in Common Pleas Court, Domestic Relations Court, and Municipal Court. Two of the judges (Dinkelacker and Hogan) did not serve in the same court for the entire three-year period.

The purpose of the newspaper's study was to evaluate the performance of the judges. Appeals are often the result of mistakes made by judges, and the newspaper wanted to know which judges were doing a good job and which were making too many mistakes. You are called in to assist in the data analysis. Use your knowledge of probability and conditional probability to help with the ranking of the judges. You also may be able to analyze the likelihood of appeal and reversal for cases handled by different courts.

Managerial Report

Prepare a report with your rankings of the judges. Also, include an analysis of the likelihood of appeal and case reversal in the three courts. At a minimum, your report should include the following:

1. The probability of cases being appealed and reversed in the three different courts.
2. The probability of a case being appealed for each judge.
3. The probability of a case being reversed for each judge.
4. The probability of reversal given an appeal for each judge.
5. Rank the judges within each court. State the criteria you used and provide a rationale for your choice.

CHAPTER 5

Discrete Probability Distributions

CONTENTS

STATISTICS in PRACTICE

CITIBANK*
LONG ISLAND CITY, NEW YORK

In March 2007, the Forbes Global 2000 report listed Citigroup as the world's largest company with assets of nearly $2 trillion and profits of more than $21 billion. Citibank, the retail banking division of Citigroup, offers a wide range of financial services including checking and saving accounts, loans and mortgages, insurance, and investment services, within the framework of a unique strategy for delivering those services called Citibanking.

Citibank was one of the first banks in the United States to introduce automatic teller machines (ATMs). Citibank's ATMs, located in Citicard Banking Centers (CBCs), let customers do all of their banking in one place with the touch of a finger, 24 hours a day, 7 days a week. More than 150 different banking functions—from deposits to managing investments—can be performed with ease. Citibank customers use ATMs for 80% of their transactions.

Each Citibank CBC operates as a waiting line system with randomly arriving customers seeking service at one of the ATMs. If all ATMs are busy, the arriving customers wait in line. Periodic CBC capacity studies are used to analyze customer waiting times and to determine whether additional ATMs are needed.

Data collected by Citibank showed that the random customer arrivals followed a probability distribution known as the Poisson distribution. Using the Poisson distribution, Citibank can compute probabilities for the number of customers arriving at a CBC during any time period and make decisions concerning the number of ATMs needed. For example, let x = the number of customers arriving during a one-minute period. Assuming that a particular CBC has a mean arrival rate of two customers per minute, the following table shows the probabilities for the number of customers arriving during a one-minute period.

A Citibank state-of-the-art ATM. © Jeff Greenberg/Photo Edit.

x	Probability
0	.1353
1	.2707
2	.2707
3	.1804
4	.0902
5 or more	.0527

Discrete probability distributions, such as the one used by Citibank, are the topic of this chapter. In addition to the Poisson distribution, you will learn about the binomial and hypergeometric distributions and how they can be used to provide helpful probability information.

*The authors are indebted to Ms. Stacey Karter, Citibank, for providing this Statistics in Practice.

In this chapter we continue the study of probability by introducing the concepts of random variables and probability distributions. The focus of this chapter is discrete probability distributions. Three special discrete probability distributions—the binomial, Poisson, and hypergeometric—are covered.

5.1 Random Variables

In Chapter 4 we defined the concept of an experiment and its associated experimental outcomes. A random variable provides a means for describing experimental outcomes using numerical values. Random variables must assume numerical values.

Random variables must assume numerical values.

RANDOM VARIABLE

A **random variable** is a numerical description of the outcome of an experiment.

In effect, a random variable associates a numerical value with each possible experimental outcome. The particular numerical value of the random variable depends on the outcome of the experiment. A random variable can be classified as being either *discrete* or *continuous* depending on the numerical values it assumes.

Discrete Random Variables

A random variable that may assume either a finite number of values or an infinite sequence of values such as 0, 1, 2, . . . is referred to as a **discrete random variable**. For example, consider the experiment of an accountant taking the certified public accountant (CPA) examination. The examination has four parts. We can define a random variable as x = the number of parts of the CPA examination passed. It is a discrete random variable because it may assume the finite number of values 0, 1, 2, 3, or 4.

As another example of a discrete random variable, consider the experiment of cars arriving at a tollbooth. The random variable of interest is x = the number of cars arriving during a one-day period. The possible values for x come from the sequence of integers 0, 1, 2, and so on. Hence, x is a discrete random variable assuming one of the values in this infinite sequence.

Although the outcomes of many experiments can naturally be described by numerical values, others cannot. For example, a survey question might ask an individual to recall the message in a recent television commercial. This experiment would have two possible outcomes: the individual cannot recall the message and the individual can recall the message. We can still describe these experimental outcomes numerically by defining the discrete random variable x as follows: let $x = 0$ if the individual cannot recall the message and $x = 1$ if the individual can recall the message. The numerical values for this random variable are arbitrary (we could use 5 and 10), but they are acceptable in terms of the definition of a random variable—namely, x is a random variable because it provides a numerical description of the outcome of the experiment.

Table 5.1 provides some additional examples of discrete random variables. Note that in each example the discrete random variable assumes a finite number of values or an infinite sequence of values such as 0, 1, 2, ... These types of discrete random variables are discussed in detail in this chapter.

TABLE 5.1 EXAMPLES OF DISCRETE RANDOM VARIABLES

Experiment	Random Variable (x)	Possible Values for the Random Variable
Contact five customers	Number of customers who place an order	0, 1, 2, 3, 4, 5
Inspect a shipment of 50 radios	Number of defective radios	0, 1, 2, . . . , 49, 50
Operate a restaurant for one day	Number of customers	0, 1, 2, 3, . . .
Sell an automobile	Gender of the customer	0 if male; 1 if female

Continuous Random Variables

A random variable that may assume any numerical value in an interval or collection of intervals is called a **continuous random variable**. Experimental outcomes based on measurement scales such as time, weight, distance, and temperature can be described by continuous random variables. For example, consider an experiment of monitoring incoming telephone calls to the claims office of a major insurance company. Suppose the random variable of interest is x = the time between consecutive incoming calls in minutes. This random variable may assume any value in the interval $x \geq 0$. Actually, an infinite number of values are possible for x, including values such as 1.26 minutes, 2.751 minutes, 4.3333 minutes, and so on. As another example, consider a 90-mile section of interstate highway I-75 north of Atlanta, Georgia. For an emergency ambulance service located in Atlanta, we might define the random variable as x = number of miles to the location of the next traffic accident along this section of I-75. In this case, x would be a continuous random variable assuming any value in the interval $0 \leq x \leq 90$. Additional examples of continuous random variables are listed in Table 5.2. Note that each example describes a random variable that may assume any value in an interval of values. Continuous random variables and their probability distributions will be the topic of Chapter 6.

TABLE 5.2 EXAMPLES OF CONTINUOUS RANDOM VARIABLES

Experiment	Random Variable (x)	Possible Values for the Random Variable
Operate a bank	Time between customer arrivals in minutes	$x \geq 0$
Fill a soft drink can (max = 12.1 ounces)	Number of ounces	$0 \leq x \leq 12.1$
Construct a new library	Percentage of project complete after six months	$0 \leq x \leq 100$
Test a new chemical process	Temperature when the desired reaction takes place (min 150° F; max 212° F)	$150 \leq x \leq 212$

NOTES AND COMMENTS

One way to determine whether a random variable is discrete or continuous is to think of the values of the random variable as points on a line segment. Choose two points representing values of the random variable. If the entire line segment between the two points also represents possible values for the random variable, then the random variable is continuous.

Exercises

Methods

SELF test

1. Consider the experiment of tossing a coin twice.
   a. List the experimental outcomes.
   b. Define a random variable that represents the number of heads occurring on the two tosses.
   c. Show what value the random variable would assume for each of the experimental outcomes.
   d. Is this random variable discrete or continuous?

2. Consider the experiment of a worker assembling a product.
   a. Define a random variable that represents the time in minutes required to assemble the product.
   b. What values may the random variable assume?
   c. Is the random variable discrete or continuous?

Applications

SELF test

3. Three students scheduled interviews for summer employment at the Brookwood Institute. In each case the interview results in either an offer for a position or no offer. Experimental outcomes are defined in terms of the results of the three interviews.
   a. List the experimental outcomes.
   b. Define a random variable that represents the number of offers made. Is the random variable continuous?
   c. Show the value of the random variable for each of the experimental outcomes.
4. In November the U.S. unemployment rate was 4.5% (*USA Today,* January 4, 2007). The Census Bureau includes nine states in the Northeast region. Assume that the random variable of interest is the number of Northeast states with an unemployment rate in November that was less than 4.5%. What values may this random variable assume?
5. To perform a certain type of blood analysis, lab technicians must perform two procedures. The first procedure requires either one or two separate steps, and the second procedure requires either one, two, or three steps.
   a. List the experimental outcomes associated with performing the blood analysis.
   b. If the random variable of interest is the total number of steps required to do the complete analysis (both procedures), show what value the random variable will assume for each of the experimental outcomes.
6. Listed is a series of experiments and associated random variables. In each case, identify the values that the random variable can assume and state whether the random variable is discrete or continuous.

Experiment	Random Variable (x)
a. Take a 20-question examination	Number of questions answered correctly
b. Observe cars arriving at a tollbooth for 1 hour	Number of cars arriving at tollbooth
c. Audit 50 tax returns	Number of returns containing errors
d. Observe an employee's work	Number of nonproductive hours in an eight-hour workday
e. Weigh a shipment of goods	Number of pounds

5.2 Discrete Probability Distributions

The **probability distribution** for a random variable describes how probabilities are distributed over the values of the random variable. For a discrete random variable x, the probability distribution is defined by a **probability function**, denoted by $f(x)$. The probability function provides the probability for each value of the random variable.

As an illustration of a discrete random variable and its probability distribution, consider the sales of automobiles at DiCarlo Motors in Saratoga, New York. Over the past 300 days of operation, sales data show 54 days with no automobiles sold, 117 days with 1 automobile sold, 72 days with 2 automobiles sold, 42 days with 3 automobiles sold, 12 days with 4 automobiles sold, and 3 days with 5 automobiles sold. Suppose we consider the experiment of selecting a day of operation at DiCarlo Motors and define the random variable of interest as x = the number of automobiles sold during a day. From historical data, we know

x is a discrete random variable that can assume the values 0, 1, 2, 3, 4, or 5. In probability function notation, $f(0)$ provides the probability of 0 automobiles sold, $f(1)$ provides the probability of 1 automobile sold, and so on. Because historical data show 54 of 300 days with 0 automobiles sold, we assign the value $54/300 = .18$ to $f(0)$, indicating that the probability of 0 automobiles being sold during a day is .18. Similarly, because 117 of 300 days had 1 automobile sold, we assign the value $117/300 = .39$ to $f(1)$, indicating that the probability of exactly 1 automobile being sold during a day is .39. Continuing in this way for the other values of the random variable, we compute the values for $f(2)$, $f(3)$, $f(4)$, and $f(5)$ as shown in Table 5.3, the probability distribution for the number of automobiles sold during a day at DiCarlo Motors.

A primary advantage of defining a random variable and its probability distribution is that once the probability distribution is known, it is relatively easy to determine the probability of a variety of events that may be of interest to a decision maker. For example, using the probability distribution for DiCarlo Motors as shown in Table 5.3, we see that the most probable number of automobiles sold during a day is 1 with a probability of $f(1) = .39$. In addition, there is an $f(3) + f(4) + f(5) = .14 + .04 + .01 = .19$ probability of selling three or more automobiles during a day. These probabilities, plus others the decision maker may ask about, provide information that can help the decision maker understand the process of selling automobiles at DiCarlo Motors.

In the development of a probability function for any discrete random variable, the following two conditions must be satisfied.

These conditions are the analogs to the two basic requirements for assigning probabilities to experimental outcomes presented in Chapter 4.

REQUIRED CONDITIONS FOR A DISCRETE PROBABILITY FUNCTION

$$f(x) \geq 0 \tag{5.1}$$

$$\Sigma f(x) = 1 \tag{5.2}$$

Table 5.3 shows that the probabilities for the random variable x satisfy equation (5.1); $f(x)$ is greater than or equal to 0 for all values of x. In addition, because the probabilities sum to 1, equation (5.2) is satisfied. Thus, the DiCarlo Motors probability function is a valid discrete probability function.

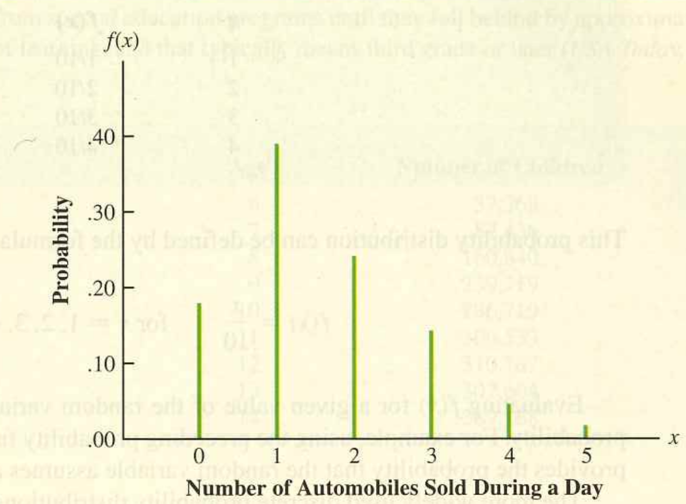
We can also present probability distributions graphically. In Figure 5.1 the values of the random variable x for DiCarlo Motors are shown on the horizontal axis and the probability associated with these values is shown on the vertical axis.

In addition to tables and graphs, a formula that gives the probability function, $f(x)$, for every value of x is often used to describe probability distributions. The simplest example of

TABLE 5.3 PROBABILITY DISTRIBUTION FOR THE NUMBER OF AUTOMOBILES SOLD DURING A DAY AT DICARLO MOTORS

x	$f(x)$
0	.18
1	.39
2	.24
3	.14
4	.04
5	.01
Total	1.00

FIGURE 5.1 GRAPHICAL REPRESENTATION OF THE PROBABILITY DISTRIBUTION FOR THE NUMBER OF AUTOMOBILES SOLD DURING A DAY AT DICARLO MOTORS



a discrete probability distribution given by a formula is the **discrete uniform probability distribution**. Its probability function is defined by equation (5.3).

DISCRETE UNIFORM PROBABILITY FUNCTION

$$f(x) = 1/n \tag{5.3}$$

where

n = the number of values the random variable may assume

For example, suppose that for the experiment of rolling a die we define the random variable x to be the number of dots on the upward face. For this experiment, $n = 6$ values are possible for the random variable; $x = 1, 2, 3, 4, 5, 6$. Thus, the probability function for this discrete uniform random variable is

$$f(x) = 1/6 \qquad x = 1, 2, 3, 4, 5, 6$$

The possible values of the random variable and the associated probabilities are shown.

x	$f(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

As another example, consider the random variable x with the following discrete probability distribution.

x	$f(x)$
1	1/10
2	2/10
3	3/10
4	4/10

This probability distribution can be defined by the formula

$$f(x) = \frac{x}{10} \qquad \text{for } x = 1, 2, 3, \text{ or } 4$$

Evaluating $f(x)$ for a given value of the random variable will provide the associated probability. For example, using the preceding probability function, we see that $f(2) = 2/10$ provides the probability that the random variable assumes a value of 2.

The more widely used discrete probability distributions generally are specified by formulas. Three important cases are the binomial, Poisson, and hypergeometric distributions; these distributions are discussed later in the chapter.

Exercises

Methods

SELF test

7. The probability distribution for the random variable x follows.

x	$f(x)$
20	.20
25	.15
30	.25
35	.40

a. Is this probability distribution valid? Explain.
b. What is the probability that $x = 30$?
c. What is the probability that x is less than or equal to 25?
d. What is the probability that x is greater than 30?

Applications

SELF test

8. The following data were collected by counting the number of operating rooms in use at Tampa General Hospital over a 20-day period: On three of the days only one operating room was used, on five of the days two were used, on eight of the days three were used, and on four days all four of the hospital's operating rooms were used.
a. Use the relative frequency approach to construct a probability distribution for the number of operating rooms in use on any given day.
b. Draw a graph of the probability distribution.
c. Show that your probability distribution satisfies the required conditions for a valid discrete probability distribution.

9. Nationally, 38% of fourth-graders cannot read an age-appropriate book. The following data show the number of children, by age, identified as learning disabled under special education. Most of these children have reading problems that should be identified and corrected before third grade. Current federal law prohibits most children from receiving extra help from special education programs until they fall behind by approximately two years' worth of learning, and that typically means third grade or later (*USA Today,* September 6, 2001).

Age	Number of Children
6	37,369
7	87,436
8	160,840
9	239,719
10	286,719
11	306,533
12	310,787
13	302,604
14	289,168

Suppose that we want to select a sample of children identified as learning disabled under special education for a program designed to improve reading ability. Let x be a random variable indicating the age of one randomly selected child.
a. Use the data to develop a probability distribution for x. Specify the values for the random variable and the corresponding values for the probability function $f(x)$.
b. Draw a graph of the probability distribution.
c. Show that the probability distribution satisfies equations (5.1) and (5.2).

10. Table 5.4 shows the percent frequency distributions of job satisfaction scores for a sample of information systems (IS) senior executives and IS middle managers. The scores range from a low of 1 (very dissatisfied) to a high of 5 (very satisfied).

TABLE 5.4 PERCENT FREQUENCY DISTRIBUTION OF JOB SATISFACTION SCORES FOR INFORMATION SYSTEMS EXECUTIVES AND MIDDLE MANAGERS

Job Satisfaction Score	IS Senior Executives (%)	IS Middle Managers (%)
1	5	4
2	9	10
3	3	12
4	42	46
5	41	28

a. Develop a probability distribution for the job satisfaction score of a senior executive.
b. Develop a probability distribution for the job satisfaction score of a middle manager.
c. What is the probability a senior executive will report a job satisfaction score of 4 or 5?
d. What is the probability a middle manager is very satisfied?
e. Compare the overall job satisfaction of senior executives and middle managers.

11. A technician services mailing machines at companies in the Phoenix area. Depending on the type of malfunction, the service call can take 1, 2, 3, or 4 hours. The different types of malfunctions occur at about the same frequency.

a. Develop a probability distribution for the duration of a service call.
b. Draw a graph of the probability distribution.
c. Show that your probability distribution satisfies the conditions required for a discrete probability function.
d. What is the probability a service call will take three hours?
e. A service call has just come in, but the type of malfunction is unknown. It is 3:00 P.M. and service technicians usually get off at 5:00 P.M. What is the probability the service technician will have to work overtime to fix the machine today?

12. The nation's two biggest cable providers are Comcast Cable Communications, with 21.5 million subscribers, and Time Warner Cable, with 11.0 million subscribers (*The New York Times 2007 Almanac*). Suppose that management of Time Warner Cable subjectively assessed a probability distribution for x, the number of new subscribers they will obtain over the next year in the state of New York, as follows:

x	$f(x)$
100,000	.10
200,000	.20
300,000	.25
400,000	.30
500,000	.10
600,000	.05

a. Is this probability distribution valid? Explain.
b. What is the probability Time Warner will obtain more than 400,000 new subscribers?
c. What is the probability Time Warner will obtain fewer than 200,000 new subscribers?

13. A psychologist determined that the number of sessions required to obtain the trust of a new patient is either 1, 2, or 3. Let x be a random variable indicating the number of sessions required to gain the patient's trust. The following probability function has been proposed.

$$f(x) = \frac{x}{6} \qquad \text{for } x = 1, 2, \text{ or } 3$$

a. Is this probability function valid? Explain.
b. What is the probability that it takes exactly 2 sessions to gain the patient's trust?
c. What is the probability that it takes at least 2 sessions to gain the patient's trust?

14. The following table is a partial probability distribution for the MRA Company's projected profits (x = profit in \$1000s) for the first year of operation (the negative value denotes a loss).

x	$f(x)$
−100	.10
0	.20
50	.30
100	.25
150	.10
200	

a. What is the proper value for $f(200)$? What is your interpretation of this value?
b. What is the probability that MRA will be profitable?
c. What is the probability that MRA will make at least \$100,000?

5.3 Expected Value and Variance

Expected Value

The **expected value**, or mean, of a random variable is a measure of the central location for the random variable. The formula for the expected value of a discrete random variable x follows.

The expected value is a weighted average of the values the random variable may assume. The weights are the probabilities.

EXPECTED VALUE OF A DISCRETE RANDOM VARIABLE

$$E(x) = \mu = \Sigma x f(x) \tag{5.4}$$

Both the notations $E(x)$ and μ are used to denote the expected value of a random variable.

Equation (5.4) shows that to compute the expected value of a discrete random variable, we must multiply each value of the random variable by the corresponding probability $f(x)$ and then add the resulting products. Using the DiCarlo Motors automobile sales example from Section 5.2, we show the calculation of the expected value for the number of automobiles sold during a day in Table 5.5. The sum of the entries in the $xf(x)$ column shows that the expected value is 1.50 automobiles per day. We therefore know that although sales of 0, 1, 2, 3, 4, or 5 automobiles are possible on any one day, over time DiCarlo can anticipate selling an average of 1.50 automobiles per day. Assuming 30 days of operation during a month, we can use the expected value of 1.50 to forecast average monthly sales of 30(1.50) = 45 automobiles.

The expected value does not have to be a value the random variable can assume.

Variance

Even though the expected value provides the mean value for the random variable, we often need a measure of variability, or dispersion. Just as we used the variance in Chapter 3 to summarize the variability in data, we now use **variance** to summarize the variability in the values of a random variable. The formula for the variance of a discrete random variable follows.

The variance is a weighted average of the squared deviations of a random variable from its mean. The weights are the probabilities.

VARIANCE OF A DISCRETE RANDOM VARIABLE

$$\text{Var}(x) = \sigma^2 = \Sigma(x - \mu)^2 f(x) \tag{5.5}$$

TABLE 5.5 CALCULATION OF THE EXPECTED VALUE FOR THE NUMBER OF AUTOMOBILES SOLD DURING A DAY AT DICARLO MOTORS

x	$f(x)$	$xf(x)$
0	.18	0(.18) = .00
1	.39	1(.39) = .39
2	.24	2(.24) = .48
3	.14	3(.14) = .42
4	.04	4(.04) = .16
5	.01	5(.01) = .05
		1.50

$$E(x) = \mu = \Sigma x f(x)$$

TABLE 5.6 CALCULATION OF THE VARIANCE FOR THE NUMBER OF AUTOMOBILES SOLD DURING A DAY AT DICARLO MOTORS

x	$x - \mu$	$(x - \mu)^2$	$f(x)$	$(x - \mu)^2 f(x)$
0	$0 - 1.50 = -1.50$	2.25	.18	$2.25(.18) = .4050$
1	$1 - 1.50 = -.50$	.25	.39	$.25(.39) = .0975$
2	$2 - 1.50 = .50$	.25	.24	$.25(.24) = .0600$
3	$3 - 1.50 = 1.50$	2.25	.14	$2.25(.14) = .3150$
4	$4 - 1.50 = 2.50$	6.25	.04	$6.25(.04) = .2500$
5	$5 - 1.50 = 3.50$	12.25	.01	$12.25(.01) = .1225$
				1.2500

$$\sigma^2 = \Sigma(x - \mu)^2 f(x)$$

As equation (5.5) shows, an essential part of the variance formula is the deviation, $x - \mu$, which measures how far a particular value of the random variable is from the expected value, or mean, μ. In computing the variance of a random variable, the deviations are squared and then weighted by the corresponding value of the probability function. The sum of these weighted squared deviations for all values of the random variable is referred to as the *variance*. The notations $\text{Var}(x)$ and σ^2 are both used to denote the variance of a random variable.

The calculation of the variance for the probability distribution of the number of automobiles sold during a day at DiCarlo Motors is summarized in Table 5.6. We see that the variance is 1.25. The **standard deviation**, σ, is defined as the positive square root of the variance. Thus, the standard deviation for the number of automobiles sold during a day is

$$\sigma = \sqrt{1.25} = 1.118$$

The standard deviation is measured in the same units as the random variable ($\sigma = 1.118$ automobiles) and therefore is often preferred in describing the variability of a random variable. The variance σ^2 is measured in squared units and is thus more difficult to interpret.

Exercises

Methods

15. The following table provides a probability distribution for the random variable x.

x	$f(x)$
3	.25
6	.50
9	.25

a. Compute $E(x)$, the expected value of x.
b. Compute σ^2, the variance of x.
c. Compute σ, the standard deviation of x.

SELF test

16. The following table provides a probability distribution for the random variable y.

y	$f(y)$
2	.20
4	.30
7	.40
8	.10

a. Compute $E(y)$.
b. Compute $\text{Var}(y)$ and σ.

Applications

17. A volunteer ambulance service handles 0 to 5 service calls on any given day. The probability distribution for the number of service calls is as follows.

Number of Service Calls	Probability	Number of Service Calls	Probability
0	.10	3	.20
1	.15	4	.15
2	.30	5	.10

a. What is the expected number of service calls?
b. What is the variance in the number of service calls? What is the standard deviation?

SELF test

18. The American Housing Survey reported the following data on the number of bedrooms in owner-occupied and renter-occupied houses in central cities (http://www.census.gov, March 31, 2003).

	Number of Houses (1000s)	
Bedrooms	**Renter-Occupied**	**Owner-Occupied**
0	547	23
1	5012	541
2	6100	3832
3	2644	8690
4 or more	557	3783

a. Define a random variable x = number of bedrooms in renter-occupied houses and develop a probability distribution for the random variable. (Let $x = 4$ represent 4 or more bedrooms.)
b. Compute the expected value and variance for the number of bedrooms in renter-occupied houses.
c. Define a random variable y = number of bedrooms in owner-occupied houses and develop a probability distribution for the random variable. (Let $y = 4$ represent 4 or more bedrooms.)
d. Compute the expected value and variance for the number of bedrooms in owner-occupied houses.
e. What observations can you make from a comparison of the number of bedrooms in renter-occupied versus owner-occupied homes?

19. The National Basketball Association (NBA) records a variety of statistics for each team. Two of these statistics are the percentage of field goals made by the team and the percentage of three-point shots made by the team. For a portion of the 2004 season, the shooting records of the 29 teams in the NBA showed the probability of scoring two points by making

a field goal was .44, and the probability of scoring three points by making a three-point shot was .34 (http://www.nba.com, January 3, 2004).

a. What is the expected value of a two-point shot for these teams?
b. What is the expected value of a three-point shot for these teams?
c. If the probability of making a two-point shot is greater than the probability of making a three-point shot, why do coaches allow some players to shoot the three-point shot if they have the opportunity? Use expected value to explain your answer.

20. The probability distribution for damage claims paid by the Newton Automobile Insurance Company on collision insurance follows.

Payment ($)	Probability
0	.85
500	.04
1000	.04
3000	.03
5000	.02
8000	.01
10000	.01

a. Use the expected collision payment to determine the collision insurance premium that would enable the company to break even.
b. The insurance company charges an annual rate of $520 for the collision coverage. What is the expected value of the collision policy for a policyholder? (*Hint:* It is the expected payments from the company minus the cost of coverage.) Why does the policyholder purchase a collision policy with this expected value?

21. The following probability distributions of job satisfaction scores for a sample of information systems (IS) senior executives and IS middle managers range from a low of 1 (very dissatisfied) to a high of 5 (very satisfied).

	Probability	
Job Satisfaction Score	IS Senior Executives	IS Middle Managers
1	.05	.04
2	.09	.10
3	.03	.12
4	.42	.46
5	.41	.28

a. What is the expected value of the job satisfaction score for senior executives?
b. What is the expected value of the job satisfaction score for middle managers?
c. Compute the variance of job satisfaction scores for executives and middle managers.
d. Compute the standard deviation of job satisfaction scores for both probability distributions.
e. Compare the overall job satisfaction of senior executives and middle managers.

22. The demand for a product of Carolina Industries varies greatly from month to month. The probability distribution in the following table, based on the past two years of data, shows the company's monthly demand.

Unit Demand	Probability
300	.20
400	.30
500	.35
600	.15

a. If the company bases monthly orders on the expected value of the monthly demand, what should Carolina's monthly order quantity be for this product?
b. Assume that each unit demanded generates $70 in revenue and that each unit ordered costs $50. How much will the company gain or lose in a month if it places an order based on your answer to part (a) and the actual demand for the item is 300 units?

23. The 2002 New York City Housing and Vacancy Survey showed a total of 59,324 rent-controlled housing units and 236,263 rent-stabilized units built in 1947 or later. For these rental units, the probability distributions for the number of persons living in the unit are given (http://www.census.gov, January 12, 2004).

Number of Persons	Rent-Controlled	Rent-Stabilized
1	.61	.41
2	.27	.30
3	.07	.14
4	.04	.11
5	.01	.03
6	.00	.01

a. What is the expected value of the number of persons living in each type of unit?
b. What is the variance of the number of persons living in each type of unit?
c. Make some comparisons between the number of persons living in rent-controlled units and the number of persons living in rent-stabilized units.

24. The J. R. Ryland Computer Company is considering a plant expansion to enable the company to begin production of a new computer product. The company's president must determine whether to make the expansion a medium- or large-scale project. Demand for the new product is uncertain, which for planning purposes may be low demand, medium demand, or high demand. The probability estimates for demand are .20, .50, and .30, respectively. Letting x and y indicate the annual profit in thousands of dollars, the firm's planners developed the following profit forecasts for the medium- and large-scale expansion projects.

		Medium-Scale Expansion Profit		Large-Scale Expansion Profit	
		x	$f(x)$	y	$f(y)$
Demand	Low	50	.20	0	.20
	Medium	150	.50	100	.50
	High	200	.30	300	.30

a. Compute the expected value for the profit associated with the two expansion alternatives. Which decision is preferred for the objective of maximizing the expected profit?
b. Compute the variance for the profit associated with the two expansion alternatives. Which decision is preferred for the objective of minimizing the risk or uncertainty?

5.4 Binomial Probability Distribution

The binomial probability distribution is a discrete probability distribution that provides many applications. It is associated with a multiple-step experiment that we call the binomial experiment.

A Binomial Experiment

A **binomial experiment** exhibits the following four properties.

PROPERTIES OF A BINOMIAL EXPERIMENT

1. The experiment consists of a sequence of n identical trials.
2. Two outcomes are possible on each trial. We refer to one outcome as a *success* and the other outcome as a *failure.*
3. The probability of a success, denoted by p, does not change from trial to trial. Consequently, the probability of a failure, denoted by $1 - p$, does not change from trial to trial.
4. The trials are independent.

Jakob Bernoulli (1654–1705), the first of the Bernoulli family of Swiss mathematicians, published a treatise on probability that contained the theory of permutations and combinations, as well as the binomial theorem.

If properties 2, 3, and 4 are present, we say the trials are generated by a Bernoulli process. If, in addition, property 1 is present, we say we have a binomial experiment. Figure 5.2 depicts one possible sequence of successes and failures for a binomial experiment involving eight trials.

In a binomial experiment, our interest is in the *number of successes occurring in the n trials*. If we let x denote the number of successes occurring in the n trials, we see that x can assume the values of 0, 1, 2, 3, . . . , n. Because the number of values is finite, x is a *discrete* random variable. The probability distribution associated with this random variable is called the **binomial probability distribution**. For example, consider the experiment of tossing a coin five times and on each toss observing whether the coin lands with a head or a tail on its upward face. Suppose we want to count the number of heads appearing over the five tosses. Does this experiment show the properties of a binomial experiment? What is the random variable of interest? Note that:

1. The experiment consists of five identical trials; each trial involves the tossing of one coin.
2. Two outcomes are possible for each trial: a head or a tail. We can designate head a success and tail a failure.
3. The probability of a head and the probability of a tail are the same for each trial, with $p = .5$ and $1 - p = .5$.
4. The trials or tosses are independent because the outcome on any one trial is not affected by what happens on other trials or tosses.

FIGURE 5.2 ONE POSSIBLE SEQUENCE OF SUCCESSES AND FAILURES FOR AN EIGHT-TRIAL BINOMIAL EXPERIMENT

Property 1: The experiment consists of $n = 8$ identical trials.

Property 2: Each trial results in either success (S) or failure (F).

Trials →	1	2	3	4	5	6	7	8
Outcomes →	S	F	F	S	S	F	S	S

Thus, the properties of a binomial experiment are satisfied. The random variable of interest is x = the number of heads appearing in the five trials. In this case, x can assume the values of 0, 1, 2, 3, 4, or 5.

As another example, consider an insurance salesperson who visits 10 randomly selected families. The outcome associated with each visit is classified as a success if the family purchases an insurance policy and a failure if the family does not. From past experience, the salesperson knows the probability that a randomly selected family will purchase an insurance policy is .10. Checking the properties of a binomial experiment, we observe that:

1. The experiment consists of 10 identical trials; each trial involves contacting one family.
2. Two outcomes are possible on each trial: the family purchases a policy (success) or the family does not purchase a policy (failure).
3. The probabilities of a purchase and a nonpurchase are assumed to be the same for each sales call, with $p = .10$ and $1 - p = .90$.
4. The trials are independent because the families are randomly selected.

Because the four assumptions are satisfied, this example is a binomial experiment. The random variable of interest is the number of sales obtained in contacting the 10 families. In this case, x can assume the values of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10.

Property 3 of the binomial experiment is called the *stationarity assumption* and is sometimes confused with property 4, independence of trials. To see how they differ, consider again the case of the salesperson calling on families to sell insurance policies. If, as the day wore on, the salesperson got tired and lost enthusiasm, the probability of success (selling a policy) might drop to .05, for example, by the tenth call. In such a case, property 3 (stationarity) would not be satisfied, and we would not have a binomial experiment. Even if property 4 held—that is, the purchase decisions of each family were made independently—it would not be a binomial experiment if property 3 was not satisfied.

In applications involving binomial experiments, a special mathematical formula, called the *binomial probability function,* can be used to compute the probability of x successes in the n trials. Using probability concepts introduced in Chapter 4, we will show in the context of an illustrative problem how the formula can be developed.

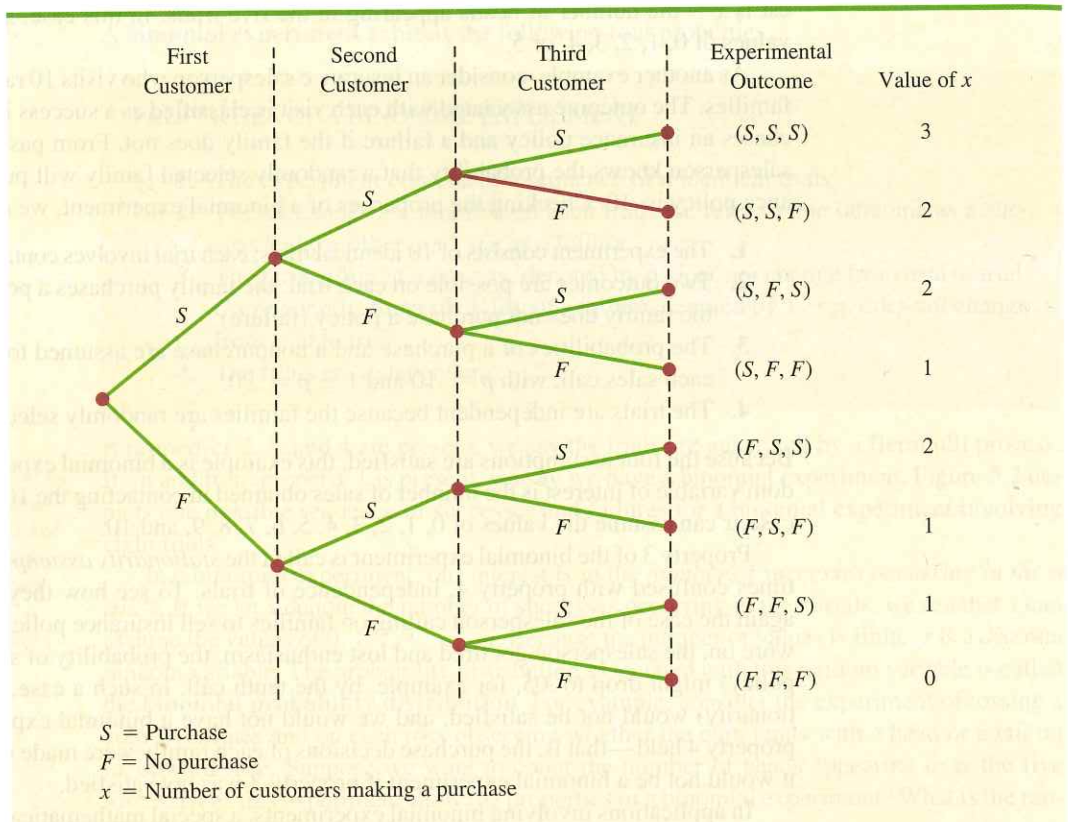
Martin Clothing Store Problem

Let us consider the purchase decisions of the next three customers who enter the Martin Clothing Store. On the basis of past experience, the store manager estimates the probability that any one customer will make a purchase is .30. What is the probability that two of the next three customers will make a purchase?

Using a tree diagram (Figure 5.3), we can see that the experiment of observing the three customers each making a purchase decision has eight possible outcomes. Using S to denote success (a purchase) and F to denote failure (no purchase), we are interested in experimental outcomes involving two successes in the three trials (purchase decisions). Next, let us verify that the experiment involving the sequence of three purchase decisions can be viewed as a binomial experiment. Checking the four requirements for a binomial experiment, we note that:

1. The experiment can be described as a sequence of three identical trials, one trial for each of the three customers who will enter the store.
2. Two outcomes—the customer makes a purchase (success) or the customer does not make a purchase (failure)—are possible for each trial.
3. The probability that the customer will make a purchase (.30) or will not make a purchase (.70) is assumed to be the same for all customers.
4. The purchase decision of each customer is independent of the decisions of the other customers.

FIGURE 5.3 TREE DIAGRAM FOR THE MARTIN CLOTHING STORE PROBLEM



Hence, the properties of a binomial experiment are present.

The number of experimental outcomes resulting in exactly x successes in n trials can be computed using the following formula.*

NUMBER OF EXPERIMENTAL OUTCOMES PROVIDING EXACTLY x SUCCESSES IN n TRIALS

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{5.6}$$

where

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

and, by definition,

$$0! = 1$$

Now let us return to the Martin Clothing Store experiment involving three customer purchase decisions. Equation (5.6) can be used to determine the number of experimental

*This formula, introduced in Chapter 4, determines the number of combinations of n objects selected x at a time. For the binomial experiment, this combinatorial formula provides the number of experimental outcomes (sequences of n trials) resulting in x successes.

outcomes involving two purchases; that is, the number of ways of obtaining $x = 2$ successes in the $n = 3$ trials. From equation (5.6) we have

$$\binom{n}{x} = \binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{(3)(2)(1)}{(2)(1)(1)} = \frac{6}{2} = 3$$

Equation (5.6) shows that three of the experimental outcomes yield two successes. From Figure 5.3 we see these three outcomes are denoted by (S, S, F), (S, F, S), and (F, S, S).

Using equation (5.6) to determine how many experimental outcomes have three successes (purchases) in the three trials, we obtain

$$\binom{n}{x} = \binom{3}{3} = \frac{3!}{3!(3-3)!} = \frac{3!}{3!0!} = \frac{(3)(2)(1)}{3(2)(1)(1)} = \frac{6}{6} = 1$$

From Figure 5.3 we see that the one experimental outcome with three successes is identified by (S, S, S).

We know that equation (5.6) can be used to determine the number of experimental outcomes that result in x successes. If we are to determine the probability of x successes in n trials, however, we must also know the probability associated with each of these experimental outcomes. Because the trials of a binomial experiment are independent, we can simply multiply the probabilities associated with each trial outcome to find the probability of a particular sequence of successes and failures.

The probability of purchases by the first two customers and no purchase by the third customer, denoted (S, S, F), is given by

$$pp(1-p)$$

With a .30 probability of a purchase on any one trial, the probability of a purchase on the first two trials and no purchase on the third is given by

$$(.30)(.30)(.70) = (.30)^2(.70) = .063$$

Two other experimental outcomes also result in two successes and one failure. The probabilities for all three experimental outcomes involving two successes follow.

Trial Outcomes 1st Customer	2nd Customer	3rd Customer	Experimental Outcome	Probability of Experimental Outcome
Purchase	Purchase	No purchase	(S, S, F)	$pp(1-p) = p^2(1-p)$ $= (.30)^2(.70) = .063$
Purchase	No purchase	Purchase	(S, F, S)	$p(1-p)p = p^2(1-p)$ $= (.30)^2(.70) = .063$
No purchase	Purchase	Purchase	(F, S, S)	$(1-p)pp = p^2(1-p)$ $= (.30)^2(.70) = .063$

Observe that all three experimental outcomes with two successes have exactly the same probability. This observation holds in general. In any binomial experiment, all sequences of trial outcomes yielding x successes in n trials have the *same probability* of occurrence. The probability of each sequence of trials yielding x successes in n trials follows.

$$\text{Probability of a particular sequence of trial outcomes with } x \text{ successes in } n \text{ trials} = p^x(1-p)^{(n-x)} \quad \textbf{(5.7)}$$

For the Martin Clothing Store, this formula shows that any experimental outcome with two successes has a probability of $p^2(1-p)^{(3-2)} = p^2(1-p)^1 = (.30)^2(.70)^1 = .063$.

Because equation (5.6) shows the number of outcomes in a binomial experiment with x successes and equation (5.7) gives the probability for each sequence involving x successes, we combine equations (5.6) and (5.7) to obtain the following **binomial probability function**.

BINOMIAL PROBABILITY FUNCTION

$$f(x) = \binom{n}{x} p^x(1-p)^{(n-x)} \quad \textbf{(5.8)}$$

where

$f(x) =$ the probability of x successes in n trials
$n =$ the number of trials
$\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$
$p =$ the probability of a success on any one trial
$1 - p =$ the probability of a failure on any one trial

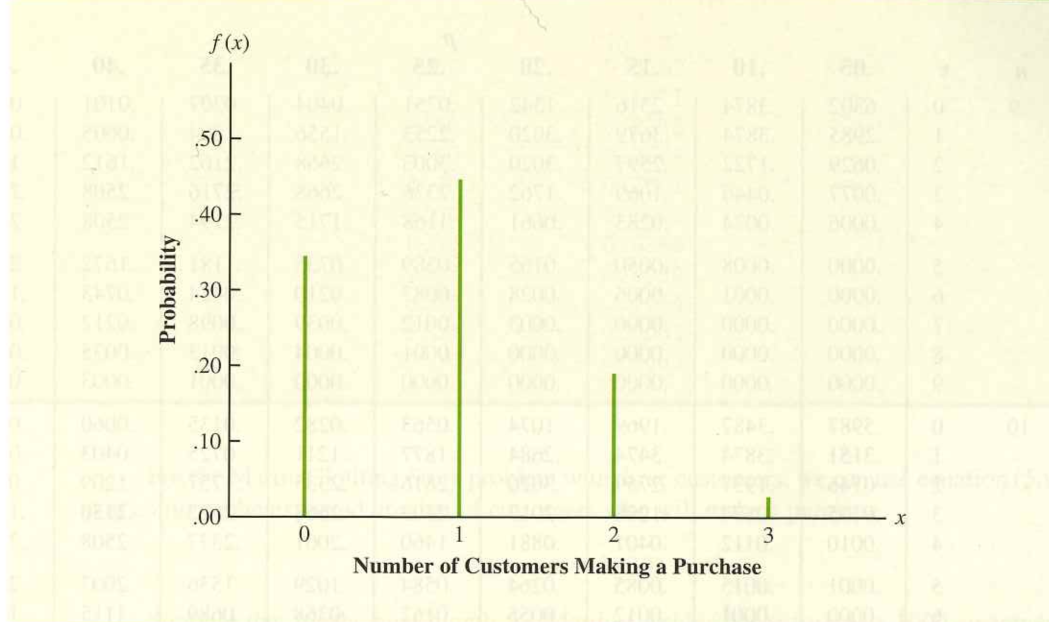
In the Martin Clothing Store example, let us compute the probability that no customer makes a purchase, exactly one customer makes a purchase, exactly two customers make a purchase, and all three customers make a purchase. The calculations are summarized in Table 5.7, which gives the probability distribution of the number of customers making a purchase. Figure 5.4 is a graph of this probability distribution.

The binomial probability function can be applied to *any* binomial experiment. If we are satisfied that a situation demonstrates the properties of a binomial experiment and if we know the values of n and p, we can use equation (5.8) to compute the probability of x successes in the n trials.

TABLE 5.7 PROBABILITY DISTRIBUTION FOR THE NUMBER OF CUSTOMERS MAKING A PURCHASE

x	$f(x)$
0	$\frac{3!}{0!3!}(.30)^0(.70)^3 = .343$
1	$\frac{3!}{1!2!}(.30)^1(.70)^2 = .441$
2	$\frac{3!}{2!1!}(.30)^2(.70)^1 = .189$
3	$\frac{3!}{3!0!}(.30)^3(.70)^0 = .027$
	1.000

FIGURE 5.4 GRAPHICAL REPRESENTATION OF THE PROBABILITY DISTRIBUTION FOR THE NUMBER OF CUSTOMERS MAKING A PURCHASE



If we consider variations of the Martin experiment, such as 10 customers rather than three entering the store, the binomial probability function given by equation (5.8) is still applicable. Suppose we have a binomial experiment with $n = 10$, $x = 4$, and $p = .30$. The probability of making exactly four sales to 10 customers entering the store is

$$f(4) = \frac{10!}{4!6!}(.30)^4(.70)^6 = .2001$$

Using Tables of Binomial Probabilities

Tables have been developed that give the probability of x successes in n trials for a binomial experiment. The tables are generally easy to use and quicker than equation (5.8). Table 5 of Appendix B provides such a table of binomial probabilities. A portion of this table appears in Table 5.8. To use this table, we must specify the values of n, p, and x for the binomial experiment of interest. In the example at the top of Table 5.8, we see that the probability of $x = 3$ successes in a binomial experiment with $n = 10$ and $p = .40$ is .2150. You can use equation (5.8) to verify that you would obtain the same answer if you used the binomial probability function directly.

Now let us use Table 5.8 to verify the probability of four successes in 10 trials for the Martin Clothing Store problem. Note that the value of $f(4) = .2001$ can be read directly from the table of binomial probabilities, with $n = 10$, $x = 4$, and $p = .30$.

With modern calculators, these tables are almost unnecessary. It is easy to evaluate equation (5.8) directly.

Even though the tables of binomial probabilities are relatively easy to use, it is impossible to have tables that show all possible values of n and p that might be encountered in a binomial experiment. However, with today's calculators, using equation (5.8) to calculate the desired probability is not difficult, especially if the number of trials is not large. In the exercises, you should practice using equation (5.8) to compute the binomial probabilities unless the problem specifically requests that you use the binomial probability table.

TABLE 5.8 SELECTED VALUES FROM THE BINOMIAL PROBABILITY TABLE
EXAMPLE: $n = 10, x = 3, p = .40; f(3) = .2150$

n	x	p .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	**.2150**	.1665	.1172
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010

Statistical software packages such as Minitab and spreadsheet packages such as Excel also provide a capability for computing binomial probabilities. Consider the Martin Clothing Store example with $n = 10$ and $p = .30$. Figure 5.5 shows the binomial probabilities generated by Minitab for all possible values of x. Note that these values are the same as those found in the $p = .30$ column of Table 5.8. Appendix 5.1 gives the step-by-step procedure for using Minitab to generate the output in Figure 5.5. Appendix 5.2 describes how Excel can be used to compute binomial probabilities.

Expected Value and Variance for the Binomial Distribution

In Section 5.3 we provided formulas for computing the expected value and variance of a discrete random variable. In the special case where the random variable has a binomial distribution with a known number of trials n and a known probability of success p, the general formulas for the expected value and variance can be simplified. The results follow.

EXPECTED VALUE AND VARIANCE FOR THE BINOMIAL DISTRIBUTION

$$E(x) = \mu = np \tag{5.9}$$

$$\text{Var}(x) = \sigma^2 = np(1 - p) \tag{5.10}$$

FIGURE 5.5 MINITAB OUTPUT SHOWING BINOMIAL PROBABILITIES FOR THE MARTIN CLOTHING STORE PROBLEM

```
    x      P(X = x)
  0.00      0.0282
  1.00      0.1211
  2.00      0.2335
  3.00      0.2668
  4.00      0.2001
  5.00      0.1029
  6.00      0.0368
  7.00      0.0090
  8.00      0.0014
  9.00      0.0001
 10.00      0.0000
```

For the Martin Clothing Store problem with three customers, we can use equation (5.9) to compute the expected number of customers who will make a purchase.

$$E(x) = np = 3(.30) = .9$$

Suppose that for the next month the Martin Clothing Store forecasts 1000 customers will enter the store. What is the expected number of customers who will make a purchase? The answer is $\mu = np = (1000)(.3) = 300$. Thus, to increase the expected number of purchases, Martin's must induce more customers to enter the store and/or somehow increase the probability that any individual customer will make a purchase after entering.

For the Martin Clothing Store problem with three customers, we see that the variance and standard deviation for the number of customers who will make a purchase are

$$\sigma^2 = np(1 - p) = 3(.3)(.7) = .63$$
$$\sigma = \sqrt{.63} = .79$$

For the next 1000 customers entering the store, the variance and standard deviation for the number of customers who will make a purchase are

$$\sigma^2 = np(1 - p) = 1000(.3)(.7) = 210$$
$$\sigma = \sqrt{210} = 14.49$$

NOTES AND COMMENTS

1. The binomial table in Appendix B shows values of p up to and including $p = .95$. Some sources show values of p only up to and including $p = .50$. It would appear that such a table cannot be used when the probability of success exceeds $p = .50$. However, they can be used by noting that the probability of $n - x$ failures is also the probability of x successes. When the probability of success is greater than $p = .50$, one can compute the probability of $n - x$ failures instead. The probability of failure, $1 - p$, will be less than .50 when $p > .50$.
2. Some sources present the binomial table in a cumulative form. In using such a table, one must subtract to find the probability of x successes in n trials. For example, $f(2) = P(x \leq 2) - P(x \leq 1)$. Our table provides these probabilities directly. To compute cumulative probabilities using our table, one simply sums the individual probabilities. For example, to compute $P(x \leq 2)$, we sum $f(0) + f(1) + f(2)$.

Exercises

Methods

SELF test

25. Consider a binomial experiment with two trials and $p = .4$.
    a. Draw a tree diagram for this experiment (see Figure 5.3).
    b. Compute the probability of one success, $f(1)$.
    c. Compute $f(0)$.
    d. Compute $f(2)$.
    e. Compute the probability of at least one success.
    f. Compute the expected value, variance, and standard deviation.
26. Consider a binomial experiment with $n = 10$ and $p = .10$.
    a. Compute $f(0)$.
    b. Compute $f(2)$.
    c. Compute $P(x \leq 2)$.
    d. Compute $P(x \geq 1)$.
    e. Compute $E(x)$.
    f. Compute $\text{Var}(x)$ and σ.
27. Consider a binomial experiment with $n = 20$ and $p = .70$.
    a. Compute $f(12)$.
    b. Compute $f(16)$.
    c. Compute $P(x \geq 16)$.
    d. Compute $P(x \leq 15)$.
    e. Compute $E(x)$.
    f. Compute $\text{Var}(x)$ and σ.

Applications

28. A Harris Interactive survey for InterContinental Hotels & Resorts asked respondents, "When traveling internationally, do you generally venture out on your own to experience culture, or stick with your tour group and itineraries?" The survey found that 23% of the respondents stick with their tour group (*USA Today,* January 21, 2004).
    a. In a sample of six international travelers, what is the probability that two will stick with their tour group?
    b. In a sample of six international travelers, what is the probability that at least two will stick with their tour group?
    c. In a sample of 10 international travelers, what is the probability that none will stick with the tour group?
29. In San Francisco, 30% of workers take public transportation daily (*USA Today,* December 21, 2005).
    a. In a sample of 10 workers, what is the probability that exactly three workers take public transportation daily?
    b. In a sample of 10 workers, what is the probability that at least three workers take public transportation daily?

SELF test

30. When a new machine is functioning properly, only 3% of the items produced are defective. Assume that we will randomly select two parts produced on the machine and that we are interested in the number of defective parts found.
    a. Describe the conditions under which this situation would be a binomial experiment.
    b. Draw a tree diagram similar to Figure 5.3 showing this problem as a two-trial experiment.
    c. How many experimental outcomes result in exactly one defect being found?
    d. Compute the probabilities associated with finding no defects, exactly one defect, and two defects.

31. Nine percent of undergraduate students carry credit card balances greater than \$7000 (*Reader's Digest,* July 2002). Suppose 10 undergraduate students are selected randomly to be interviewed about credit card usage.
    a. Is the selection of 10 students a binomial experiment? Explain.
    b. What is the probability that two of the students will have a credit card balance greater than \$7000?
    c. What is the probability that none will have a credit card balance greater than \$7000?
    d. What is the probability that at least three will have a credit card balance greater than \$7000?
32. Military radar and missile detection systems are designed to warn a country of an enemy attack. A reliability question is whether a detection system will be able to identify an attack and issue a warning. Assume that a particular detection system has a .90 probability of detecting a missile attack. Use the binomial probability distribution to answer the following questions.
    a. What is the probability that a single detection system will detect an attack?
    b. If two detection systems are installed in the same area and operate independently, what is the probability that at least one of the systems will detect the attack?
    c. If three systems are installed, what is the probability that at least one of the systems will detect the attack?
    d. Would you recommend that multiple detection systems be used? Explain.
33. Fifty percent of Americans believed the country was in a recession, even though technically the economy had not shown two straight quarters of negative growth (*BusinessWeek,* July 30, 2001). For a sample of 20 Americans, make the following calculations.
    a. Compute the probability that exactly 12 people believed the country was in a recession.
    b. Compute the probability that no more than five people believed the country was in a recession.
    c. How many people would you expect to say the country was in a recession?
    d. Compute the variance and standard deviation of the number of people who believed the country was in a recession.
34. The Census Bureau's Current Population Survey shows 28% of individuals, ages 25 and older, have completed four years of college (*The New York Times Almanac,* 2006). For a sample of 15 individuals, ages 25 and older, answer the following questions:
    a. What is the probability four will have completed four years of college?
    b. What is the probability three or more will have completed four years of college?
35. A university found that 20% of its students withdraw without completing the introductory statistics course. Assume that 20 students registered for the course.
    a. Compute the probability that two or fewer will withdraw.
    b. Compute the probability that exactly four will withdraw.
    c. Compute the probability that more than three will withdraw.
    d. Compute the expected number of withdrawals.
36. According to a survey conducted by TD Ameritrade, one out of four investors have exchange-traded funds in their portfolios (*USA Today,* January 11, 2007). For a sample of 20 investors, answer the following questions:
    a. Compute the probability that exactly four investors have exchange-traded funds in their portfolio.
    b. Compute the probability that at least two of the investors have exchange-traded funds in their portfolio.
    c. If you found that exactly twelve of the investors have exchange-traded funds in their portfolio, would you doubt the accuracy of the survey results?
    d. Compute the expected number of investors who have exchange-traded funds in their portfolio.
37. Twenty-three percent of automobiles are not covered by insurance (CNN, February 23, 2006). On a particular weekend, 35 automobiles are involved in traffic accidents.
    a. What is the expected number of these automobiles that are not covered by insurance?
    b. What is the variance and standard deviation?

5.5 Poisson Probability Distribution

The Poisson probability distribution is often used to model random arrivals in waiting line situations.

In this section we consider a discrete random variable that is often useful in estimating the number of occurrences over a specified interval of time or space. For example, the random variable of interest might be the number of arrivals at a car wash in one hour, the number of repairs needed in 10 miles of highway, or the number of leaks in 100 miles of pipeline. If the following two properties are satisfied, the number of occurrences is a random variable described by the **Poisson probability distribution**.

PROPERTIES OF A POISSON EXPERIMENT

1. The probability of an occurrence is the same for any two intervals of equal length.
2. The occurrence or nonoccurrence in any interval is independent of the occurrence or nonoccurrence in any other interval.

The **Poisson probability function** is defined by equation (5.11).

Siméon Poisson taught mathematics at the Ecole Polytechnique in Paris from 1802 to 1808. In 1837, he published a work entitled, "Researches on the Probability of Criminal and Civil Verdicts," which includes a discussion of what later became known as the Poisson distribution.

POISSON PROBABILITY FUNCTION

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{5.11}$$

where

$f(x)$ = the probability of x occurrences in an interval
μ = expected value or mean number of occurrences in an interval
e = 2.71828

Before we consider a specific example to see how the Poisson distribution can be applied, note that the number of occurrences, x, has no upper limit. It is a discrete random variable that may assume an infinite sequence of values ($x = 0, 1, 2, \ldots$).

An Example Involving Time Intervals

Bell Labs used the Poisson distribution to model the arrival of phone calls.

Suppose that we are interested in the number of arrivals at the drive-up teller window of a bank during a 15-minute period on weekday mornings. If we can assume that the probability of a car arriving is the same for any two time periods of equal length and that the arrival or nonarrival of a car in any time period is independent of the arrival or nonarrival in any other time period, the Poisson probability function is applicable. Suppose these assumptions are satisfied and an analysis of historical data shows that the average number of cars arriving in a 15-minute period of time is 10; in this case, the following probability function applies.

$$f(x) = \frac{10^x e^{-10}}{x!}$$

The random variable here is x = number of cars arriving in any 15-minute period.

If management wanted to know the probability of exactly five arrivals in 15 minutes, we would set $x = 5$ and thus obtain

$$\text{Probability of exactly 5 arrivals in 15 minutes} = f(5) = \frac{10^5 e^{-10}}{5!} = .0378$$

TABLE 5.9 SELECTED VALUES FROM THE POISSON PROBABILITY TABLES
EXAMPLE: $\mu = 10, x = 5; f(5) = .0378$

	μ									
x	**9.1**	**9.2**	**9.3**	**9.4**	**9.5**	**9.6**	**9.7**	**9.8**	**9.9**	**10**
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0000
1	.0010	.0009	.0009	.0008	.0007	.0007	.0006	.0005	.0005	.0005
2	.0046	.0043	.0040	.0037	.0034	.0031	.0029	.0027	.0025	.0023
3	.0140	.0131	.0123	.0115	.0107	.0100	.0093	.0087	.0081	.0076
4	.0319	.0302	.0285	.0269	.0254	.0240	.0226	.0213	.0201	.0189
5	.0581	.0555	.0530	.0506	.0483	.0460	.0439	.0418	.0398	**.0378**
6	.0881	.0851	.0822	.0793	.0764	.0736	.0709	.0682	.0656	.0631
7	.1145	.1118	.1091	.1064	.1037	.1010	.0982	.0955	.0928	.0901
8	.1302	.1286	.1269	.1251	.1232	.1212	.1191	.1170	.1148	.1126
9	.1317	.1315	.1311	.1306	.1300	.1293	.1284	.1274	.1263	.1251
10	.1198	.1210	.1219	.1228	.1235	.1241	.1245	.1249	.1250	.1251
11	.0991	.1012	.1031	.1049	.1067	.1083	.1098	.1112	.1125	.1137
12	.0752	.0776	.0799	.0822	.0844	.0866	.0888	.0908	.0928	.0948
13	.0526	.0549	.0572	.0594	.0617	.0640	.0662	.0685	.0707	.0729
14	.0342	.0361	.0380	.0399	.0419	.0439	.0459	.0479	.0500	.0521
15	.0208	.0221	.0235	.0250	.0265	.0281	.0297	.0313	.0330	.0347
16	.0118	.0127	.0137	.0147	.0157	.0168	.0180	.0192	.0204	.0217
17	.0063	.0069	.0075	.0081	.0088	.0095	.0103	.0111	.0119	.0128
18	.0032	.0035	.0039	.0042	.0046	.0051	.0055	.0060	.0065	.0071
19	.0015	.0017	.0019	.0021	.0023	.0026	.0028	.0031	.0034	.0037
20	.0007	.0008	.0009	.0010	.0011	.0012	.0014	.0015	.0017	.0019
21	.0003	.0003	.0004	.0004	.0005	.0006	.0006	.0007	.0008	.0009
22	.0001	.0001	.0002	.0002	.0002	.0002	.0003	.0003	.0004	.0004
23	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

Although this probability was determined by evaluating the probability function with $\mu = 10$ and $x = 5$, it is often easier to refer to a table for the Poisson distribution. The table provides probabilities for specific values of x and μ. We included such a table as Table 7 of Appendix B. For convenience, we reproduced a portion of this table as Table 5.9. Note that to use the table of Poisson probabilities, we need know only the values of x and μ. From Table 5.9 we see that the probability of five arrivals in a 15-minute period is found by locating the value in the row of the table corresponding to $x = 5$ and the column of the table corresponding to $\mu = 10$. Hence, we obtain $f(5) = .0378$.

A property of the Poisson distribution is that the mean and variance are equal.

In the preceding example, the mean of the Poisson distribution is $\mu = 10$ arrivals per 15-minute period. A property of the Poisson distribution is that the mean of the distribution and the variance of the distribution are *equal*. Thus, the variance for the number of arrivals during 15-minute periods is $\sigma^2 = 10$. The standard deviation is $\sigma = \sqrt{10} = 3.16$.

Our illustration involves a 15-minute period, but other time periods can be used. Suppose we want to compute the probability of one arrival in a 3-minute period. Because 10 is the expected number of arrivals in a 15-minute period, we see that 10/15 = 2/3 is the expected number of arrivals in a 1-minute period and that (2/3)(3 minutes) = 2 is the expected number of arrivals in a 3-minute period. Thus, the probability of x arrivals in a 3-minute time period with $\mu = 2$ is given by the following Poisson probability function.

$$f(x) = \frac{2^x e^{-2}}{x!}$$

The probability of one arrival in a 3-minute period is calculated as follows:

$$\text{Probability of exactly 1 arrival in 3 minutes} = f(1) = \frac{2^1 e^{-2}}{1!} = .2707$$

Earlier we computed the probability of five arrivals in a 15-minute period; it was .0378. Note that the probability of one arrival in a three-minute period (.2707) is not the same. When computing a Poisson probability for a different time interval, we must first convert the mean arrival rate to the time period of interest and then compute the probability.

An Example Involving Length or Distance Intervals

Let us illustrate an application not involving time intervals in which the Poisson distribution is useful. Suppose we are concerned with the occurrence of major defects in a highway one month after resurfacing. We will assume that the probability of a defect is the same for any two highway intervals of equal length and that the occurrence or nonoccurrence of a defect in any one interval is independent of the occurrence or nonoccurrence of a defect in any other interval. Hence, the Poisson distribution can be applied.

Suppose we learn that major defects one month after resurfacing occur at the average rate of two per mile. Let us find the probability of no major defects in a particular three-mile section of the highway. Because we are interested in an interval with a length of three miles, $\mu = (2 \text{ defects/mile})(3 \text{ miles}) = 6$ represents the expected number of major defects over the three-mile section of highway. Using equation (5.11), the probability of no major defects is $f(0) = 6^0 e^{-6}/0! = .0025$. Thus, it is unlikely that no major defects will occur in the three-mile section. In fact, this example indicates a $1 - .0025 = .9975$ probability of at least one major defect in the three-mile highway section.

Exercises

Methods

38. Consider a Poisson distribution with $\mu = 3$.
    a. Write the appropriate Poisson probability function.
    b. Compute $f(2)$.
    c. Compute $f(1)$.
    d. Compute $P(x \geq 2)$.

SELF test

39. Consider a Poisson distribution with a mean of two occurrences per time period.
    a. Write the appropriate Poisson probability function.
    b. What is the expected number of occurrences in three time periods?
    c. Write the appropriate Poisson probability function to determine the probability of x occurrences in three time periods.
    d. Compute the probability of two occurrences in one time period.
    e. Compute the probability of six occurrences in three time periods.
    f. Compute the probability of five occurrences in two time periods.

Applications

40. Phone calls arrive at the rate of 48 per hour at the reservation desk for Regional Airways.
    a. Compute the probability of receiving three calls in a 5-minute interval of time.
    b. Compute the probability of receiving exactly 10 calls in 15 minutes.
    c. Suppose no calls are currently on hold. If the agent takes 5 minutes to complete the current call, how many callers do you expect to be waiting by that time? What is the probability that none will be waiting?
    d. If no calls are currently being processed, what is the probability that the agent can take 3 minutes for personal time without being interrupted by a call?

41. During the period of time that a local university takes phone-in registrations, calls come in at the rate of one every two minutes.
    a. What is the expected number of calls in one hour?
    b. What is the probability of three calls in five minutes?
    c. What is the probability of no calls in a five-minute period?

SELF test

42. More than 50 million guests stay at bed and breakfasts (B&Bs) each year. The Web site for the Bed and Breakfast Inns of North America (http://www.bestinns.net), which averages approximately seven visitors per minute, enables many B&Bs to attract guests (*Time*, September 2001).
    a. Compute the probability of no Web site visitors in a one-minute period.
    b. Compute the probability of two or more Web site visitors in a one-minute period.
    c. Compute the probability of one or more Web site visitors in a 30-second period.
    d. Compute the probability of five or more Web site visitors in a one-minute period.
43. Airline passengers arrive randomly and independently at the passenger-screening facility at a major international airport. The mean arrival rate is 10 passengers per minute.
    a. Compute the probability of no arrivals in a one-minute period.
    b. Compute the probability that three or fewer passengers arrive in a one-minute period.
    c. Compute the probability of no arrivals in a 15-second period.
    d. Compute the probability of at least one arrival in a 15-second period.
44. An average of 15 aircraft accidents occur each year (*The World Almanac and Book of Facts*, 2004).
    a. Compute the mean number of aircraft accidents per month.
    b. Compute the probability of no accidents during a month.
    c. Compute the probability of exactly one accident during a month.
    d. Compute the probability of more than one accident during a month.
45. The National Safety Council (NSC) estimates that off-the-job accidents cost U.S. businesses almost $200 billion annually in lost productivity (National Safety Council, March 2006). Based on NSC estimates, companies with 50 employees are expected to average three employee off-the-job accidents per year. Answer the following questions for companies with 50 employees.
    a. What is the probability of no off-the-job accidents during a one-year period?
    b. What is the probability of at least two off-the-job accidents during a one-year period?
    c. What is the expected number of off-the-job accidents during six months?
    d. What is the probability of no off-the-job accidents during the next six months?

5.6 Hypergeometric Probability Distribution

The **hypergeometric probability distribution** is closely related to the binomial distribution. The two probability distributions differ in two key ways. With the hypergeometric distribution, the trials are not independent; and the probability of success changes from trial to trial.

In the usual notation for the hypergeometric distribution, r denotes the number of elements in the population of size N labeled success, and $N - r$ denotes the number of elements in the population labeled failure. The **hypergeometric probability function** is used to compute the probability that in a random selection of n elements, selected without replacement, we obtain x elements labeled success and $n - x$ elements labeled failure. For this outcome to occur, we must obtain x successes from the r successes in the population and $n - x$ failures from the $N - r$ failures. The following hypergeometric probability function provides $f(x)$, the probability of obtaining x successes in a sample of size n.

HYPERGEOMETRIC PROBABILITY FUNCTION

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \quad \text{for } 0 \le x \le r \tag{5.12}$$

where

$f(x)$ = probability of x successes in n trials
n = number of trials
N = number of elements in the population
r = number of elements in the population labeled success

Note that $\binom{N}{n}$ represents the number of ways a sample of size n can be selected from a population of size N; $\binom{r}{x}$ represents the number of ways that x successes can be selected from a total of r successes in the population; and $\binom{N-r}{n-x}$ represents the number of ways that $n - x$ failures can be selected from a total of $N - r$ failures in the population.

To illustrate the computations involved in using equation (5.12), let us consider the following quality control application. Electric fuses produced by Ontario Electric are packaged in boxes of 12 units each. Suppose an inspector randomly selects three of the 12 fuses in a box for testing. If the box contains exactly five defective fuses, what is the probability that the inspector will find exactly one of the three fuses defective? In this application, $n = 3$ and $N = 12$. With $r = 5$ defective fuses in the box the probability of finding $x = 1$ defective fuse is

$$f(1) = \frac{\binom{5}{1}\binom{7}{2}}{\binom{12}{3}} = \frac{\left(\frac{5!}{1!4!}\right)\left(\frac{7!}{2!5!}\right)}{\left(\frac{12!}{3!9!}\right)} = \frac{(5)(21)}{220} = .4773$$

Now suppose that we wanted to know the probability of finding *at least* 1 defective fuse. The easiest way to answer this question is to first compute the probability that the inspector does not find any defective fuses. The probability of $x = 0$ is

$$f(0) = \frac{\binom{5}{0}\binom{7}{3}}{\binom{12}{3}} = \frac{\left(\frac{5!}{0!5!}\right)\left(\frac{7!}{3!4!}\right)}{\left(\frac{12!}{3!9!}\right)} = \frac{(1)(35)}{220} = .1591$$

With a probability of zero defective fuses $f(0) = .1591$, we conclude that the probability of finding at least one defective fuse must be $1 - .1591 = .8409$. Thus, there is a reasonably high probability that the inspector will find at least 1 defective fuse.

The mean and variance of a hypergeometric distribution are as follows.

$$E(x) = \mu = n\left(\frac{r}{N}\right) \tag{5.13}$$

$$\text{Var}(x) = \sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right) \tag{5.14}$$

In the preceding example $n = 3$, $r = 5$, and $N = 12$. Thus, the mean and variance for the number of defective fuses is

$$\mu = n\left(\frac{r}{N}\right) = 3\left(\frac{5}{12}\right) = 1.25$$

$$\sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right) = 3\left(\frac{5}{12}\right)\left(1 - \frac{5}{12}\right)\left(\frac{12-3}{12-1}\right) = .60$$

The standard deviation is $\sigma = \sqrt{.60} = .77$.

NOTES AND COMMENTS

Consider a hypergeometric distribution with n trials. Let $p = (r/N)$ denote the probability of a success on the first trial. If the population size is large, the term $(N - n)/(N - 1)$ in equation (5.14) approaches 1. As a result, the expected value and variance can be written $E(x) = np$ and $\text{Var}(x) = np(1 - p)$. Note that these expressions are the same as the expressions used to compute the expected value and variance of a binomial distribution, as in equations (5.9) and (5.10). When the population size is large, a hypergeometric distribution can be approximated by a binomial distribution with n trials and a probability of success $p = (r/N)$.

Exercises

Methods

SELF test

46. Suppose $N = 10$ and $r = 3$. Compute the hypergeometric probabilities for the following values of n and x.
    a. $n = 4, x = 1$.
    b. $n = 2, x = 2$.
    c. $n = 2, x = 0$.
    d. $n = 4, x = 2$.
47. Suppose $N = 15$ and $r = 4$. What is the probability of $x = 3$ for $n = 10$?

Applications

48. In a survey conducted by the Gallup Organization, respondents were asked, "What is your favorite sport to watch?" Football and basketball ranked number one and two in terms of preference (http://www.gallup.com, January 3, 2004). Assume that in a group of 10 individuals, seven preferred football and three preferred basketball. A random sample of three of these individuals is selected.
    a. What is the probability that exactly two preferred football?
    b. What is the probability that the majority (either two or three) preferred football?
49. Blackjack, or twenty-one as it is frequently called, is a popular gambling game played in Las Vegas casinos. A player is dealt two cards. Face cards (jacks, queens, and kings) and tens have a point value of 10. Aces have a point value of 1 or 11. A 52-card deck contains 16 cards with a point value of 10 (jacks, queens, kings, and tens) and four aces.

a. What is the probability that both cards dealt are aces or 10-point cards?
b. What is the probability that both of the cards are aces?
c. What is the probability that both of the cards have a point value of 10?
d. A blackjack is a 10-point card and an ace for a value of 21. Use your answers to parts (a), (b), and (c) to determine the probability that a player is dealt blackjack. (*Hint:* Part (d) is not a hypergeometric problem. Develop your own logical relationship as to how the hypergeometric probabilities from parts (a), (b), and (c) can be combined to answer this question.)

SELF test

50. Axline Computers manufactures personal computers at two plants, one in Texas and the other in Hawaii. The Texas plant has 40 employees; the Hawaii plant has 20. A random sample of 10 employees is to be asked to fill out a benefits questionnaire.
a. What is the probability that none of the employees in the sample work at the plant in Hawaii?
b. What is the probability that one of the employees in the sample works at the plant in Hawaii?
c. What is the probability that two or more of the employees in the sample work at the plant in Hawaii?
d. What is the probability that nine of the employees in the sample work at the plant in Texas?

51. The *2003 Zagat Restaurant Survey* provides food, decor, and service ratings for some of the top restaurants across the United States. For 15 top-ranking restaurants located in Boston, the average price of a dinner, including one drink and tip, was \$48.60. You are leaving for a business trip to Boston and will eat dinner at three of these restaurants. Your company will reimburse you for a maximum of \$50 per dinner. Business associates familiar with these restaurants have told you that the meal cost at one-third of these restaurants will exceed \$50. Suppose that you randomly select three of these restaurants for dinner.
a. What is the probability that none of the meals will exceed the cost covered by your company?
b. What is the probability that one of the meals will exceed the cost covered by your company?
c. What is the probability that two of the meals will exceed the cost covered by your company?
d. What is the probability that all three of the meals will exceed the cost covered by your company?

52. A shipment of 10 items has two defective and eight nondefective items. In the inspection of the shipment, a sample of items will be selected and tested. If a defective item is found, the shipment of 10 items will be rejected.
a. If a sample of three items is selected, what is the probability that the shipment will be rejected?
b. If a sample of four items is selected, what is the probability that the shipment will be rejected?
c. If a sample of five items is selected, what is the probability that the shipment will be rejected?
d. If management would like a .90 probability of rejecting a shipment with two defective and eight nondefective items, how large a sample would you recommend?

Summary

A random variable provides a numerical description of the outcome of an experiment. The probability distribution for a random variable describes how the probabilities are distributed over the values the random variable can assume. For any discrete random variable x, the probability distribution is defined by a probability function, denoted by $f(x)$, which provides the probability associated with each value of the random variable. Once the probability function is defined, we can compute the expected value, variance, and standard deviation for the random variable.

The binomial distribution can be used to determine the probability of x successes in n trials whenever the experiment has the following properties:

1. The experiment consists of a sequence of n identical trials.
2. Two outcomes are possible on each trial, one called success and the other failure.
3. The probability of a success p does not change from trial to trial. Consequently, the probability of failure, $1 - p$, does not change from trial to trial.
4. The trials are independent.

When the four properties hold, the binomial probability function can be used to determine the probability of obtaining x successes in n trials. Formulas were also presented for the mean and variance of the binomial distribution.

The Poisson distribution is used when it is desirable to determine the probability of obtaining x occurrences over an interval of time or space. The following assumptions are necessary for the Poisson distribution to be applicable.

1. The probability of an occurrence of the event is the same for any two intervals of equal length.
2. The occurrence or nonoccurrence of the event in any interval is independent of the occurrence or nonoccurrence of the event in any other interval.

A third discrete probability distribution, the hypergeometric, was introduced in Section 5.6. Like the binomial, it is used to compute the probability of x successes in n trials. But, in contrast to the binomial, the probability of success changes from trial to trial.

Glossary

Random variable A numerical description of the outcome of an experiment.
Discrete random variable A random variable that may assume either a finite number of values or an infinite sequence of values.
Continuous random variable A random variable that may assume any numerical value in an interval or collection of intervals.
Probability distribution A description of how the probabilities are distributed over the values of the random variable.
Probability function A function, denoted by $f(x)$, that provides the probability that x assumes a particular value for a discrete random variable.
Discrete uniform probability distribution A probability distribution for which each possible value of the random variable has the same probability.
Expected value A measure of the central location of a random variable.
Variance A measure of the variability, or dispersion, of a random variable.
Standard deviation The positive square root of the variance.
Binomial experiment An experiment having the four properties stated at the beginning of Section 5.4.
Binomial probability distribution A probability distribution showing the probability of x successes in n trials of a binomial experiment.
Binomial probability function The function used to compute binomial probabilities.
Poisson probability distribution A probability distribution showing the probability of x occurrences of an event over a specified interval of time or space.
Poisson probability function The function used to compute Poisson probabilities.
Hypergeometric probability distribution A probability distribution showing the probability of x successes in n trials from a population with r successes and $N - r$ failures.
Hypergeometric probability function The function used to compute hypergeometric probabilities.

Key Formulas

Discrete Uniform Probability Function

$$f(x) = 1/n \tag{5.3}$$

Expected Value of a Discrete Random Variable

$$E(x) = \mu = \Sigma x f(x) \tag{5.4}$$

Variance of a Discrete Random Variable

$$\text{Var}(x) = \sigma^2 = \Sigma(x - \mu)^2 f(x) \tag{5.5}$$

Number of Experimental Outcomes Providing Exactly x Successes in n Trials

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{5.6}$$

Binomial Probability Function

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} \tag{5.8}$$

Expected Value for the Binomial Distribution

$$E(x) = \mu = np \tag{5.9}$$

Variance for the Binomial Distribution

$$\text{Var}(x) = \sigma^2 = np(1-p) \tag{5.10}$$

Poisson Probability Function

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{5.11}$$

Hypergeometric Probability Function

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \quad \text{for } 0 \le x \le r \tag{5.12}$$

Expected Value for the Hypergeometric Distribution

$$E(x) = \mu = n\left(\frac{r}{N}\right) \tag{5.13}$$

Variance for the Hypergeometric Distribution

$$\text{Var}(x) = \sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right) \tag{5.14}$$

Supplementary Exercises

53. Through the week ending September 16, 2001, Tiger Woods was the leading money winner on the PGA Tour, with total earnings of $5,517,777. Of the top 10 money winners, seven players used a Titleist brand golf ball (http://www.pgatour.com). Suppose that we randomly select two of the top 10 money winners.
    a. What is the probability that exactly one uses a Titleist golf ball?
    b. What is the probability that both use Titleist golf balls?
    c. What is the probability that neither uses a Titleist golf ball?

54. The unemployment rate in the state of Arizona is 4.1% (http://money.cnn.com, May 2, 2007). Assume that 100 employable people in Arizona are selected randomly.
    a. What is the expected number of people who are unemployed?
    b. What are the variance and standard deviation of the number of people who are unemployed?

55. A deck of playing cards contains 52 cards, four of which are aces. What is the probability that the deal of a five-card hand provides:
    a. A pair of aces?
    b. Exactly one ace?
    c. No aces?
    d. At least one ace?

56. Many companies use a quality control technique called acceptance sampling to monitor incoming shipments of parts, raw materials, and so on. In the electronics industry, component parts are commonly shipped from suppliers in large lots. Inspection of a sample of n components can be viewed as the n trials of a binomial experiment. The outcome for each component tested (trial) will be that the component is classified as good or defective. Reynolds Electronics accepts a lot from a particular supplier if the defective components in the lot do not exceed 1%. Suppose a random sample of five items from a recent shipment is tested.
    a. Assume that 1% of the shipment is defective. Compute the probability that no items in the sample are defective.
    b. Assume that 1% of the shipment is defective. Compute the probability that exactly one item in the sample is defective.
    c. What is the probability of observing one or more defective items in the sample if 1% of the shipment is defective?
    d. Would you feel comfortable accepting the shipment if one item was found to be defective? Why or why not?

57. Customer arrivals at a bank are random and independent; the probability of an arrival in any one-minute period is the same as the probability of an arrival in any other one-minute period. Answer the following questions, assuming a mean arrival rate of three customers per minute.
    a. What is the probability of exactly three arrivals in a one-minute period?
    b. What is the probability of at least three arrivals in a one-minute period?

58. A company is planning to interview Internet users to learn how its proposed Web site will be received by different age groups. According to the Census Bureau, 40% of individuals ages 18 to 54 and 12% of individuals age 55 and older use the Internet (*Statistical Abstract of the United States,* 2000).
    a. How many people from the 18–54 age group must be contacted to find an expected number of at least 10 Internet users?
    b. How many people from the age group 55 and older must be contacted to find an expected number of at least 10 Internet users?
    c. If you contact the number of 18- to 54-year-old people suggested in part (a), what is the standard deviation of the number who will be Internet users?

d. If you contact the number of people age 55 and older suggested in part (b), what is the standard deviation of the number who will be Internet users?

59. A regional director responsible for business development in the state of Pennsylvania is concerned about the number of small business failures. If the mean number of small business failures per month is 10, what is the probability that exactly four small businesses will fail during a given month? Assume that the probability of a failure is the same for any two months and that the occurrence or nonoccurrence of a failure in any month is independent of failures in any other month.

60. A survey conducted by the Bureau of Transportation Statistics (BTS) showed that the average commuter spends about 26 minutes on a one-way door-to-door trip from home to work. In addition, 5% of commuters reported a one-way commute of more than one hour (http://www.bts.gov, January 12, 2004).
    a. If 20 commuters are surveyed on a particular day, what is the probability that three will report a one-way commute of more than one hour?
    b. If 20 commuters are surveyed on a particular day, what is the probability that none will report a one-way commute of more than one hour?
    c. If a company has 2000 employees, what is the expected number of employees that have a one-way commute of more than one hour?
    d. If a company has 2000 employees, what is the variance and standard deviation of the number of employees that have a one-way commute of more than one hour?

61. A new automated production process averages 1.5 breakdowns per day. Because of the cost associated with a breakdown, management is concerned about the possibility of having three or more breakdowns during a day. Assume that breakdowns occur randomly, that the probability of a breakdown is the same for any two time intervals of equal length, and that breakdowns in one period are independent of breakdowns in other periods. What is the probability of having three or more breakdowns during a day?

62. The budgeting process for a midwestern college resulted in expense forecasts for the coming year (in \$ millions) of \$9, \$10, \$11, \$12, and \$13. Because the actual expenses are unknown, the following respective probabilities are assigned: .3, .2, .25, .05, and .2.
    a. Show the probability distribution for the expense forecast.
    b. What is the expected value of the expense forecast for the coming year?
    c. What is the variance of the expense forecast for the coming year?
    d. If income projections for the year are estimated at \$12 million, comment on the financial position of the college.

63. Cars arrive at a car wash randomly and independently; the probability of an arrival is the same for any two time intervals of equal length. The mean arrival rate is 15 cars per hour. What is the probability that 20 or more cars will arrive during any given hour of operation?

64. The American Association of Individual Investors publishes an annual guide to the top mutual funds (*The Individual Investor's Guide to the Top Mutual Funds,* 22e, American Association of Individual Investors, 2003). Table 5.10 contains their ratings of the total risk for 29 categories of mutual funds.
    a. Let $x = 1$ for low risk up through $x = 5$ for high risk, and develop a probability distribution for level of risk.
    b. What are the expected value and variance for total risk?
    c. It turns out that 11 of the fund categories were bond funds. For the bond funds, seven categories were rated low and four were rated below average. Compare the total risk of the bond funds with the 18 categories of stock funds.

TABLE 5.10 RISK RATING FOR 29 CATEGORIES OF MUTUAL FUNDS

Total Risk	Number of Fund Categories
Low	7
Below Average	6
Average	3
Above Average	6
High	7

65. A poll conducted by Zogby International showed that of those Americans who said music plays a "very important" role in their lives, 30% said their local radio stations "always" play the kind of music they like (http://www.zogby.com, January 12, 2004). Suppose a sample of 800 people who say music plays an important role in their lives is taken.
    a. How many would you expect to say that their local radio stations always play the kind of music they like?
    b. What is the standard deviation of the number of respondents who think their local radio stations always play the kind of music they like?
    c. What is the standard deviation of the number of respondents who do not think their local radio stations always play the kind of music they like?

66. The *Barron's* Big Money Poll asked 131 investment managers across the United States about their short-term investment outlook (*Barron's,* October 28, 2002). Their responses showed 4% were very bullish, 39% were bullish, 29% were neutral, 21% were bearish, and 7% were very bearish. Let x be the random variable reflecting the level of optimism about the market. Set $x = 5$ for very bullish down through $x = 1$ for very bearish.
    a. Develop a probability distribution for the level of optimism of investment managers.
    b. Compute the expected value for the level of optimism.
    c. Compute the variance and standard deviation for the level of optimism.
    d. Comment on what your results imply about the level of optimism and its variability.

Appendix 5.1 Discrete Probability Distributions with Minitab

Statistical packages such as Minitab offer a relatively easy and efficient procedure for computing binomial probabilities. In this appendix, we show the step-by-step procedure for determining the binomial probabilities for the Martin Clothing Store problem in Section 5.4. Recall that the desired binomial probabilities are based on $n = 10$ and $p = .30$. Before beginning the Minitab routine, the user must enter the desired values of the random variable x into a column of the worksheet. We entered the values 0, 1, 2, . . . , 10 in column 1 (see Figure 5.5) to generate the entire binomial probability distribution. The Minitab steps to obtain the desired binomial probabilities follow.

Step 1. Select the **Calc** menu
Step 2. Choose **Probability Distributions**
Step 3. Choose **Binomial**
Step 4. When the Binomial Distribution dialog box appears:
  Select **Probability**
  Enter 10 in the **Number of trials** box
  Enter .3 in the **Event probability** box
  Enter C1 in the **Input column** box
  Click **OK**

The Minitab output with the binomial probabilities will appear as shown in Figure 5.5.

Minitab provides Poisson and hypergeometric probabilities in a similar manner. For instance, to compute Poisson probabilities the only differences are in step 3, where the **Poisson** option would be selected, and step 4, where the **Mean** would be entered rather than the number of trials and the probability of success.

Appendix 5.2 Discrete Probability Distributions with Excel

Excel provides functions for computing probabilities for the binomial, Poisson, and hypergeometric distributions introduced in this chapter. The Excel function for computing binomial probabilities is BINOMDIST. It has four arguments: x (the number of successes), n (the number of trials), p (the probability of success), and cumulative. FALSE is used for the fourth argument (cumulative) if we want the probability of x successes, and TRUE is used for the fourth argument if we want the cumulative probability of x or fewer successes. Here we show how to compute the probabilities of 0 through 10 successes for the Martin Clothing Store problem in Section 5.4 (see Figure 5.5).

As we describe the worksheet development, refer to Figure 5.6; the formula worksheet is set in the background, and the value worksheet appears in the foreground. We entered

FIGURE 5.6 EXCEL WORKSHEET FOR COMPUTING BINOMIAL PROBABILITIES

	A	B	C	D
1	Number of Trials (n)	10		
2	Probability of Success (p)	0.3		
3				
4		x	$f(x)$	
5		0	=BINOMDIST(B5,B1,B2,FALSE)	
6		1	=BINOMDIST(B6,B1,B2,FALSE)	
7		2	=BINOMDIST(B7,B1,B2,FALSE)	
8		3	=BINOMDIST(B8,B1,B2,FALSE)	
9		4	=BINOMDIST(B9,B1,B2,FALSE)	
10		5	=BINOMDIST(B10,B1,B2,FALSE)	
11		6	=BINOMDIST(B11,B1,B2,FALSE)	
12		7	=BINOMDIST(B12,B1,B2,FALSE)	
13		8	=BINOMDIST(B13,B1,B2,FALSE)	
14		9	=BINOMDIST(B14,B1,B2,FALSE)	
15		10	=BINOMDIST(B15,B1,B2,FALSE)	
16				

	A	B	C	D
1	Number of Trials (n)	10		
2	Probability of Success (p)	0.3		
3				
4		x	$f(x)$	
5		0	0.0282	
6		1	0.1211	
7		2	0.2335	
8		3	0.2668	
9		4	0.2001	
10		5	0.1029	
11		6	0.0368	
12		7	0.0090	
13		8	0.0014	
14		9	0.0001	
15		10	0.0000	
16				

the number of trials (10) into cell B1, the probability of success into cell B2, and the values for the random variable into cells B5:B15. The following steps will generate the desired probabilities:

Step 1. Use the BINOMDIST function to compute the probability of $x = 0$ by entering the following formula into cell C5:

=BINOMDIST(B5,B1,B2,FALSE)

Step 2. Copy the formula in cell C5 into cells C6:C15

The value worksheet in Figure 5.6 shows that the probabilities obtained are the same as in Figure 5.5. Poisson and hypergeometric probabilities can be computed in a similar fashion. The POISSON and HYPGEOMDIST functions are used. Excel's Insert Function dialog box can help the user in entering the proper arguments for these functions (see Appendix E).