Use the Section EG5.1 instructions to compute the variance and standard deviation of a discrete variable.

STANDARD DEVIATION OF A DISCRETE VARIABLE

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum_{i=1}^{N}[x_i - E(X)]^2 P(X = x_i)} \quad \textbf{(5.3)}$$

The variance and the standard deviation of the number of interruptions per day are computed as follows and in Table 5.3, using Equations (5.2) and (5.3):

$$\begin{aligned}
\sigma^2 &= \sum_{i=1}^{N}[x_i - E(X)]^2 P(X = x_i)\\
&= (0 - 1.4)^2(0.35) + (1 - 1.4)^2(0.25) + (2 - 1.4)^2(0.20) + (3 - 1.4)^2(0.10)\\
&\quad + (4 - 1.4)^2(0.05) + (5 - 1.4)^2(0.05)\\
&= 0.686 + 0.040 + 0.072 + 0.256 + 0.338 + 0.648\\
&= 2.04
\end{aligned}$$

and

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.04} = 1.4283$$

TABLE 5.3

Computing the Variance and Standard Deviation of the Number of Interruptions per Day

Interruptions per Day (x_i)	$P(X = x_i)$	$x_iP(X = x_i)$	$[x_i\text{-}E(X)]^2$	$[x_i - E(X)]^2P(X = x_i)$
0	0.35	0.00	$(0 - 1.4)^2 = 1.96$	$(1.96)(0.35) = 0.686$
1	0.25	0.25	$(1 - 1.4)^2 = 0.16$	$(0.16)(0.25) = 0.040$
2	0.20	0.40	$(2 - 1.4)^2 = 0.36$	$(0.36)(0.20) = 0.072$
3	0.10	0.30	$(3 - 1.4)^2 = 2.56$	$(2.56)(0.10) = 0.256$
4	0.05	0.20	$(4 - 1.4)^2 = 6.76$	$(6.76)(0.05) = 0.338$
5	0.05	0.25	$(5 - 1.4)^2 = 12.96$	$(12.96)(0.05) = 0.648$
	1.00	$\mu = E(X) = 1.40$		$\sigma^2 = 2.04$
				$\sigma = \sqrt{\sigma^2} = 1.4283$

Thus, the mean number of interruptions per day is 1.4, the variance is 2.04, and the standard deviation is approximately 1.43 interruptions per day.

Problems for Section 5.1

LEARNING THE BASICS

5.1 Given the following probability distributions:

Distribution *A*		Distribution *B*	
x_i	$P(X = x_i)$	x_i	$P(X = x_i)$
0	0.50	0	0.05
1	0.20	1	0.10
2	0.15	2	0.15
3	0.10	3	0.20
4	0.05	4	0.50

a. Compute the expected value for each distribution.
b. Compute the standard deviation for each distribution.
c. Compare the results of distributions *A* and *B*.

APPLYING THE CONCEPTS

SELF Test **5.2** The following table contains the probability distribution for the number of traffic accidents daily in a small town:

Number of Accidents Daily (X)	$P(X = x_i)$
0	0.10
1	0.20
2	0.45
3	0.15
4	0.05
5	0.05

a. Compute the mean number of accidents per day.
b. Compute the standard deviation.

5.3 Recently, a regional automobile dealership sent out fliers to perspective customers indicating that they had already won one of three different prizes: an automobile valued at $25,000, a $100 gas card, or a $5 Walmart shopping card. To claim his or her prize, a prospective customer needed to present the flier at the dealership's showroom. The fine print on the back of the flier listed the probabilities of winning. The chance of winning the car was 1 out of 31,478, the chance of winning the gas card was 1 out of 31,478, and the chance of winning the shopping card was 31,476 out of 31,478.
a. How many fliers do you think the automobile dealership sent out?
b. Using your answer to (a) and the probabilities listed on the flier, what is the expected value of the prize won by a prospective customer receiving a flier?
c. Using your answer to (a) and the probabilities listed on the flier, what is the standard deviation of the value of the prize won by a prospective customer receiving a flier?
d. Do you think this is an effective promotion? Why or why not?

5.4 In the carnival game Under-or-Over-Seven, a pair of fair dice is rolled once, and the resulting sum determines whether the player wins or loses his or her bet. For example, the player can bet $1 that the sum will be under 7—that is, 2, 3, 4, 5, or 6. For this bet, the player wins $1 if the result is under 7 and loses $1 if the outcome equals or is greater than 7. Similarly, the player can bet $1 that the sum will be over 7—that is, 8, 9, 10, 11, or 12. Here, the player wins $1 if the result is over 7 but loses $1 if the result is 7 or under. A third method of play is to bet $1 on the outcome 7. For this bet, the player wins $4 if the result of the roll is 7 and loses $1 otherwise.
a. Construct the probability distribution representing the different outcomes that are possible for a $1 bet on under 7.
b. Construct the probability distribution representing the different outcomes that are possible for a $1 bet on over 7.
c. Construct the probability distribution representing the different outcomes that are possible for a $1 bet on 7.
d. Show that the expected long-run profit (or loss) to the player is the same, no matter which method of play is used.

5.5 The number of arrivals per minute at a bank located in the central business district of a large city was recorded over a period of 200 minutes, with the following results:

Arrivals	Frequency
0	14
1	31
2	47
3	41
4	29
5	21
6	10
7	5
8	2

a. Compute the expected number of arrivals per minute.
b. Compute the standard deviation.

5.6 The manager of the commercial mortgage department of a large bank has collected data during the past two years concerning the number of commercial mortgages approved per week. The results from these two years (104 weeks) are as follows:

Number of Commercial Mortgages Approved	Frequency
0	13
1	25
2	32
3	17
4	9
5	6
6	1
7	1

a. Compute the expected number of mortgages approved per week.
b. Compute the standard deviation.

5.2 Covariance of a Probability Distribution and Its Application in Finance

Section 5.1 defined the expected value, variance, and standard deviation for a single discrete variable. In this section, the covariance between two variables is introduced and applied to portfolio management, a topic of great interest to financial analysts.

Problems for Section 5.2

LEARNING THE BASICS

5.7 Given the following probability distributions for variables X and Y:

$P(x, y)$	X	Y
0.4	100	200
0.6	200	100

Compute

a. $E(X)$ and $E(Y)$.
b. σ_X and σ_Y.
c. σ_{XY}.
d. $E(X + Y)$.

5.8 Given the following probability distributions for variables X and Y:

$P(x, y)$	X	Y
0.2	−100	50
0.4	50	30
0.3	200	20
0.1	300	20

Compute

a. $E(X)$ and $E(Y)$.
b. σ_X and σ_Y.
c. σ_{XY}.
d. $E(X + Y)$.

5.9 Two investments, X and Y, have the following characteristics:

$$E(X) = \$50, E(Y) = \$100, \sigma_X^2 = 9{,}000,$$
$$\sigma_Y^2 = 15{,}000, \text{ and } \sigma_{XY} = 7{,}500.$$

If the weight of portfolio assets assigned to investment X is 0.4, compute the

a. portfolio expected return.
b. portfolio risk.

APPLYING THE CONCEPTS

5.10 The process of being served at a bank consists of two independent parts—the time waiting in line and the time it takes to be served by the teller. Suppose that the time waiting in line has an expected value of 4 minutes, with a standard deviation of 1.2 minutes, and the time it takes to be served by the teller has an expected value of 5.5 minutes, with a standard deviation of 1.5 minutes. Compute the

a. expected value of the total time it takes to be served at the bank.
b. standard deviation of the total time it takes to be served at the bank.

5.11 In the portfolio example in this section (see page 192), half the portfolio assets are invested in the Dow Jones fund and half in a weak-economy fund. Recalculate the portfolio expected return and the portfolio risk if

a. 30% of the portfolio assets are invested in the Dow Jones fund and 70% in a weak-economy fund.
b. 70% of the portfolio assets are invested in the Dow Jones fund and 30% in a weak-economy fund.
c. Which of the three investment strategies (30%, 50%, or 70% in the Dow Jones fund) would you recommend? Why?

SELF Test **5.12** You are trying to develop a strategy for investing in two different stocks. The anticipated annual return for a $1,000 investment in each stock under four different economic conditions has the following probability distribution:

		Returns	
Probability	**Economic Condition**	**Stock X**	**Stock Y**
0.1	Recession	−100	50
0.3	Slow growth	0	150
0.3	Moderate growth	80	−20
0.3	Fast growth	150	−100

Compute the

a. expected return for stock X and for stock Y.
b. standard deviation for stock X and for stock Y.
c. covariance of stock X and stock Y.
d. Would you invest in stock X or stock Y? Explain.

5.13 Suppose that in Problem 5.12 you wanted to create a portfolio that consists of stock X and stock Y. Compute the portfolio expected return and portfolio risk for each of the following percentages invested in stock X:

a. 30%
b. 50%
c. 70%
d. On the basis of the results of (a) through (c), which portfolio would you recommend? Explain.

5.14 You are trying to develop a strategy for investing in two different stocks. The anticipated annual return for a $1,000 investment in each stock under four different economic conditions has the following probability distribution:

		Returns	
Probability	**Economic Condition**	**Stock X**	**Stock Y**
0.1	Recession	−50	−100
0.3	Slow growth	20	50
0.4	Moderate growth	100	130
0.2	Fast growth	150	200

Compute the

a. expected return for stock X and for stock Y.
b. standard deviation for stock X and for stock Y.
c. covariance of stock X and stock Y.
d. Would you invest in stock X or stock Y? Explain.

5.15 Suppose that in Example 5.1 on page 193, you wanted to create a portfolio that consists of the Black Swan fund and the Good Times fund. Compute the portfolio expected return and portfolio risk for each of the following percentages invested in the Black Swan fund:

a. 30%
b. 50%
c. 70%
d. On the basis of the results of (a) through (c), which portfolio would you recommend? Explain.

5.16 You plan to invest $1,000 in a corporate bond fund or in a common stock fund. The following table presents the annual return (per $1,000) of each of these investments under various economic conditions and the probability that each of those economic conditions will occur. Compute the

Probability	Economic Condition	Corporate Bond Fund	Common Stock Fund
0.01	Extreme recession	−200	−999
0.09	Recession	−70	−300
0.15	Stagnation	30	−100
0.35	Slow growth	80	100
0.30	Moderate growth	100	150
0.10	High growth	120	350

a. expected return for the corporate bond fund and for the common stock fund.
b. standard deviation for the corporate bond fund and for the common stock fund.
c. covariance of the corporate bond fund and the common stock fund.
d. Would you invest in the corporate bond fund or the common stock fund? Explain.
e. If you chose to invest in the common stock fund in (d), what do you think about the possibility of losing $999 of every $1,000 invested if there is an extreme recession?

5.17 Suppose that in Problem 5.16 you wanted to create a portfolio that consists of the corporate bond fund and the common stock fund. Compute the portfolio expected return and portfolio risk for each of the following situations:

a. $300 in the corporate bond fund and $700 in the common stock fund.
b. $500 in each fund.
c. $700 in the corporate bond fund and $300 in the common stock fund.
d. On the basis of the results of (a) through (c), which portfolio would you recommend? Explain.

5.3 Binomial Distribution

This is the first of three sections that considers mathematical models. A **mathematical model** is a mathematical expression that represents a variable of interest. When a mathematical model exists, you can compute the exact probability of occurrence of any particular value of the variable. For discrete random variables, the mathematical model is a **probability distribution function**.

The **binomial distribution** is an important mathematical model used in many business situations. You use the binomial distribution when the discrete variable is the number of events of interest in a sample of n observations. The binomial distribution has four important properties.

Student Tip

Do not confuse this use of the Greek letter pi, π, to represent the probability of an event of interest with the mathematical constant that is the ratio of the circumference to a diameter of a circle—approximately 3.14159—which is also known by the same Greek letter.

PROPERTIES OF THE BINOMIAL DISTRIBUTION

- The sample consists of a fixed number of observations, n.
- Each observation is classified into one of two mutually exclusive and collectively exhaustive categories.
- The probability of an observation being classified as the event of interest, π, is constant from observation to observation. Thus, the probability of an observation being classified as not being the event of interest, $1 - \pi$, is constant over all observations.
- The value of any observation is independent of the value of any other observation.

Returning to the Ricknel Home Improvement scenario presented on page 185 concerning the accounting information system, suppose the event of interest is defined as a tagged order form. You want to determine the number of tagged order forms in a given sample of orders.

What results can occur? If the sample contains four orders, there could be none, one, two, three, or four tagged order forms. No other value can occur because the number of tagged

$$\begin{aligned}
P(X = 0 \mid n = 3, \pi = 0.889) &= \frac{3!}{0!(3-0)!}(0.889)^0(1-0.889)^{3-0} \\
&= \frac{3!}{0!(3-0)!}(0.889)^0(0.111)^3 \\
&= 1(1)(0.111)(0.111)(0.111) = 0.0014 \\
P(X = 2 \mid n = 3, \pi = 0.889) &= \frac{3!}{2!(3-2)!}(0.889)^2(1-0.889)^{3-2} \\
&= \frac{3!}{2!(3-2)!}(0.889)^2(0.111)^1 \\
&= 3(0.889)(0.889)(0.111) = 0.2632
\end{aligned}$$

$$\begin{aligned}
P(X \geq 2) &= P(X = 2) + P(X = 3) \\
&= 0.2632 + 0.7026 \\
&= 0.9658
\end{aligned}$$

Using Equations (5.12) and (5.13),

$$\mu = E(X) = n\pi = 3(0.889) = 2.667$$

$$\begin{aligned}
\sigma = \sqrt{\sigma^2} &= \sqrt{Var(X)} = \sqrt{n\pi(1-\pi)} \\
&= \sqrt{3(0.889)(0.111)} \\
&= \sqrt{0.2960} = 0.5441
\end{aligned}$$

The mean number of orders filled correctly in a sample of three orders is 2.667, and the standard deviation is 0.5441. The probability that all three orders are filled correctly is 0.7026, or 70.26%. The probability that none of the orders are filled correctly is 0.0014, or 0.14%. The probability that at least two orders are filled correctly is 0.9658, or 96.58%.

Problems for Section 5.3

LEARNING THE BASICS

5.18 Determine the following:

a. For $n = 4$ and $\pi = 0.12$, what is $P(X = 0)$?
b. For $n = 10$ and $\pi = 0.40$, what is $P(X = 9)$?
c. For $n = 10$ and $\pi = 0.50$, what is $P(X = 8)$?
d. For $n = 6$ and $\pi = 0.83$, what is $P(X = 5)$?

5.19 If $n = 5$ and $\pi = 0.40$, what is the probability that

a. $X = 4$?
b. $X \leq 3$?
c. $X < 2$?
d. $X > 1$?

5.20 Determine the mean and standard deviation of the variable X in each of the following binomial distributions:

a. $n = 4$ and $\pi = 0.10$
b. $n = 4$ and $\pi = 0.40$
c. $n = 5$ and $\pi = 0.80$
d. $n = 3$ and $\pi = 0.50$

APPLYING THE CONCEPTS

5.21 The increase or decrease in the price of a stock between the beginning and the end of a trading day is assumed to be an equally likely random event. What is the probability that a stock will show an increase in its closing price on five consecutive days?

5.22 A recent YouGov (UK) survey reported that 27% of under-25-year-olds in the United Kingdom own tablets. (Data extracted from "Tablets Spur News Consumption," **bit.ly/12XpmR0**). Using the binomial distribution, what is the probability that in the next six under-25-year-olds surveyed,

a. four will own a tablet?
b. all six will own a tablet?
c. at least four will own a tablet?
d. What are the mean and standard deviation of the number of under-25-year-olds who will own a tablet in a survey of six?
e. What assumptions do you need to make in (a) through (c)?

5.23 A student is taking a multiple-choice exam in which each question has four choices. Assume that the student has no knowledge of the correct answers to any of the questions. She has decided on a strategy in which she will place four balls (marked *A*, *B*, *C*, and *D*) into a box. She randomly selects one ball for each question and replaces the ball in the box. The marking on the ball will determine her answer to the question. There are five multiple-choice questions on the exam. What is the probability that she will get

a. five questions correct?
b. at least four questions correct?
c. no questions correct?
d. no more than two questions correct?

5.24 A manufacturing company regularly conducts quality control checks at specified periods on the products it manufactures. Historically, the failure rate for LED light bulbs that the company manufactures is 5%. Suppose a random sample of 10 LED light bulbs is selected. What is the probability that

a. none of the LED light bulbs are defective?
b. exactly one of the LED light bulbs is defective?
c. two or fewer of the LED light bulbs are defective?
d. three or more of the LED light bulbs are defective?

5.25 When a customer places an order with Rudy's On-Line Office Supplies, a computerized accounting information system (AIS) automatically checks to see if the customer has exceeded his or her credit limit. Past records indicate that the probability of customers exceeding their credit limit is 0.05. Suppose that, on a given day, 20 customers place orders. Assume that the number of customers that the AIS detects as having exceeded their credit limit is distributed as a binomial random variable.

a. What are the mean and standard deviation of the number of customers exceeding their credit limits?
b. What is the probability that zero customers will exceed their credit limits?
c. What is the probability that one customer will exceed his or her credit limit?
d. What is the probability that two or more customers will exceed their credit limits?

SELF Test **5.26** In Example 5.5 on page 200, you and two friends decided to go to Wendy's. Now, suppose that instead you go to Burger King, which recently filled approximately 83% of orders correctly. What is the probability that

a. all three orders will be filled correctly?
b. none of the three will be filled correctly?
c. at least two of the three will be filled correctly?
d. What are the mean and standard deviation of the binomial distribution used in (a) through (c)? Interpret these values.

5.27 In Example 5.5 on page 200, you and two friends decided to go to Wendy's. Now, suppose that instead you go to McDonald's, which recently filled approximately 90.9% of the orders correctly. What is the probability that

a. all three orders will be filled correctly?
b. none of the three will be filled correctly?
c. at least two of the three will be filled correctly?
d. What are the mean and standard deviation of the binomial distribution used in (a) through (c)? Interpret these values.
e. Compare the result of (a) through (d) with those of Burger King in Problem 5.26 and Wendy's in Example 5.5 on page 200.

5.4 Poisson Distribution

Many studies are based on counts of the occurrences of a particular event in a given interval of time or space (often referred to as an *area of opportunity*). In such an **area of opportunity** there can be more than one occurrence of an event. The Poisson distribution can be used to compute probabilities in such situations. Examples of variables that follow the Poisson distribution are the surface defects on a new refrigerator, the number of network failures in a day, the number of people arriving at a bank, and the number of fleas on the body of a dog. You can use the **Poisson distribution** to calculate probabilities in situations such as these if the following properties hold:

- You are interested in counting the number of times a particular event occurs in a given area of opportunity. The area of opportunity is defined by time, length, surface area, and so forth.
- The probability that an event occurs in a given area of opportunity is the same for all the areas of opportunity.
- The number of events that occur in one area of opportunity is independent of the number of events that occur in any other area of opportunity.
- The probability that two or more events will occur in an area of opportunity approaches zero as the area of opportunity becomes smaller.

Consider the number of customers arriving during the lunch hour at a bank located in the central business district in a large city. You are interested in the number of customers who arrive each minute. Does this situation match the four properties of the Poisson distribution given earlier?

Problems for Section 5.4

LEARNING THE BASICS

5.28 Assume a Poisson distribution.
a. If $\lambda = 2.5$, find $P(X = 2)$.
b. If $\lambda = 8.0$, find $P(X = 8)$.
c. If $\lambda = 0.5$, find $P(X = 1)$.
d. If $\lambda = 3.7$, find $P(X = 0)$.

5.29 Assume a Poisson distribution.
a. If $\lambda = 2.0$, find $P(X \geq 2)$.
b. If $\lambda = 8.0$, find $P(X \geq 3)$.
c. If $\lambda = 0.5$, find $P(X \leq 1)$.
d. If $\lambda = 4.0$, find $P(X \geq 1)$.
e. If $\lambda = 5.0$, find $P(X \leq 3)$.

5.30 Assume a Poisson distribution with $\lambda = 5.0$. What is the probability that
a. $X = 1$?
b. $X < 1$?
c. $X > 1$?
d. $X \leq 1$?

APPLYING THE CONCEPTS

5.31 Assume that the number of network errors experienced in a day on a local area network (LAN) is distributed as a Poisson variable. The mean number of network errors experienced in a day is 2.4. What is the probability that in any given day
a. zero network errors will occur?
b. exactly one network error will occur?
c. two or more network errors will occur?
d. fewer than three network errors will occur?

SELF Test **5.32** The quality control manager of Marilyn's Cookies is inspecting a batch of chocolate-chip cookies that has just been baked. If the production process is in control, the mean number of chocolate-chip parts per cookie is 6.0. What is the probability that in any particular cookie being inspected
a. fewer than five chocolate-chip parts will be found?
b. exactly five chocolate-chip parts will be found?
c. five or more chocolate-chip parts will be found?
d. either four or five chocolate-chip parts will be found?

5.33 Refer to Problem 5.32. How many cookies in a batch of 100 should the manager expect to discard if company policy requires that all chocolate-chip cookies sold have at least four chocolate-chip parts?

5.34 The U.S. Department of Transportation maintains statistics for mishandled bags per 1,000 airline passengers. In February 2013, Delta mishandled 2.05 bags per 1,000 passengers. What is the probability that in the next 1,000 passengers, Delta will have
a. no mishandled bags?
b. at least one mishandled bag?
c. at least two mishandled bags?

5.35 The U.S. Department of Transportation maintains statistics for involuntary denial of boarding. In February 2013, the American Airlines rate of involuntarily denying boarding was 0.74 per 10,000 passengers. What is the probability that in the next 10,000 passengers, there will be
a. no one involuntarily denied boarding?
b. at least one person involuntarily denied boarding?
c. at least two persons involuntarily denied boarding?

5.36 The Consumer Financial Protection Bureau's consumer response team hears directly from consumers about the challenges they face in the marketplace, brings their concerns to the attention of financial institutions, and assists in addressing their complaints. The consumer response team accepts complaints related to mortgages, bank accounts and services, private student loans, other consumer loans, and credit reporting. An analysis of complaints over time indicates that the mean number of credit reporting complaints registered by consumers is 2.15 per day. (Source: *Consumer Response: A Snapshot of Complaints Received*, **1.usa.gov/WZ9N8Q**.) Assume that the number of credit reporting complaints registered by consumers is distributed as a Poisson random variable. What is the probability that on a given day
a. no credit reporting complaints will be registered by consumers?
b. exactly one credit reporting complaint will be registered by consumers?
c. more than one credit reporting complaint will be registered by consumers?
d. fewer than two credit reporting complaints will be registered by consumers?

5.37 J.D. Power and Associates calculates and publishes various statistics concerning car quality. The dependability score measures problems experienced during the past 12 months by original owners of three-year-old vehicles (those that were introduced for the 2010 model year). For these models of cars, Ford had 1.27 problems per car and Toyota had 1.12 problems per car. (Data extracted from "2013 U.S. Vehicle Dependability Study," J.D. Power and Associates, February 13, 2013, **bit.ly/101aR9l**.) Let X be equal to the number of problems with a three-year-old Ford.
a. What assumptions must be made in order for X to be distributed as a Poisson random variable? Are these assumptions reasonable?

Making the assumptions as in (a), if you purchased a Ford in the 2010 model year, what is the probability that in the past 12 months, the car had
b. zero problems?
c. two or fewer problems?
d. Give an operational definition for *problem*. Why is the operational definition important in interpreting the initial quality score?

5.38 Refer to Problem 5.37. If you purchased a Toyota in the 2010 model year, what is the probability that in the past 12 months the car had
a. zero problems?
b. two or fewer problems?
c. Compare your answers in (a) and (b) to those for the Ford in Problem 5.37 (b) and (c).

5.39 Refer to Problem 5.37. Another press release reported in 2012 that for 2009 model cars, Ford had 1.24 problems per car and Toyota had 1.04 problems per car. (Data extracted from "2013 U.S. Vehicle Dependability Study," J.D. Power and Associates, February 13, 2013, **bit.ly/101aR9l**.) If you purchased a 2009 Ford, what is the probability that in the past 12 months the car had
a. zero problems?
b. two or fewer problems?
c. Compare your answers in (a) and (b) to those for the 2010 model year Ford in Problem 5.37 (b) and (c).

5.40 Refer to Problem 5.39. If you purchased a 2009 Toyota, what is the probability that in the past 12 months, the car had

a. zero problems?
b. two or fewer problems?
c. Compare your answers in (a) and (b) to those for the 2010 model year Toyota in Problem 5.38 (a) and (b).

5.41 A toll-free phone number is available from 9 A.M. to 9 P.M. for your customers to register complaints about a product purchased from your company. Past history indicates that an average of 0.8 calls is received per minute.

a. What properties must be true about the situation described here in order to use the Poisson distribution to calculate probabilities concerning the number of phone calls received in a one-minute period?
Assuming that this situation matches the properties discussed in (a), what is the probability that during a one-minute period
b. zero phone calls will be received?
c. three or more phone calls will be received?
d. What is the maximum number of phone calls that will be received in a one-minute period 99.99% of the time?

5.5 Hypergeometric Distribution

Both the binomial distribution and the **hypergeometric distribution** use the number of events of interest in a sample containing n observations. One of the differences in these two probability distributions is in the way the samples are selected. For the binomial distribution, the sample data are selected *with* replacement from a *finite* population or *without* replacement from an *infinite* population. Thus, the probability of an event of interest, π, is constant over all observations, and the result of any particular observation is independent of the result of any other observation. For the hypergeometric distribution, the sample data are selected *without* replacement from a *finite* population. Thus, the result of one observation is dependent on the results of the previous observations.

Consider a population of size N. Let E represent the total number of events of interest in the population. The hypergeometric distribution is then used to find the probability of x events of interest in a sample of size n, selected without replacement. Equation (5.15) represents the mathematical expression of the hypergeometric distribution for finding x events of interest, given a knowledge of n, N, and E.

HYPERGEOMETRIC DISTRIBUTION

$$P(X = x \mid n, N, E) = \frac{\binom{E}{x}\binom{N-E}{n-x}}{\binom{N}{n}} \tag{5.15}$$

where

$P(X = x \mid n, N, E)$ = probability of x events of interest, given knowledge of n, N, and E
n = sample size
N = population size
E = number of events of interest in the population
$N - E$ = number of events that are not of interest in the population
x = number of events of interest in the sample
$\binom{E}{x} = {}_EC_x$ = number of combinations [see Equation (5.10) on page 196]
$x \le E$
$x \le n$

Because the number of events of interest in the sample, represented by x, cannot be greater than the number of events of interest in the population, E, nor can x be greater than the sample size, n, the range of the hypergeometric random variable is limited to the sample size or to the number of events of interest in the population, whichever is smaller.

Example 5.7 shows an application of the hypergeometric distribution in portfolio selection.

EXAMPLE 5.7

Computing Hypergeometric Probabilities

You are a financial analyst facing the task of selecting mutual funds to purchase for a client's portfolio. You have narrowed the funds to be selected to 10 different funds. In order to diversify your client's portfolio, you will recommend the purchase of 4 different funds. Six of the funds are growth funds. What is the probability that of the 4 funds selected, 3 are growth funds?

SOLUTION Using Equation (5.15) with $X = 3, n = 4, N = 10$, and $E = 6$,

$$
\begin{aligned}
P(X = 3 \mid n = 4, N = 10, E = 6) &= \frac{\binom{6}{3}\binom{4}{1}}{\binom{10}{4}} \\
&= \frac{\left(\dfrac{6!}{3!(3)!}\right)\left(\dfrac{(4)!}{(1)!(3)!}\right)}{\left(\dfrac{10!}{4!(6)!}\right)} \\
&= 0.3810
\end{aligned}
$$

The probability that of the 4 funds selected, 3 are growth funds, is 0.3810, or 38.10%.

Problems for Section 5.5

LEARNING THE BASICS

5.42 Determine the following:

a. If $n = 4, N = 10$, and $E = 5$, find $P(X = 3)$.
b. If $n = 4, N = 6$, and $E = 3$, find $P(X = 1)$.
c. If $n = 5, N = 12$, and $E = 3$, find $P(X = 0)$.
d. If $n = 3, N = 10$, and $E = 3$, find $P(X = 3)$.

5.43 Referring to Problem 5.42, compute the mean and standard deviation for the hypergeometric distributions described in (a) through (d).

APPLYING THE CONCEPTS

SELF Test **5.44** An auditor for the Internal Revenue Service is selecting a sample of 6 tax returns for an audit. If 2 or more of these returns are "improper," the entire population of 100 tax returns will be audited. What is the probability that the entire population will be audited if the true number of improper returns in the population is

a. 25?
b. 30?
c. 5?
d. 10?
e. Discuss the differences in your results, depending on the true number of improper returns in the population.

5.45 KSDLDS-Pros, an IT project management consulting firm, is forming an IT project management team of 5 professionals. In the firm of 50 professionals, 8 are considered to be data analytics specialists. If the professionals are selected at random, what is the probability that the team will include

a. no data analytics specialist?
b. at least one data analytics specialist?
c. no more than two data analytics specialists?
d. What is your answer to (a) if the team consists of 7 members?

5.46 From an inventory of 30 cars being shipped to a local automobile dealer, 4 are SUVs. What is the probability that if 4 cars arrive at a particular dealership,

a. all 4 are SUVs?
b. none are SUVs?
c. at least 1 is an SUV?
d. What are your answers to (a) through (c) if 6 cars being shipped are SUVs?

5.47 As a quality control manager, you are responsible for checking the quality level of AC adapters for tablet PCs that your company manufactures. You must reject a shipment if you find 4 defective units. Suppose a shipment of 40 AC adapters has 8 defective units and 32 nondefective units. If you sample 12 AC adapters, what's the probability that

a. there will be no defective units in the shipment?
b. there will be at least 1 defective unit in the shipment?
c. there will be 4 defective units in the shipment?
d. the shipment will be accepted?

5.48 In Example 5.7 above, a financial analyst was facing the task of selecting mutual funds to purchase for a client's portfolio. Suppose that the number of funds had been narrowed to 12 funds instead of the 10 funds (still with 6 growth funds) in Example 5.7. What is the probability that of the 4 funds selected,

a. exactly 1 is a growth fund?
b. at least 1 is a growth fund?
c. 3 are growth fund?
d. Compare the result of (c) to the result of Example 5.7.

KEY EQUATIONS

Expected Value, μ, of a Discrete Variable

$$\mu = E(X) = \sum_{i=1}^{N} x_i P(X = x_i) \quad \textbf{(5.1)}$$

Variance of a Discrete Variable

$$\sigma^2 = \sum_{i=1}^{N} [x_i - E(X)]^2 P(X = x_i) \quad \textbf{(5.2)}$$

Standard Deviation of a Discrete Variable

$$\sigma = \sqrt{\sigma^2} = x\sqrt{\sum_{i=1}^{N} x_i [x_i - E(X)]^2 P(X = x_i)} \quad \textbf{(5.3)}$$

Covariance

$$\sigma_{XY} = \sum_{i=1}^{N} [x_i - E(X)][y_i - E(Y)]\, P(x_i, y_i) \quad \textbf{(5.4)}$$

Expected Value of the Sum of Two Variables

$$E(X + Y) = E(X) + E(Y) \quad \textbf{(5.5)}$$

Variance of the Sum of Two Variables

$$Var(X + Y) = \sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\sigma_{XY} \quad \textbf{(5.6)}$$

Standard Deviation of the Sum of Two Variables

$$\sigma_{X+Y} = \sqrt{\sigma^2_{X+Y}} \quad \textbf{(5.7)}$$

Portfolio Expected Return

$$E(P) = wE(X) + (1 - w)E(Y) \quad \textbf{(5.8)}$$

Portfolio Risk

$$\sigma_p = \sqrt{w^2\sigma^2_X + (1 - w)^2\sigma^2_Y + 2w(1 - w)\sigma_{XY}} \quad \textbf{(5.9)}$$

Combinations

$${}_nC_x = \frac{n!}{x!(n - x)!} \quad \textbf{(5.10)}$$

Binomial Distribution

$$P(X = x \mid n, \pi) = \frac{n!}{x!(n - x)!}\pi^x(1 - \pi)^{n-x} \quad \textbf{(5.11)}$$

Mean of the Binomial Distribution

$$\mu = E(X) = n\pi \quad \textbf{(5.12)}$$

Standard Deviation of the Binomial Distribution

$$\sigma = \sqrt{\sigma^2} = \sqrt{Var(X)} = \sqrt{n\pi(1 - \pi)} \quad \textbf{(5.13)}$$

Poisson Distribution

$$P(X = x \mid \lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \textbf{(5.14)}$$

Hypergeometric Distribution

$$P(X = x \mid n, N, E) = \frac{\binom{E}{x}\binom{N - E}{n - x}}{\binom{N}{n}} \quad \textbf{(5.15)}$$

Mean of the Hypergeometric Distribution

$$\mu = E(X) = \frac{nE}{N} \quad \textbf{(5.16)}$$

Standard Deviation of the Hypergeometric Distribution

$$\sigma = \sqrt{\frac{nE(N - E)}{N^2}}\sqrt{\frac{N - n}{N - 1}} \quad \textbf{(5.17)}$$

KEY TERMS

CHECKING YOUR UNDERSTANDING

5.49 What is the meaning of the expected value of a random variable?

5.50 What are the four properties that must be present in order to use the binomial distribution?

5.51 What are the four properties that must be present in order to use the Poisson distribution?

5.52 When do you use the hypergeometric distribution instead of the binomial distribution?

CHAPTER REVIEW PROBLEMS

5.53 Darwin Head, a 35-year-old sawmill worker, won $1 million and a Chevrolet Malibu Hybrid by scoring 15 goals within 24 seconds at the Vancouver Canucks National Hockey League game (B. Ziemer, "Darwin Evolves into an Instant Millionaire," *Vancouver Sun*, February 28, 2008, p. 1). Head said he would use the money to pay off his mortgage and provide for his children, and he had no plans to quit his job. The contest was part of the Chevrolet Malibu Million Dollar Shootout, sponsored by General Motors Canadian Division. Did GM-Canada risk the $1 million? No! GM-Canada purchased event insurance from a company specializing in promotions at sporting events such as a half-court basketball shot or a hole-in-one giveaway at the local charity golf outing. The event insurance company estimates the probability of a contestant winning the contest, and for a modest charge, insures the event. The promoters pay the insurance premium but take on no added risk as the insurance company will make the large payout in the unlikely event that a contestant wins. To see how it works, suppose that the insurance company estimates that the probability a contestant would win a million-dollar shootout is 0.001 and that the insurance company charges $4,000.

a. Calculate the expected value of the profit made by the insurance company.
b. Many call this kind of situation a win–win opportunity for the insurance company and the promoter. Do you agree? Explain.

5.54 Between 1896—when the Dow Jones index was created—and 2012, the index rose in 65% of the years. (Sources: M. Hulbert, "What the Past Can't Tell Investors," *The New York Times*, January 3, 2010, p. BU2 and **bit.ly/100zwvT**.) Based on this information, and assuming a binomial distribution, what do you think is the probability that the stock market will rise

a. next year?
b. the year after next?
c. in four of the next five years?
d. in none of the next five years?
e. For this situation, what assumption of the binomial distribution might not be valid?

5.55 Smartphone adoption among American teens has increased substantially, and mobile access to the Internet is pervasive. One in four teenagers are "cell mostly" Internet users—that is, they *mostly go online using their phone* and not using some other device such as a desktop or laptop computer. (Source: *Teens and Technology 2013*, Pew Research Center, **bit.ly/101ciF1**.)

If a sample of 10 American teens is selected, what is the probability that

a. 4 are "cell mostly" Internet users?
b. at least 4 are "cell mostly" Internet users?
c. at most 8 are "cell mostly" Internet users?
d. If you selected the sample in a particular geographical area and found that none of the 10 respondents are "cell mostly" Internet users, what conclusions might you reach about whether the percentage of "cell mostly" Internet users in this area was 25%?

5.56 One theory concerning the Dow Jones Industrial Average is that it is likely to increase during U.S. presidential election years. From 1964 through 2012, the Dow Jones Industrial Average increased in 10 of the 13 U.S. presidential election years. Assuming that this indicator is a random event with no predictive value, you would expect that the indicator would be correct 50% of the time.

a. What is the probability of the Dow Jones Industrial Average increasing in 10 or more of the 13 U.S. presidential election years if the probability of an increase in the Dow Jones Industrial Average is 0.50?
b. What is the probability that the Dow Jones Industrial Average will increase in 10 or more of the 13 U.S. presidential election years if the probability of an increase in the Dow Jones Industrial Average in any year is 0.75?

5.57 Medical billing errors and fraud are on the rise. According to Medical Billing Advocates of America, 8 out of 10 times, the medical bills that you get are not right. (Data extracted from "Services Diagnose, Treat Medical Billing Errors," *USA Today*, June 20, 2012.) If a sample of 10 medical bills is selected, what is the probability that

a. 0 medical bills will contain errors?
b. exactly 5 medical bills will contain errors?
c. more than 5 medical bills will contain errors?
d. What are the mean and standard deviation of the probability distribution?

5.58 Refer to Problem 5.57. Suppose that a quality improvement initiative has reduced the percentage of medical bills containing errors to 40%. If a sample of 10 medical bills is selected, what is the probability that

a. 0 medical bills will contain errors?
b. exactly 5 medical bills will contain errors?
c. more than 5 medical bills contain errors?
d. What are the mean and standard deviation of the probability distribution?
e. Compare the results of (a) through (c) to those of Problem 5.57 (a) through (c).

5.59 Social log-ins involve recommending or sharing an article that you read online. According to Janrain, in the first quarter of 2013, 46% signed in via Facebook compared with 34% for Google.

(Source: "Social Login Trends Across the Web for Q1 2013," **bit.ly/ZQCRSF**.) If a sample of 10 social log-ins is selected, what is the probability that

a. more than 5 signed in using Facebook?
b. more than 5 signed in using Google?
c. none signed in using Facebook?
d. What assumptions did you have to make to answer (a) through (c)?

5.60 The Consumer Financial Protection Bureau's consumer response team hears directly from consumers about the challenges they face in the marketplace, brings their concerns to the attention of financial institutions, and assists in addressing their complaints. Consumer response accepts complaints related to mortgages, bank accounts and services, private student loans, other consumer loans, and credit reporting. Of the consumers who registered a bank account and service complaint, 41% cited "account management" as the type of complaint; these complaints are related to opening, closing, or managing the account and address issues, such as confusing marketing, denial, fees, statements, and joint accounts. (Source: *Consumer Response: A Snapshot of Complaints Received*, **1.usa.gov/WZ9N8Q**.) Consider a sample of 20 consumers who registered bank account and service complaints. Use the binomial model to answer the following questions:

a. What is the expected value, or mean, of the binomial distribution?
b. What is the standard deviation of the binomial distribution?
c. What is the probability that 10 of the 20 consumers cited "account management" as the type of complaint?
d. What is the probability that no more than 5 of the consumers cited "account management" as the type of complaint?
e. What is the probability that 5 or more of the consumers cited "account management" as the type of complaint?

5.61 Refer to Problem 5.60. In the same time period, 27% of the consumers registering a bank account and service compliant cited "deposit and withdrawal" as the type of complaint; these are issues such as transaction holds and unauthorized transactions.

a. What is the expected value, or mean, of the binomial distribution?
b. What is the standard deviation of the binomial distribution?
c. What is the probability that none of the 20 consumers cited "deposit and withdrawal" as the type of complaint?
d. What is the probability that no more than 2 of the consumers cited "deposit and withdrawal" as the type of complaint?
e. What is the probability that 3 or more of the consumers cited "deposit and withdrawal" as the type of complaint?

5.62 One theory concerning the S&P 500 Index is that if it increases during the first five trading days of the year, it is likely to increase during the entire year. From 1950 through 2012, the S&P 500 Index had these early gains in 40 years (in 2011 there was virtually no change). In 35 of these 40 years, the S&P 500 Index increased for the entire year. Assuming that this indicator is a random event with no predictive value, you would expect that the indicator would be correct 50% of the time. What is the probability of the S&P 500 Index increasing in 35 or more years if the true probability of an increase in the S&P 500 Index is

a. 0.50?
b. 0.70?
c. 0.90?
d. Based on the results of (a) through (c), what do you think is the probability that the S&P 500 Index will increase if there is an early gain in the first five trading days of the year? Explain.

5.63 *Spurious correlation* refers to the apparent relationship between variables that either have no true relationship or are related to other variables that have not been measured. One widely publicized stock market indicator in the United States that is an example of spurious correlation is the relationship between the winner of the National Football League Super Bowl and the performance of the Dow Jones Industrial Average in that year. The "indicator" states that when a team that existed before the National Football League merged with the American Football League wins the Super Bowl, the Dow Jones Industrial Average will increase in that year. (Of course, any correlation between these is spurious as one thing has absolutely nothing to do with the other!) Since the first Super Bowl was held in 1967 through 2012, the indicator has been correct 37 out of 46 times. Assuming that this indicator is a random event with no predictive value, you would expect that the indicator would be correct 50% of the time.

a. What is the probability that the indicator would be correct 37 or more times in 46 years?
b. What does this tell you about the usefulness of this indicator?

5.64 The National Insurance Crime Bureau says that Miami-Dade, Broward, and Palm Beach counties account for a substantial number of questionable insurance claims referred to investigators. (Source: "United Auto Courts Reports," **bit.ly/100DZi9**.) Assume that the number of questionable insurance claims referred to investigators by Miami-Dade, Broward, and Palm Beach counties is distributed as a Poisson random variable with a mean of 10 per day.

a. What assumptions need to be made so that the number of questionable insurance claims referred to investigators by Miami-Dade, Broward, and Palm Beach counties is distributed as a Poisson random variable?

Making the assumptions given in (a), what is the probability that

b. 5 questionable insurance claims will be referred to investigators by Miami-Dade, Broward, and Palm Beach counties in a day?
c. 10 or fewer questionable insurance claims will be referred to investigators by Miami-Dade, Broward, and Palm Beach counties in a day?
d. 11 or more questionable insurance claims will be referred to investigators by Miami-Dade, Broward, and Palm Beach counties in a day?

5.65 In the Florida lottery Lotto game, you select six numbers from a pool of numbers from 1 to 53 (see **flalottery.com**). Each wager costs $1. You win the jackpot if you match all six numbers that you have selected.

Find the probability of

a. winning the jackpot.
b. matching five numbers.
c. matching four numbers.
d. matching three numbers.
e. matching two numbers.
f. matching one number.
g. matching none of the numbers.
h. If you match zero, one, or two numbers, you do not win anything. What is the probability that you will not win anything?
i. The Lotto ticket gives complete game rules and probabilities of matching zero through six numbers. The lottery ticket has the saying "A Win for Education" on the back of the ticket. Do you think Florida's slogan and the printed complete game rules and probabilities of matching zero through six numbers is an ethical approach to running the lottery game?