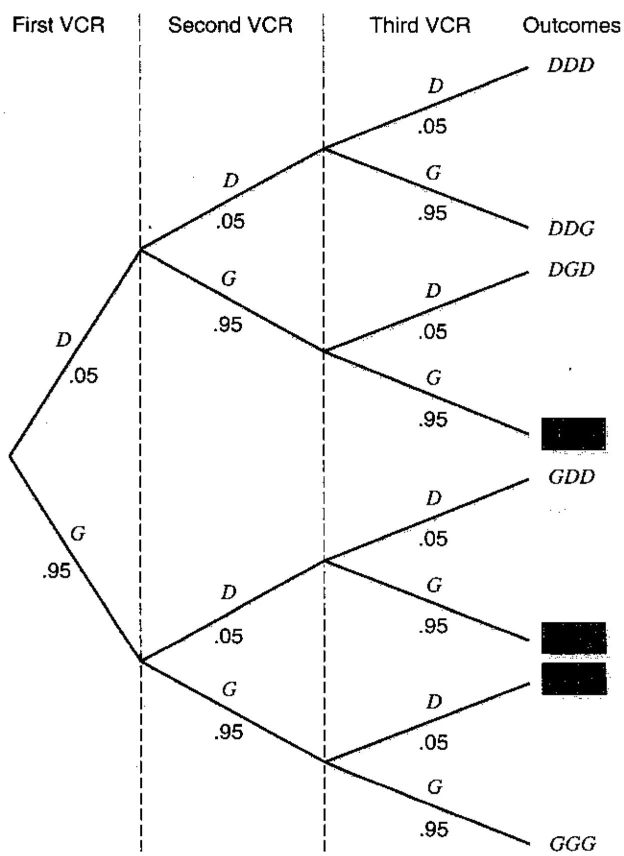


Figure 5.4 Tree diagram for selecting three VCRs.

We know that 5% of all VCRs manufactured at this company are defective. As a result, 95% of all VCRs are good. So the probability that a randomly selected VCR is defective is .05 and the probability that it is good is .95.

$$P(D) = .05 \quad \text{and} \quad P(G) = .95$$

Because the size of the population is large (note that it is a large company), the selections can be considered to be independent. The probability of each of the three outcomes, which give exactly one defective VCR, is calculated as follows:

$$P(DGG) = P(D) \cdot P(G) \cdot P(G) = (.05)(.95)(.95) = .0451$$
$$P(GDG) = P(G) \cdot P(D) \cdot P(G) = (.95)(.05)(.95) = .0451$$
$$P(GGD) = P(G) \cdot P(G) \cdot P(D) = (.95)(.95)(.05) = .0451$$

Note that *DGG* is simply the intersection of the three events *D*, *G*, and *G*. In other words, *P*(*DGG*) is the joint probability of three events: the first VCR selected is defective, the second is good, and the third is good. To calculate this probability, we use the multiplication rule for independent events we learned in Chapter 4. The same is true about the probabilities of the other two outcomes: *GDG* and *GGD*.

Exactly one defective VCR will be selected if *DGG* or *GDG* or *GGD* occurs. These are three mutually exclusive outcomes. Therefore, from the addition rule of Chapter 4, the probability of the union of these three outcomes is simply the sum of their individual probabilities.

$$\begin{aligned} P(1 \text{ VCR in 3 is defective}) &= P(DGG \text{ or } GDG \text{ or } GGD) \\ &= P(DGG) + P(GDG) + P(GGD) \\ &= .0451 + .0451 + .0451 = \mathbf{.1353} \end{aligned}$$

Now let us use the binomial formula to compute this probability. Let us call the selection of a defective VCR a *success* and the selection of a good VCR a *failure*. The reason we have called a defective VCR a *success* is that the question refers to selecting exactly one defective VCR. Then,

$$\begin{aligned} n &= \text{total number of trials} = 3 \text{ VCRs} \\ x &= \text{number of successes} = \text{number of defective VCRs} = 1 \\ n - x &= \text{number of failures} = \text{number of good VCRs} = 3 - 1 = 2 \\ p &= P(\text{success}) = .05 \\ q &= P(\text{failure}) = 1 - p = .95 \end{aligned}$$

The probability of 1 success is denoted by $P(x = 1)$ or simply by $P(1)$. By substituting all the values in the binomial formula, we obtain

Number of ways to obtain 1 success in 3 trials — Number of successes — Number of failures

$$P(x = 1) = {}_3C_1(.05)^1(.95)^2 = (3)(.05)(.9025) = \mathbf{.1354}$$

Probability of success — Probability of failure

Note that the value of ${}_3C_1$ in the formula either can be read from Table III of Appendix C or can be computed as follows:

$$_3C_1 = \frac{3!}{1!(3-1)!} = \frac{3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 1} = 3$$

In the above computation, $_3C_1$ gives the three ways to select one defective VCR in three selections. As listed previously, these three ways to select one defective VCR are *DGG*, *GDG*, and *GGD*. The probability .1354 is slightly different from the earlier calculation (.1353) because of rounding.

EXAMPLE 5–19

Calculating the probability using the binomial formula.

At the Express House Delivery Service, providing high-quality service to customers is the top priority of the management. The company guarantees a refund of all charges if a package it is delivering does not arrive at its destination by the specified time. It is known from past data that despite all efforts, 2% of the packages mailed through this company do not arrive at their destinations within the specified time. Suppose a corporation mails 10 packages through Express House Delivery Service on a certain day.

(a) Find the probability that exactly 1 of these 10 packages will not arrive at its destination within the specified time.

(b) Find the probability that at most 1 of these 10 packages will not arrive at its destination within the specified time.

Solution Let us call it a success if a package does not arrive at its destination within the specified time and a failure if it does arrive within the specified time. Then,

$$n = \text{total number of packages mailed} = 10$$

$$p = P(\text{success}) = .02$$

$$q = P(\text{failure}) = 1 - .02 = .98$$

(a) For this part,

$$x = \text{number of successes} = 1$$

$$n - x = \text{number of failures} = 10 - 1 = 9$$

Substituting all values in the binomial formula, we obtain

$$P(x = 1) = {}_{10}C_1(.02)^1(.98)^9 = \frac{10!}{1!(10-1)!}(.02)^1(.98)^9$$

$$= (10)(.02)(.83374776) = \mathbf{.1667}$$

Thus, there is a .1667 probability that exactly one of the 10 packages mailed will not arrive at its destination within the specified time.

(b) The probability that at most one of the 10 packages will not arrive at its destination within the specified time is given by the sum of the probabilities of $x = 0$ and $x = 1$. Thus,

$$P(x \leq 1) = P(x = 0) + P(x = 1)$$

$$= {}_{10}C_0(.02)^0(.98)^{10} + {}_{10}C_1(.02)^1(.98)^9$$

$$= (1)(1)(.81707281) + (10)(.02)(.83374776)$$

$$= .8171 + .1667 = \mathbf{.9838}$$

Thus, the probability that at most one of the 10 packages will not arrive at its destination within the specified time is .9838.

EXAMPLE 5–20

Constructing a binomial probability distribution and its graph.

According to an Allstate Survey, 56% of Baby Boomers have car loans and are making payments on these loans (*USA TODAY*, October 28, 2002). Assume that this result holds true for the current population of all Baby Boomers. Let x denote the number in a random sample of three Baby Boomers who are making payments on their car loans. Write the probability distribution of x and draw a bar graph for this probability distribution.

Solution Let x be the number of Baby Boomers in a sample of three who are making payments on their car loans. Then, $n - x$ is the number of Baby Boomers who do not have car loans and, hence, are not making such payments. From the given information,

$$n = \text{total Baby Boomers in the sample} = 3$$

$$p = P(\text{a Baby Boomer is making car loan payments}) = .56$$

$$q = P(\text{a Baby Boomer is not making car loan payments}) = 1 - .56 = .44$$

The possible values that x can assume are 0, 1, 2, and 3. In other words, the number of Baby Boomers who are making car loan payments can be 0, 1, 2, or 3. The probability of each of these four outcomes is calculated as follows.

If $x = 0$, then $n - x = 3$. From the binomial formula, the probability of $x = 0$ is

$$P(x = 0) = {}_3C_0(.56)^0(.44)^3 = (1)(1)(.085184) = .0852$$

Note that ${}_3C_0$ is equal to 1 by definition and $(.56)^0$ is equal to 1 because any number raised to the power zero is always 1.

If $x = 1$, then $n - x = 2$. From the binomial formula, the probability of $x = 1$ is

$$P(x = 1) = {}_3C_1(.56)^1(.44)^2 = (3)(.56)(.1936) = .3252$$

Similarly, if $x = 2$, then $n - x = 1$, and if $x = 3$, then $n - x = 0$. The probabilities of $x = 2$ and $x = 3$ are

$$P(x = 2) = {}_3C_2(.56)^2(.44)^1 = (3)(.3136)(.44) = .4140$$

$$P(x = 3) = {}_3C_3(.56)^3(.44)^0 = (1)(.175616)(1) = .1756$$

These probabilities are written in Table 5.10. Figure 5.5 shows the bar graph for the probability distribution of Table 5.10.

Table 5.10 Probability Distribution of x

x	$P(x)$
0	.0852
1	.3252
2	.4140
3	.1756

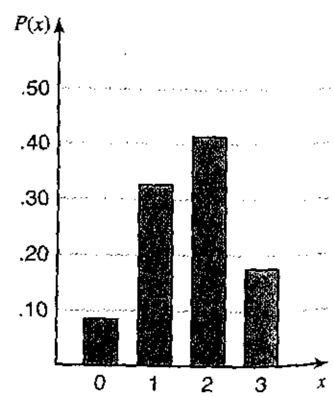


Figure 5.5 Bar graph of the probability distribution of x.

5.6.3 Using the Table of Binomial Probabilities

The probabilities for a binomial experiment can also be read from Table IV, the table of binomial probabilities, in Appendix C. That table lists the probabilities of x for $n = 1$ to $n = 25$ and for selected values of p. Example 5–21 illustrates how to read Table IV.

EXAMPLE 5–21

Using the binomial table to find probabilities and construct the probability distribution and graph.

According to a 2001 study of college students by Harvard University's School of Public Health, 19.3% of those included in the study abstained from drinking (*USA TODAY*, April 3, 2002). Suppose that of all current college students in the United States, 20% abstain from drinking. A random sample of six college students is selected. Using Table IV of Appendix C, answer the following.

(a) Find the probability that exactly three college students in this sample abstain from drinking.
(b) Find the probability that at most two college students in this sample abstain from drinking.
(c) Find the probability that at least three college students in this sample abstain from drinking.
(d) Find the probability that one to three college students in this sample abstain from drinking.
(e) Let x be the number of college students in this sample who abstain from drinking. Write the probability distribution of x and draw a bar graph for this probability distribution.

Solution

(a) To read the required probability from Table IV of Appendix C, we first determine the values of n, x, and p. For this example,

$$n = \text{number of college students in the sample} = 6$$

$$x = \text{number of college students in six who abstain from drinking} = 3$$

$$p = P(\text{a college student abstains from drinking}) = .20$$

Then we locate $n = 6$ in the column labeled n in Table IV. The relevant portion of Table IV with $n = 6$ is reproduced here as Table 5.11. Next, we locate 3 in the column for x in the portion of the table for $n = 6$ and locate $p = .20$ in the row for p at the top of the table. The entry at the intersection of the row for $x = 3$ and the column for $p = .20$ gives the probability of 3 successes in 6 trials when the probability of success is .20. From Table IV or Table 5.11,

$$P(x = 3) = \mathbf{.0819}$$

Table 5.11 Determining $P(x = 3)$ for $n = 6$ and $p = .20$

			p			
n	*x*	.05	.10	.20 ($p = .20$)	...	.95
$n = 6 \longrightarrow$ 6	0	.7351	.5314	.2621	...	.0000
	1	.2321	.3543	.3932	...	.0000
	2	.0305	.0984	.2458	...	.0001
	$x = 3 \longrightarrow$ 3	.0021	.0146	.0819 ← $P(x = 3) = .0819$	...	.0021
	4	.0001	.0012	.0154	...	.0305
	5	.0000	.0001	.0015	...	.2321
	6	.0000	.0000	.0001	...	.7351

Using Table IV or Table 5.11, we write Table 5.12, which can be used to answer the remaining parts of this example.

Table 5.12 Portion of Table IV for $n = 6$ and $p = .20$

		p
n	x	.20
6	0	.2621
	1	.3932
	2	.2458
	3	.0819
	4	.0154
	5	.0015
	6	.0001

(b) The event that at most two college students in this sample abstain from drinking will occur if x is equal to 0, 1, or 2. From Table IV of Appendix C or Table 5.12, the required probability is

$$P(\text{at most } 2) = P(0 \text{ or } 1 \text{ or } 2) = P(x = 0) + P(x = 1) + P(x = 2)$$
$$= .2621 + .3932 + .2458 = \mathbf{.9011}$$

(c) The probability that at least three college students in this sample abstain from drinking is given by the sum of the probabilities of 3, 4, 5, or 6. Using Table IV of Appendix C or Table 5.12,

$$P(\text{at least } 3) = P(3 \text{ or } 4 \text{ or } 5 \text{ or } 6)$$
$$= P(x = 3) + P(x = 4) + P(x = 5) + P(x = 6)$$
$$= .0819 + .0154 + .0015 + .0001 = \mathbf{.0989}$$

(d) The probability that one to three college students in this sample abstain from drinking is given by the sum of the probabilities $x = 1, 2,$ or 3. Using Table IV of Appendix C or Table 5.12,

$$P(1 \text{ to } 3) = P(x = 1) + P(x = 2) + P(x = 3)$$
$$= .3932 + .2458 + .0819 = \mathbf{.7209}$$

(e) Using Table IV of Appendix C or Table 5.12, we list the probability distribution of x for $n = 6$ and $p = .20$ in Table 5.13. Figure 5.6 shows the bar graph of the probability distribution of x.

Table 5.13 Probability Distribution of x for $n = 6$ and $p = .20$

x	$P(x)$
0	.2621
1	.3932
2	.2458
3	.0819
4	.0154
5	.0015
6	.0001

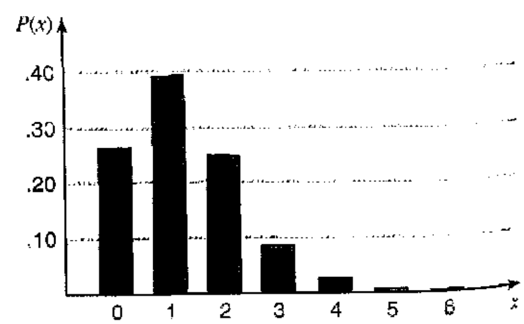


Figure 5.6 Bar graph for the probability distribution of x.

5.6.4 Probability of Success and the Shape of the Binomial Distribution

For any number of trials n:

1. The binomial probability distribution is symmetric if $p = .50$.
2. The binomial probability distribution is skewed to the right if p is less than .50.
3. The binomial probability distribution is skewed to the left if p is greater than .50.

These three cases are illustrated next with examples and graphs.

1. Let $n = 4$ and $p = .50$. Using Table IV of Appendix C, we have written the probability distribution of x in Table 5.14 and plotted it in Figure 5.7. As we can observe from Table 5.14 and Figure 5.7, the probability distribution of x is symmetric.

Table 5.14 Probability Distribution of x for $n = 4$ and $p = .50$

x	$P(x)$
0	.0625
1	.2500
2	.3750
3	.2500
4	.0625

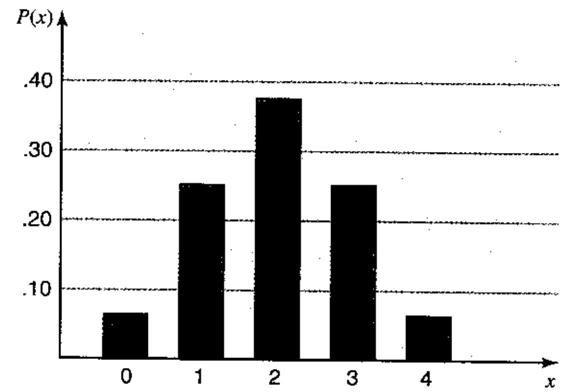


Figure 5.7 Bar graph for the probability distribution of Table 5.14.

2. Let $n = 4$ and $p = .30$ (which is less than .50). Table 5.15, which is written by using Table IV of Appendix C, and the graph of the probability distribution in Figure 5.8 show that the probability distribution of x for $n = 4$ and $p = .30$ is skewed to the right.

Table 5.15 Probability Distribution of x for $n = 4$ and $p = .30$

[illegible]	$P(x)$
[illegible]	.2401
[illegible]	.4116
[illegible]	.2646
[illegible]	.0756
[illegible]	.0081

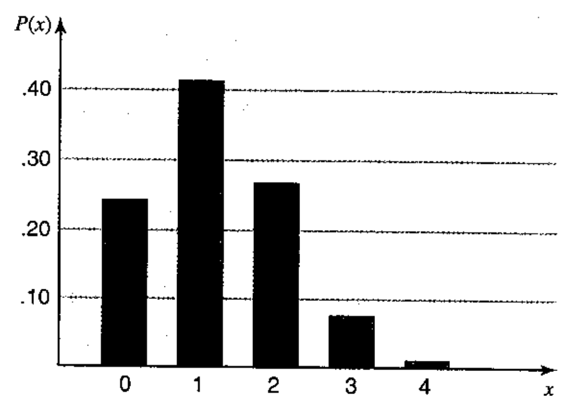


Figure 5.8 Bar graph for the probability distribution of Table 5.15.

3. Let $n = 4$ and $p = .80$ (which is greater than .50). Table 5.16, which is written by using Table IV of Appendix C, and the graph of the probability distribution in Figure 5.9 show that the probability distribution of x for $n = 4$ and $p = .80$ is skewed to the left.

Table 5.16 Probability Distribution of x for $n = 4$ and $p = .80$

x	$P(x)$
0	.0016
1	.0256
2	.1536
3	.4096
4	.4096

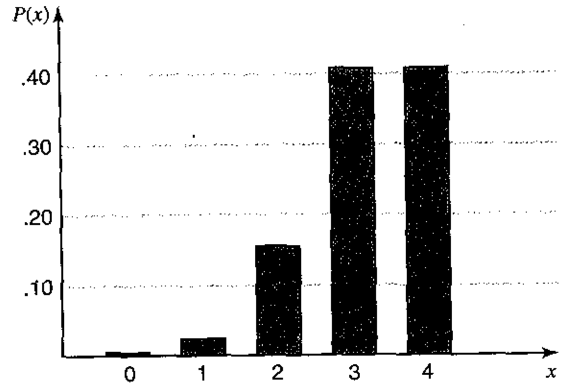


Figure 5.9 Bar graph for the probability distribution of Table 5.16.

5.6.5 Mean and Standard Deviation of the Binomial Distribution

Sections 5.3 and 5.4 explained how to compute the mean and standard deviation, respectively, for a probability distribution of a discrete random variable. When a discrete random variable has a binomial distribution, the formulas learned in Sections 5.3 and 5.4 could still be used to compute its mean and standard deviation. However, it is simpler and more convenient to use the following formulas to find the mean and standard deviation in such cases.

Mean and Standard Deviation of a Binomial Distribution The *mean* and *standard deviation of a binomial distribution* are

$$\mu = np \quad \text{and} \quad \sigma = \sqrt{npq}$$

where n is the total number of trials, p is the probability of success, and q is the probability of failure.

Example 5–22 describes the calculation of the mean and standard deviation of a binomial distribution.

EXAMPLE 5–22

Calculating the mean and standard deviation of a binomial random variable.

In a Maritz poll of adult drivers conducted in July 2002, 45% said that they "often" or "sometimes" eat or drink while driving (*USA TODAY*, October 23, 2002). Assume that this result is true for the current population of all adult drivers. A sample of 40 adult drivers is selected. Let x be the number of drivers in this sample who "often" or "sometimes" eat or drink while driving. Find the mean and standard deviation of the probability distribution of x.

Solution This is a binomial experiment with a total of 40 trials. Each trial has two outcomes: (1) the selected driver eats or drinks "often" or "sometimes" while driving, or (2) the selected driver never eats or drinks while driving. Assume that these are the only two possible outcomes for a driver. The probabilities p and q for these two outcomes are .45 and .55, respectively. Thus,

$$n = 40, \quad p = .45, \quad \text{and} \quad q = .55$$

Using the formulas for the mean and standard deviation of the binomial distribution, we obtain

$$\mu = np = 40(.45) = \mathbf{18}$$

$$\sigma = \sqrt{npq} = \sqrt{(40)(.45)(.55)} = \mathbf{3.146}$$

Thus, the mean of the probability distribution of x is 18 and the standard deviation is 3.146. The value of the mean is what we expect to obtain, on average, per repetition of the experiment. In this example, if we select many samples of 40 adult drivers, we expect that each sample will contain an average of 18 drivers, with a standard deviation of 3.146, who would say that they eat or drink "often" or "sometimes" while driving.

EXERCISES

CONCEPTS AND PROCEDURES

5.49 Briefly explain the following.
- **a.** A binomial experiment **b.** A trial **c.** A binomial random variable

5.50 What are the parameters of the binomial probability distribution and what do they mean?

5.51 Which of the following are binomial experiments? Explain why.
- **a.** Rolling a die many times and observing the number of spots
- **b.** Rolling a die many times and observing whether the number obtained is even or odd
- **c.** Selecting a few voters from a very large population of voters and observing whether or not each of them favors a certain proposition in an election when 54% of all voters are known to be in favor of this proposition.

5.52 Which of the following are binomial experiments? Explain why.
- **a.** Drawing 3 balls with replacement from a box that contains 10 balls, 6 of which are red and 4 blue, and observing the colors of the drawn balls
- **b.** Drawing 3 balls without replacement from a box that contains 10 balls, 6 of which are red and 4 blue, and observing the colors of the drawn balls
- **c.** Selecting a few households from New York City and observing whether or not they own stocks when it is known that 28% of all households in New York City own stocks

5.53 Let x be a discrete random variable that possesses a binomial distribution. Using the binomial formula, find the following probabilities.
- **a.** $P(x = 5)$ for $n = 8$ and $p = .70$
- **b.** $P(x = 3)$ for $n = 4$ and $p = .40$
- **c.** $P(x = 2)$ for $n = 6$ and $p = .30$

Verify your answers by using Table IV of Appendix C.

5.54 Let x be a discrete random variable that possesses a binomial distribution. Using the binomial formula, find the following probabilities.
- **a.** $P(x = 0)$ for $n = 5$ and $p = .05$
- **b.** $P(x = 4)$ for $n = 7$ and $p = .90$
- **c.** $P(x = 7)$ for $n = 10$ and $p = .60$

Verify your answers by using Table IV of Appendix C.

5.72 Let $N = 14$, $r = 6$, and $n = 5$. Using the hypergeometric probability distribution formula, find
a. $P(x = 4)$ **b.** $P(x = 5)$ **c.** $P(x \leq 1)$

5.73 Let $N = 11$, $r = 4$, and $n = 4$. Using the hypergeometric probability distribution formula, find
a. $P(x = 2)$ **b.** $P(x = 4)$ **c.** $P(x \leq 1)$

5.74 Let $N = 16$, $r = 10$, and $n = 5$. Using the hypergeometric probability distribution formula, find
a. $P(x = 5)$ **b.** $P(x = 0)$ **c.** $P(x \leq 1)$

APPLICATIONS

5.75 An Internal Revenue Service inspector is to select 3 corporations from a list of 15 for tax audit purposes. Of the 15 corporations, 6 earned profits and 9 incurred losses during the year for which the tax returns are to be audited. If the IRS inspector decides to select three corporations randomly, find the probability that the number of corporations in these three that incurred losses during the year for which the tax returns are to be audited is
a. exactly 2 **b.** none **c.** at most 1

5.76 Six jurors are to be selected from a pool of 20 potential candidates to hear a civil case involving a lawsuit between two families. Unknown to the judge or any of the attorneys, 4 of the 20 prospective jurors are potentially prejudiced by being acquainted with one or more of the litigants, but they will not disclose this during the jury selection process. If 6 jurors are selected at random from this group of 20, find the probability that the number of potentially prejudiced jurors among the 6 selected jurors is
a. exactly 1 **b.** none **c.** at most 2

5.77 A shop has 11 video games to choose from, and 4 of them contain extreme violence. A customer picks 3 of these 11 games at random. What is the probability that the number of extremely violent games among the three selected games is
a. exactly two **b.** more than one **c.** none

5.78 Bender Electronics buys keyboards for its computers from another company. The keyboards are received in shipments of 100 boxes, each box containing 20 keyboards. The quality control department at Bender Electronics first randomly selects one box from each shipment and then randomly selects five keyboards from that box. The shipment is accepted if not more than one of the five keyboards is defective. The quality control inspector at Bender Electronics selected a box from a recently received shipment of keyboards. Unknown to the inspector, this box contains six defective keyboards.
a. What is the probability that this shipment will be accepted?
b. What is the probability that this shipment will not be accepted?

THE POISSON PROBABILITY DISTRIBUTION

The **Poisson probability distribution**, named after the French mathematician Simeon D. Poisson, is another important probability distribution of a discrete random variable that has a large number of applications. Suppose a washing machine in a laundromat breaks down an average of three times a month. We may want to find the probability of exactly two breakdowns during the next month. This is an example of a Poisson probability distribution problem. Each breakdown is called an *occurrence* in Poisson probability distribution terminology. The Poisson probability distribution is applied to experiments with random and independent occurrences. The occurrences are random in the sense that they do not follow any pattern, and, hence, they are unpredictable. Independence of occurrences means that one occurrence (or nonoccurrence) of an event does not influence the successive occurrences or nonoccurrences of that event. The occurrences are always considered with respect to an interval. In the example of the washing machine, the interval is one month. The interval may be a time interval, a space interval, or a volume interval. The actual number of occurrences within an interval is random and independent. If the average number of occurrences for a given interval is known, then by using the Poisson probability distribution, we can compute the probability of a certain number of occurrences, x, in that interval. Note that the number of actual occurrences in an interval is denoted by x.

Conditions to Apply the Poisson Probability Distribution The following three conditions must be satisfied to apply the Poisson probability distribution.

1. x is a discrete random variable.
2. The occurrences are random.
3. The occurrences are independent.

The following are three examples of discrete random variables for which the occurrences are random and independent. Hence, these are examples to which the Poisson probability distribution can be applied.

1. Consider the number of telemarketing phone calls received by a household during a given day. In this example, the receiving of a telemarketing phone call by a household is called an occurrence, the interval is one day (an interval of time), and the occurrences are random (that is, there is no specified time for such a phone call to come in). The total number of telemarketing phone calls received by a household during a given day may be 0, 1, 2, 3, 4, ... The independence of occurrences in this example means that the telemarketing phone calls are received individually and none of two (or more) of these phone calls are related.
2. Consider the number of defective items in the next 100 items manufactured on a machine. In this case, the interval is a volume interval (100 items). The occurrences (number of defective items) are random because there may be 0, 1, 2, 3 . . . 100 defective items in 100 items. We can assume the occurrence of defective items to be independent of one another.
3. Consider the number of defects in a five-foot-long iron rod. The interval, in this example, is a space interval (five feet). The occurrences (defects) are random because there may be any number of defects in a five-foot iron rod. We can assume that these defects are independent of one another.

The following examples also qualify for the application of the Poisson probability distribution.

1. The number of accidents that occur on a given highway during a one-week period
2. The number of customers entering a grocery store during a one-hour interval
3. The number of television sets sold at a department store during a given week

In contrast, consider the arrival of patients at a physician's office. These arrivals are nonrandom if the patients have to make appointments to see the doctor. The arrival of commercial airplanes at an airport is nonrandom because all planes are scheduled to arrive at certain times, and airport authorities know the exact number of arrivals for any period (although this number may change slightly because of late or early arrivals and cancellations). The Poisson probability distribution cannot be applied to these examples.

In the Poisson probability distribution terminology, the average number of occurrences in an interval is denoted by λ (Greek letter *lambda*). The actual number of occurrences in that interval is denoted by x. Then, using the Poisson probability distribution, we find the probability of x occurrences during an interval given that the mean occurrences during that interval are λ.

Poisson Probability Distribution Formula According to the *Poisson probability distribution*, the probability of x occurrences in an interval is

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where λ (pronounced *lambda*) is the mean number of occurrences in that interval and the value of e is approximately 2.71828.

The mean number of occurrences in an interval, denoted by λ, is called the *parameter of the Poisson probability distribution* or the **Poisson parameter**. As is obvious from the Poisson probability distribution formula, we need to know only the value of λ to compute the probability of any given value of x. We can read the value of $e^{-\lambda}$ for a given λ from Table V of Appendix C. Examples 5–25 through 5–27 illustrate the use of the Poisson probability distribution formula.

EXAMPLE 5–25

Using the Poisson formula: x equals a specific value.

On average, a household receives 9.5 telemarketing phone calls per week. Using the Poisson distribution formula, find the probability that a randomly selected household receives exactly six telemarketing phone calls during a given week.

Solution Let λ be the mean number of telemarketing phone calls received by a household per week. Then, $\lambda = 9.5$. Let x be the number of telemarketing phone calls received by a household during a given week. We are to find the probability of $x = 6$. Substituting all the values in the Poisson formula, we obtain

$$P(x = 6) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(9.5)^6 e^{-9.5}}{6!} = \frac{(735{,}091.8906)(.00007485)}{720} = \mathbf{.0764}$$

To do these calculations, we can find the value of 6! from Table II of Appendix C, and we can find the value of $e^{-9.5}$ by using the e^x key on a calculator or from Table V.

EXAMPLE 5–26

Calculating probabilities using the Poisson formula.

A washing machine in a laundromat breaks down an average of three times per month. Using the Poisson probability distribution formula, find the probability that during the next month this machine will have

(a) exactly two breakdowns **(b)** at most one breakdown

Solution Let λ be the mean number of breakdowns per month, and let x be the actual number of breakdowns observed during the next month for this machine. Then,

$$\lambda = 3$$

(a) The probability that exactly two breakdowns will be observed during the next month is

$$P(x = 2) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(3)^2 e^{-3}}{2!} = \frac{(9)(.04978707)}{2} = \mathbf{.2240}$$

(b) The probability that at most one breakdown will be observed during the next month is given by the sum of the probabilities of zero and one breakdown. Thus,

$$P(\text{at most 1 breakdown}) = P(0 \text{ or } 1 \text{ breakdown}) = P(x = 0) + P(x = 1)$$
$$= \frac{(3)^0 e^{-3}}{0!} + \frac{(3)^1 e^{-3}}{1!}$$
$$= \frac{(1)(.04978707)}{1} + \frac{(3)(.04978707)}{1}$$
$$= .0498 + .1494 = \mathbf{.1992}$$

Remember ► One important point about the Poisson probability distribution is that *the intervals for λ and x must be equal.* If they are not, the mean λ should be redefined to make them equal. Example 5–27 illustrates this point.

EXAMPLE 5–27

Calculating a probability using the Poisson formula.

Cynthia's Mail Order Company provides free examination of its products for seven days. If not completely satisfied, a customer can return the product within that period and get a full refund. According to past records of the company, an average of 2 of every 10 products sold by this company are returned for a refund. Using the Poisson probability distribution formula, find the probability that exactly 6 of the 40 products sold by this company on a given day will be returned for a refund.

Solution Let x denote the number of products in 40 that will be returned for a refund. We are to find $P(x = 6)$. The given mean is defined per 10 products, but x is defined for 40 products. As a result, we should first find the mean for 40 products. Because, on average, 2 out of 10 products are returned, the mean number of products returned out of 40 will be 8. Thus, $\lambda = 8$. Substituting $x = 6$ and $\lambda = 8$ in the Poisson probability distribution formula, we obtain

$$P(x = 6) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{(8)^6 e^{-8}}{6!} = \frac{(262{,}144)(.00033546)}{720} = \mathbf{.1221}$$

Thus, the probability is .1221 that exactly 6 products out of 40 sold on a given day will be returned.

Note that Example 5–27 is actually a binomial problem with $p = 2/10 = .20$, $n = 40$, and $x = 6$. In other words, the probability of success (that is, the probability that a product is returned) is .20 and the number of trials (products sold) is 40. We are to find the probability of six successes (returns). However, we used the Poisson distribution to solve this problem. This is referred to as *using the Poisson distribution as an approximation to the binomial distribution.* We can also use the binomial distribution to find this probability as follows:

$$P(x = 6) = {}_{40}C_6\,(.20)^6\,(.80)^{34} = \frac{40!}{6!(40-6)!}(.20)^6\,(.80)^{34}$$
$$= (3{,}838{,}380)(.000064)(.00050706) = \mathbf{.1246}$$

5–3 ASK MR. STATISTICS

CASE STUDY

Fortune magazine used to publish a column titled *Ask Mr. Statistics,* which contained questions and answers to statistical problems. The following excerpts are reprinted from one such column.

Dear Oddgiver: I am in the seafood distribution business and find myself endlessly wrangling with supermarkets about appropriate order sizes, especially with high-end tidbit products like our matjes herring in superspiced wine, which we let them have for $4.25, and still they take only a half-dozen jars, thereby running the risk of getting sold out early in the week and causing the better class of customers to storm out empty-handed. How do I get them to realize that lowballing on inventories is usually bad business, also to at least try a few jars of our pickled crappie balls?

—HEADED FOR A BREAKDOWN

Dear Picklehead: The science of statistics has much to offer people puzzled by seafood inventory problems. Your salvation lies in the Poisson distribution, "poisson" being French for fish and, of arguably greater relevance, the surname of a 19th-century French probabilist.

Simeon Poisson's contribution was to develop a method for calculating the likelihood that a specified number of successes will occur given that (a) the probability of success on any one trial is very low but (b) the number of trials is very high. A real world example often mentioned in the literature concerns the distribution of Prussian cavalry deaths from getting kicked by horses in the period 1875–94.

As you would expect of Teutons, the Prussian military kept meticulous records on horse-kick deaths in each of its army corps, and the data are neatly summarized in a 1963 book called *Lady Luck,* by the late Warren Weaver. There were a total of 196 kicking deaths—these being the, er, "successes." The "trials" were each army corps' observations on the number of kicking deaths sustained in the year. So with 14 army corps and data for 20 years, there were 280 trials. We shall not detain you with the Poisson formula, but it predicts, for example, that there will be 34.1 instances of a corps' having exactly two deaths in a year. In fact, there were 32 such cases. Pretty good, eh?

Back to seafood. The Poisson calculation is appropriate to your case, since the likelihood of any one customer's buying your overspiced herring is extremely small, but the number of trials—i.e., customers in the store during a typical week—is very large. Let us say that one customer in 1,000 deigns to buy the herring, and 6,000 customers visit the store in a week. So six jars are sold in an average week.

But the store manager doesn't care about average weeks. What he's worried about is having too much or not enough. He needs to know the probabilities assigned to different sales levels. Our Poisson distribution shows the following morning line: The chance of fewer than three sales—only 6.2%. Of four to six sales: 45.5%. Chances of losing some sales if the store elects to start the week with six jars because that happens to be the average: 39.4%. If the store wants to be 90% sure of not losing sales, it needs to start with nine jars.

There is no known solution to the problem of pickled crappie balls.

Using the Poisson probability distribution, calculate the probabilities mentioned at the end of this case study.

Source: Daniel Seligman, "Ask Mr. Statistics," *Fortune*, March 7, 1994.

Thus the probability $P(x = 6)$ is .1246 when we use the binomial distribution.

As we can observe, simplifying the above calculations for the binomial formula is quite complicated when n is large. It is much easier to solve this problem using the Poisson distribution. As a general rule, if it is a binomial problem with $n > 25$ but $\mu \leq 25$, then we can use the Poisson distribution as an approximation to the binomial distribution. However, if $n > 25$ and $\mu > 25$, we prefer to use the normal distribution as an approximation to the binomial. The latter case will be discussed in Chapter 6.

Case Study 5–3 presents applications of the binomial and Poisson probability distributions.

5.8.1 Using the Table of Poisson Probabilities

The probabilities for a Poisson distribution can also be read from Table VI, the table of Poisson probabilities, in Appendix C. The following example describes how to read that table.

EXAMPLE 5–28

Using the table of Poisson probabilities.

On average, two new accounts are opened per day at an Imperial Savings Bank branch. Using Table VI of Appendix C, find the probability that on a given day the number of new accounts opened at this bank will be

(a) exactly 6 (b) at most 3 (c) at least 7

Solution Let

λ = mean number of new accounts opened per day at this bank

x = number of new accounts opened at this bank on a given day

(a) The values of λ and x are

$$\lambda = 2 \quad \text{and} \quad x = 6$$

In Table VI of Appendix C, we first locate the column that corresponds to $\lambda = 2$. In this column, we then read the value for $x = 6$. The relevant portion of that table is shown here as Table 5.17. The probability that exactly 6 new accounts will be opened on a given day is .0120. Therefore,

$$P(x = 6) = \mathbf{.0120}$$

Table 5.17 Portion of Table VI for λ = 2.0

		λ		
x	1.1	1.2	...	2.0 ← λ = 2.0
0				.1353
1				.2707
2				.2707
3				.1804
4				.0902
5				.0361
x = 6 → 6				.0120 ← $P(x = 6)$
7				.0034
8				.0009
9				.0002

Actually, Table 5.17 gives the probability distribution of x for $\lambda = 2.0$. Note that the sum of the 10 probabilities given in Table 5.17 is .9999 and not 1.0. This is so for two reasons. First, these probabilities are rounded to four decimal places. Second, on a given day more than nine new accounts might be opened at this bank. However, the probabilities of 10, 11, 12 . . . new accounts are very small and they are not listed in the table.

(b) The probability that at most three new accounts are opened on a given day is obtained by adding the probabilities of 0, 1, 2, and 3 new accounts. Thus, using Table VI of Appendix C or Table 5.17, we obtain

$$P(\text{at most } 3) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3)$$
$$= .1353 + .2707 + .2707 + .1804 = \mathbf{.8571}$$

(c) The probability that at least 7 new accounts are opened on a given day is obtained by adding the probabilities of 7, 8, and 9 new accounts. Note that 9 is the last value of x for $\lambda = 2.0$ in Table VI of Appendix C or Table 5.17. Hence, 9 is the last value of x whose probability is included in the sum. However, this does not mean that on a given day more than nine new accounts cannot be opened. It simply means that the probability of 10 or more accounts is close to zero. Thus,

$$P(\text{at least } 7) = P(x = 7) + P(x = 8) + P(x = 9)$$
$$= .0034 + .0009 + .0002 = \mathbf{.0045}$$

EXAMPLE 5–29

Constructing a Poisson probability distribution and graphing it.

An auto salesperson sells an average of .9 car per day. Let x be the number of cars sold by this salesperson on any given day. Using the Poisson probability distribution table, write the probability distribution of x. Draw a graph of the probability distribution.

Solution Let λ be the mean number of cars sold per day by this salesperson. Hence, $\lambda = .9$. Using the portion of Table VI corresponding to $\lambda = .9$, we write the probability distribution of x in Table 5.18. Figure 5.10 shows the bar graph for the probability distribution of Table 5.18.

Table 5.18 Probability Distribution of x for $\lambda = .9$

x	$P(x)$
0	.4066
1	.3659
2	.1647
3	.0494
4	.0111
5	.0020
6	.0003

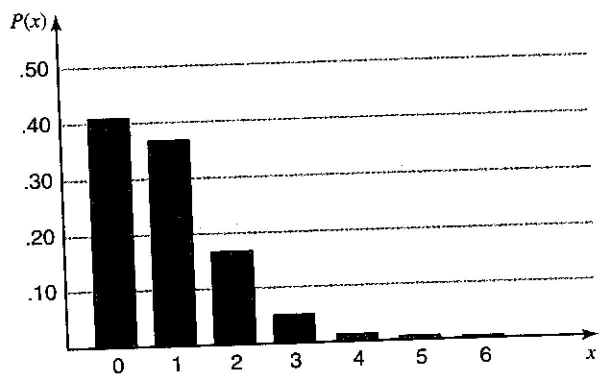


Figure 5.10 Bar graph for the probability distribution of Table 5.18.

Note that 6 is the largest value of x for $\lambda = .9$ listed in Table VI for which the probability is greater than zero. However, this does not mean that this salesperson cannot sell more than six cars on a given day. What this means is that the probability of selling seven or more cars is very small. Actually, the probability of $x = 7$ for $\lambda = .9$ calculated by using the Poisson formula is .000039. When rounded to four decimal places, this probability is .0000, as listed in Table VI.

5.8.2 Mean and Standard Deviation of the Poisson Probability Distribution

For the Poisson probability distribution, the mean and variance both are equal to λ, and the standard deviation is equal to $\sqrt{\lambda}$. That is, for the Poisson probability distribution,

$$\mu = \lambda$$
$$\sigma^2 = \lambda$$
$$\sigma = \sqrt{\lambda}$$

For Example 5–29, $\lambda = .9$. Therefore, for the probability distribution of x in Table 5.18, the mean, variance, and standard deviation are

$$\mu = \lambda = .9 \text{ car}$$
$$\sigma^2 = \lambda = .9$$
$$\sigma = \sqrt{\lambda} = \sqrt{.9} = .949 \text{ car}$$

EXERCISES

CONCEPTS AND PROCEDURES

5.79 What are the conditions that must be satisfied to apply the Poisson probability distribution?

5.80 What is the parameter of the Poisson probability distribution, and what does it mean?

5.81 Using the Poisson formula, find the following probabilities.
a. $P(x \leq 1)$ for $\lambda = 5$ **b.** $P(x = 2)$ for $\lambda = 2.5$

Verify these probabilities using Table VI of Appendix C.

5.82 Using the Poisson formula, find the following probabilities.
a. $P(x < 2)$ for $\lambda = 3$ **b.** $P(x = 8)$ for $\lambda = 5.5$

Verify these probabilities using Table VI of Appendix C.

5.83 Let x be a Poisson random variable. Using the Poisson probabilities table, write the probability distribution of x for each of the following. Find the mean, variance, and standard deviation for each of these probability distributions. Draw a graph for each of these probability distributions.
a. $\lambda = 1.3$ **b.** $\lambda = 2.1$

5.84 Let x be a Poisson random variable. Using the Poisson probabilities table, write the probability distribution of x for each of the following. Find the mean, variance, and standard deviation for each of these probability distributions. Draw a graph for each of these probability distributions.
a. $\lambda = .6$ **b.** $\lambda = 1.8$

APPLICATIONS

5.85 A household receives an average of 1.7 pieces of junk mail per day. Find the probability that this household will receive exactly three pieces of junk mail on a certain day. Use the Poisson probability distribution formula.

5.86 A commuter airline receives an average of 9.7 complaints per day from its passengers. Using the Poisson formula, find the probability that on a certain day this airline will receive exactly six complaints.

5.89 A university police department receives an average of 3.7 reports per week of lost student ID cards.

a. Find the probability that at most one such report will be received during a given week by this police department. Use the Poisson probability distribution formula.

b. Using the Poisson probabilities table, find the probability that during a given week the number of such reports received by this police department is

i. 1 to 4 **ii.** at least 6 **iii.** at most 3

5.90 A large proportion of small businesses in the United States fail during the first few years of operation. On average, 1.6 businesses file for bankruptcy per day in a large city.

a. Using the Poisson formula, find the probability that exactly three businesses will file for bankruptcy on a given day in this city.

b. Using the Poisson probabilities table, find the probability that the number of businesses that will file for bankruptcy on a given day in this city is

i. 2 to 3 **ii.** more than 3 **iii.** less than 3

5.91 Despite all efforts by the quality control department, the fabric made at Benton Corporation always contains a few defects. A certain type of fabric made at this corporation contains an average of .5 defect per 500 yards.

a. Using the Poisson formula, find the probability that a given piece of 500 yards of this fabric will contain exactly one defect.

b. Using the Poisson probabilities table, find the probability that the number of defects in a given 500-yard piece of this fabric will be

i. 2 to 4 **ii.** more than 3 **iii.** less than 2

5.92 A large self-service gas station experiences an average of 1.6 "drive-offs" (a customer drives away without paying) per week.

a. Using the Poisson probability distribution formula, find the probability that exactly two drive-offs will occur during a given week.

b. Using the Poisson probabilities table, find the probability that the number of drive-offs experienced by this gas station during a given week will be

i. less than 3 **ii.** more than 5 **iii.** 2 to 5

5.93 An average of 4.8 customers come to Columbia Savings and Loan every half hour.

a. Find the probability that exactly two customers will come to this savings and loan during a given hour.

b. Find the probability that during a given hour, the number of customers who will come to this savings and loan is

i. 2 or fewer **ii.** 10 or more

5.94 A certain newspaper contains an average of 1.1 typographical errors per page.

a. Using the Poisson formula, find the probability that a randomly selected page of this newspaper will contain exactly four typographical errors.

b. Using the Poisson probabilities table, find the probability that the number of typographical errors on a randomly selected page will be

i. more than 3 **ii.** less than 4

5.95 An insurance salesperson sells an average of 1.4 policies per day.

a. Using the Poisson formula, find the probability that this salesperson will sell no insurance policy on a certain day.

b. Let x denote the number of insurance policies that this salesperson will sell on a given day. Using the Poisson probabilities table, write the probability distribution of x.

c. Find the mean, variance, and standard deviation of the probability distribution developed in part b.

5.96 An average of .8 accident occurs per day in a large city.

a. Find the probability that no accident will occur in this city on a given day.

b. Let x denote the number of accidents that will occur in this city on a given day. Write the probability distribution of x.

c. Find the mean, variance, and standard deviation of the probability distribution developed in part b.