Discrete Probability Distributions

CHAPTER 6

Outline

DECISIONS

A woman who was shopping in Los Angeles had her purse stolen by a young, blonde female who was wearing a ponytail. Because there were no eyewitnesses and no real evidence, the prosecution used probability to make its case against the defendant. Your job is to play the role of both the prosecution and defense attorneys to make probabilistic arguments both for and against the defendant. See the Decisions project on page 309.

Putting It All Together

In Chapter 5, we discussed the idea of probability. The probability of an event is the long-term proportion with which the event is observed. That is, if we conduct an experiment 1000 times and observe the outcome 300 times, the probability of the outcome is 0.3. The more times we conduct the experiment, the more accurate the empirical probability. This is the Law of Large Numbers. We learned that we can use counting techniques to obtain theoretical probabilities provided that the outcomes in the experiment are equally likely. This is called classical probability.

We also learned that a probability model lists the possible outcomes to a probability experiment and each outcome's probability. A probability model must satisfy the rules of probability. In particular, all probabilities must be between 0 and 1, inclusive, and the sum of the probabilities must equal 1.

In this chapter, we introduce probability models for *random variables*. A random variable is a numerical measure of the outcome to a probability experiment. So, rather than listing specific outcomes to a probability experiment such as heads or tails, we might list the number of heads obtained in, say, three flips of a coin. In Section 6.1, we discuss random variables and describe the distribution of discrete random variables (shape, center, and spread). Then we discuss a specific discrete probability distribution, *the binomial probability distribution*.

6.1 Discrete Random Variables

Preparing for This Section Before getting started, review the following:

- Discrete versus continuous variables (Section 1.1, pp. 7–9)
- Relative frequency histograms for discrete data (Section 2.2, pp. 72–73)
- Mean (Section 3.1, pp. 107–110)
- Standard deviation (Section 3.2, pp. 129–130)
- Mean from grouped data (Section 3.3, pp. 142–143)
- Standard deviation from grouped data (Section 3.3, pp. 144–146)

Objectives

1. **Distinguish between discrete and continuous random variables**
2. **Identify discrete probability distributions**
3. **Construct probability histograms**
4. **Compute and interpret the mean of a discrete random variable**
5. **Interpret the mean of a discrete random variable as an expected value**
6. **Compute the variance and standard deviation of a discrete random variable**

1 Distinguish between Discrete and Continuous Random Variables

In Chapter 5, we presented the concept of a probability experiment and its outcomes. Suppose we flip a coin two times. The possible outcomes of the experiment are {HH, HT, TH, TT}. Rather than being interested in the outcome, we might be interested in the number of heads. When experiments are conducted in a way such that the outcome is a numerical result, we say the outcome is a *random variable*.

Definition

A **random variable** is a numerical measure of the outcome of a probability experiment, so its value is determined by chance. Random variables are denoted using letters such as X.

So, in our coin-flipping example, if the random variable X represents the number of heads in two flips of a coin, the possible values of X are 0, 1, or 2.

We will follow the practice of using a capital letter to identify the random variable and a small letter to list the possible values of the random variable, that is, the sample space of the experiment. For example, if an experiment is conducted in which a single die is cast, then X represents the number of pips showing on the die and the possible values of X are $x = 1, 2, 3, 4, 5,$ or 6. As another example, suppose an experiment is conducted in which the time between arrivals of cars at a drive-through is measured. The random variable T might describe the time between arrivals, so the sample space of the experiment is $t > 0$.

There are two types of random variables, *discrete* and *continuous*.

Definitions

A **discrete random variable** has either a finite or countable number of values. The values of a discrete random variable can be plotted on a number line with space between each point. See Figure 1(a) on the next page.

A **continuous random variable** has infinitely many values. The values of a continuous random variable can be plotted on a line in an uninterrupted fashion. See Figure 1(b) on the next page.

In Other Words

Discrete random variables typically result from counting, such as 0, 1, 2, 3, and so on. Continuous random variables are variables that result from measurement.

Figure 1

(a) Discrete Random Variable

(b) Continuous Random Variable

EXAMPLE 1 Distinguishing between Discrete and Continuous Random Variables

(a) The number of A's earned in a section of statistics with 15 students enrolled is a discrete random variable because the value of the random variable results from counting. If we let the random variable X represent the number of A's, then the possible values of X are $x = 0, 1, 2, \ldots, 15$.

(b) The number of cars that travel through a McDonald's drive-through in the next hour is a discrete random variable because the value of the random variable results from counting. If we let the random variable X represent the number of cars through the drive-through in the next hour, the possible values of X are $x = 0, 1, 2, \ldots$.

(c) The speed of the next car that passes a state trooper is a continuous random variable because speed is measured. If we let the random variable S represent the speed of the next car, the possible values of S are all positive real numbers; that is, $s > 0$.

CAUTION
Even though a radar gun may report the speed of a car as 37 miles per hour, it is actually any number greater than or equal to 36.5 mph and less than 37.5 mph. That is, $36.5 \le s < 37.5$.

Now Work Problem 7.

In this chapter, we will concentrate on probabilities of discrete random variables. Probabilities for certain continuous random variables will be discussed in the next chapter.

2 Identify Discrete Probability Distributions

Because the value of a random variable is determined by chance, there are probabilities that correspond to the possible values of the random variable.

Definition

The **probability distribution** of a discrete random variable X provides the possible values of the random variable and their corresponding probabilities. A probability distribution can be in the form of a table, graph, or mathematical formula.

EXAMPLE 2 A Discrete Probability Distribution

Table 1

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

Suppose we ask a basketball player to shoot three free throws. Let the random variable X represent the number of shots made, so that $x = 0, 1, 2,$ or 3. Table 1 shows a probability distribution for the random variable X assuming the player historically makes 80% of her free-throw attempts.

From the probability distribution in Table 1, we can see that the probability the player makes all three free-throw attempts is 0.51.

We will denote probabilities using the notation $P(x)$, where x is a specific value of the random variable. We read $P(x)$ as "the probability that the random variable X equals x." For example, $P(3) = 0.51$ is read "the probability that the random variable X equals 3 is 0.51."

Recall from Section 5.1 that probabilities must obey certain rules. We repeat the rules for a discrete probability distribution using the notation just introduced.

In Other Words
The first rule states that the sum of the probabilities must equal 1. The second rule states that each probability must be greater than or equal to 0 and less than or equal to 1.

Rules for a Discrete Probability Distribution

Let $P(x)$ denote the probability that the random variable X equals x; then

1. $\sum P(x) = 1$
2. $0 \leq P(x) \leq 1$

Table 1 from Example 2 is a probability distribution because the sum of the probabilities equals 1 and each probability is between 0 and 1, inclusive. You are encouraged to verify this.

EXAMPLE 3 Identifying Discrete Probability Distributions

Problem: Which of the following is a discrete probability distribution?

(a)

x	$P(x)$
1	0.20
2	0.35
3	0.12
4	0.40
5	−0.07

(b)

x	$P(x)$
1	0.20
2	0.25
3	0.10
4	0.14
5	0.49

(c)

x	$P(x)$
1	0.20
2	0.25
3	0.10
4	0.14
5	0.31

Approach: In a discrete probability distribution, the sum of the probabilities must equal 1, and all probabilities must be greater than or equal to 0 and less than or equal to 1.

Solution

(a) This is not a discrete probability distribution because $P(5) = -0.07$, which is less than 0.

(b) This is not a discrete probability distribution because

$$\sum P(x) = 0.2 + 0.25 + 0.10 + 0.14 + 0.49 = 1.18 \neq 1$$

(c) This is a discrete probability distribution because the sum of the probabilities equals 1, and each probability is greater than or equal to 0 and less than or equal to 1.

Now Work Problem 11.

Table 1 is a discrete probability distribution in table form. Probability distributions can also be represented through graphs or mathematical formulas. We discuss discrete probability distributions using graphs now and discuss probability distributions as mathematical formulas in the next section.

3 Construct Probability Histograms

A graphical depiction of a discrete probability distribution is typically done with a *probability histogram*.

Definition

A **probability histogram** is a histogram in which the horizontal axis corresponds to the value of the random variable and the vertical axis represents the probability of each value of the random variable.

EXAMPLE 4 Constructing a Probability Histogram

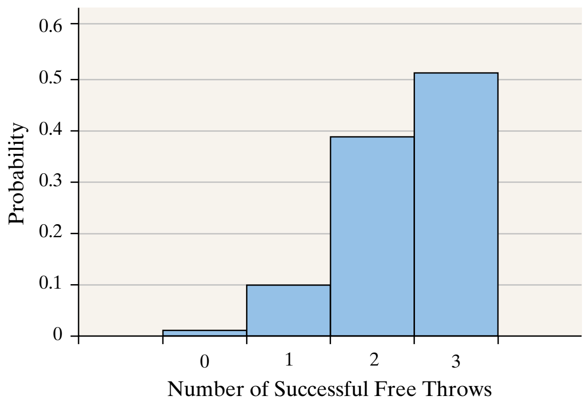
Problem: Construct a probability histogram of the discrete probability distribution given in Table 1 from Example 2.

Approach: Probability histograms are constructed like relative frequency histograms, except that the vertical axis represents the probability of the random variable, rather than its relative frequency. Each rectangle is centered at the value of the discrete random variable.

Solution: Figure 2 presents the probability histogram.

In Other Words
A probability histogram is constructed the same way as a relative frequency histogram for discrete data. The only difference is that the vertical axis is a probability, rather than a relative frequency.

Figure 2



Notice that the area of each rectangle in the probability histogram equals the probability that the random variable assumes the particular value. For example, the area of the rectangle corresponding to the random variable $X = 2$ is $1 \cdot (0.38) = 0.38$, where 1 represents the width of the rectangle and 0.38 represents its height.

Probability histograms help us to determine the shape of the distribution. Recall that we describe distributions as skewed left, skewed right, or symmetric. For example, the probability histogram presented in Figure 2 is skewed left.

Now Work Problems 19(a) and (b).

4 Compute and Interpret the Mean of a Discrete Random Variable

Remember, when we describe the distribution of a variable, we describe its center, spread, and shape. We now introduce methods for identifying the center and spread of a discrete random variable. We will use the mean to describe the center of a random variable. The variance and standard deviation are used to describe the spread of a random variable.

To help see where the formula for computing the mean of a discrete random variable comes from, consider the following. One semester I had a small statistics class of 10 students. I asked them to disclose the number of people living in their household and obtained the following:

$$2, 4, 6, 6, 4, 4, 2, 3, 5, 5$$

What is the mean number of people in the 10 households? Of course, we could find the mean by adding the observations and dividing by 10. But we will take a different approach. Let the random variable X represent the number of people in the household and obtain the probability distribution in Table 2.

Table 2

x	$P(x)$
2	$\frac{2}{10} = 0.2$
3	$\frac{1}{10} = 0.1$
4	$\frac{3}{10} = 0.3$
5	$\frac{2}{10} = 0.2$
6	$\frac{2}{10} = 0.2$

Now we compute the mean as follows:

$$
\begin{aligned}
\mu = \frac{\sum x_i}{N} &= \frac{2+4+6+6+4+4+2+3+5+5}{10} \\
&= \frac{\overbrace{2+2}^{2} + \overbrace{3}^{1} + \overbrace{4+4+4}^{3} + \overbrace{5+5}^{2} + \overbrace{6+6}^{2}}{10} \\
&= \frac{2\cdot 2 + 3\cdot 1 + 4\cdot 3 + 5\cdot 2 + 6\cdot 2}{10} \\
&= 2\cdot\frac{2}{10} + 3\cdot\frac{1}{10} + 4\cdot\frac{3}{10} + 5\cdot\frac{2}{10} + 6\cdot\frac{2}{10} \\
&= 2\cdot P(2) + 3\cdot P(3) + 4\cdot P(4) + 5\cdot P(5) + 6\cdot P(6) \\
&= 2(0.2) + 3(0.1) + 4(0.3) + 5(0.2) + 6(0.2) \\
&= 4.1
\end{aligned}
$$

Based on the preceding computations, we conclude that the mean of a discrete random variable is found by multiplying each possible value of the random variable by its corresponding probability and adding these products.

In Other Words
To find the mean of a discrete random variable, multiply the value of each random variable by its probability. Then add these products.

The Mean of a Discrete Random Variable

The mean of a discrete random variable is given by the formula

$$\mu_X = \sum [x \cdot P(x)] \tag{1}$$

where x is the value of the random variable and $P(x)$ is the probability of observing the random variable x.

EXAMPLE 5 Computing the Mean of a Discrete Random Variable

Problem: Compute the mean of the discrete random variable given in Table 1 from Example 2.

Approach: The mean of a discrete random variable is found by multiplying each value of the random variable by its probability and adding these products.

Solution: Refer to Table 3. The first two columns represent the discrete probability distribution. The third column represents $x \cdot P(x)$.

Table 3

x	$P(x)$	$x \cdot P(x)$
0	0.01	$0 \cdot 0.01 = 0$
1	0.10	$1 \cdot 0.1 = 0.1$
2	0.38	0.76
3	0.51	1.53

We substitute into Formula (1) to find the mean number of free throws made.

$$\mu_X = \sum [x \cdot P(x)] = 0(0.01) + 1(0.10) + 2(0.38) + 3(0.51) = 2.39 \approx 2.4$$

We will follow the practice of rounding the mean, variance, and standard deviation to one more decimal place than the values of the random variable.

How to Interpret the Mean of a Discrete Random Variable

The mean of a discrete random variable can be thought of as the mean outcome of the probability experiment if we repeated the experiment many times. Consider the result of Example 5. If we repeated the experiment of shooting three free throws many times and recorded the number of free throws made, we would expect the average number of free throws made to be around 2.4.

In Other Words
We can think of the mean of a discrete random variable as the average outcome if the experiment is repeated many, many times.

Interpretation of the Mean of a Discrete Random Variable

Suppose an experiment is repeated n independent times and the value of the random variable X is recorded. As the number of repetitions of the experiment, n, increases, the mean value of the n trials will approach μ_X, the mean of the random variable X. In other words, let x_1 be the value of the random variable X after the first experiment, x_2 be the value of the random variable X after the second experiment, and so on. Then

$$\overline{X} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

The difference between $\overline{X}$ and μ_X gets closer to 0 as n increases.

EXAMPLE 6 Illustrating the Interpretation of the Mean of a Discrete Random Variable

Problem: The basketball player from Example 2 is asked to shoot three free throws 100 times. Compute the mean number of free throws made.

Approach: The player shoots three free throws and the number made is recorded. We repeat this experiment 99 more times and then compute the mean number of free throws made.

Solution: The results are presented in Table 4.

Table 4

3	2	3	3	3	3	1	2	3	2
2	3	3	1	2	2	2	2	2	3
3	3	2	2	3	2	3	2	2	2
3	3	2	3	2	3	3	2	3	1
3	2	2	2	2	0	2	3	1	2
3	3	2	3	2	3	2	1	3	2
2	3	3	3	1	3	3	1	3	3
3	2	2	1	3	2	2	2	3	2
3	2	2	2	3	3	2	2	3	3
2	3	2	1	2	3	3	2	3	3

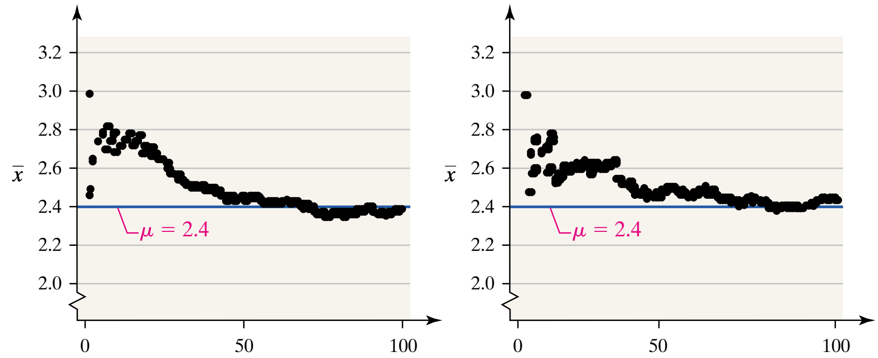
The first time the experiment was conducted, the player made all three free throws. The second time the experiment was conducted, the player made two out of three free throws. The hundredth time the experiment was conducted, the player made three out of three free throws. The mean number of free throws made was

$$\overline{X} = \frac{3 + 2 + 3 + \cdots + 3}{100} = 2.35$$

This is close to the mean of 2.4 (from Example 5). As the number of repetitions of the experiment increases, we expect $\bar{x}$ to get even closer to 2.4.

Figure 3(a) and Figure 3(b) further demonstrate the interpretation of the mean of a discrete random variable. Figure 3(a) shows the mean number of free throws made versus the number of repetitions of the experiment for the data in Table 4. Figure 3(b) shows the mean number of free throws made versus the number of repetitions of the experiment when the same experiment of shooting three free throws 100 times is conducted a second time. In both plots the player starts off "hot," since the mean number of free throws made is above the theoretical level of 2.4. However, both graphs approach the theoretical mean of 2.4 as the number of repetitions of the experiment increases.

Figure 3



(a) Number of Repetitions **(b)** Number of Repetitions

Now Work Problem 19(c).

5 Interpret the Mean of a Discrete Random Variable as an Expected Value

Because the mean of a random variable represents what we would expect to happen in the long run, the mean of a random variable is also called the **expected value**. The interpretation of expected value is the same as the interpretation of the mean of a discrete random variable.

EXAMPLE 7 Finding the Expected Value

In Other Words
The expected value of a discrete random variable is the mean of the discrete random variable.

Problem: A term life insurance policy will pay a beneficiary a certain sum of money upon the death of the policyholder. These policies have premiums that must be paid annually. Suppose a life insurance company sells a \$250,000 one-year term life insurance policy to an 18-year-old male for \$350. According to the *National Vital Statistics Report*, Vol. 47, No. 28, the probability that the male will survive the year is 0.998789. Compute the expected value of this policy to the insurance company.

Approach: There are two possible outcomes to the experiment: survival or death. Let the random variable X represent the *payout* (money lost or gained), depending on survival or death of the insured. We assign probabilities to each of these random variables and substitute these values into Formula (1).

Solution

Step 1: We have $P(\text{survives}) = 0.998789$, so $P(\text{dies}) = 0.001211$. From the point of view of the insurance company, if the client survives the year, the insurance company makes \$350. Therefore, we let $x = \$350$ if the client survives the year. If the client dies during the year, the insurance company must pay \$250,000 to the client's beneficiary. However, the company still keeps the \$350, so we let $x = \$350 - \$250{,}000 = -\$249{,}650$. The value is negative because it is money paid out by the insurance company. The probability distribution is listed in Table 5.

Table 5

x	$P(x)$
\$350 (survives)	0.998789
−\$249,650 (dies)	0.001211

Step 2: Substituting into Formula (1), we obtain the expected value (from the point of view of the insurance company) of the policy.

$$E(X) = \mu_X = \sum xP(x) = \$350(0.998789) + (-\$249{,}650)(0.001211) = \$47.25$$

Interpretation: The company expects to make \$47.25 for each 18-year-old male client it insures. The \$47.25 profit of the insurance company is a long-term result. It does not make \$47.25 on each person it insures, but rather the average profit per person insured is \$47.25. Because this is a long-term result, the insurance "idea" will not work with only a few insured.

Now Work Problem 29.

Historical Note

Christiaan Huygens was born on April 14, 1629, into an influential Dutch family. He studied Law and Mathematics at the University of Leiden from 1645 to 1647. From 1647 to 1649, he continued to study Law and Mathematics at the College of Orange at Breda. Among his many great accomplishments, Huygens discovered the first moon of Saturn in 1655 and the shape of the rings of Saturn in 1656. While in Paris sharing his discoveries, he learned about probability through the correspondence of Fermat and Pascal. In 1657, Huygens published the first book on probability theory. In that text, Huygens introduced the idea of expected value.

In-Class Activity: Expected Value

Consider the following game of chance. A player pays \$1 and rolls a pair of fair dice. If the player rolls a 2, 3, 4, 10, 11, or 12, the player loses the \$1 bet. If the player rolls 5, 6, 8, or 9, there is a "push" and the player gets his or her dollar back. If the player rolls a 7, the player wins \$1.

(a) Construct a probability distribution that describes the game.
(b) Compute the expected value of the game from the player's point of view.
(c) Actually play the game in a small group. Keep track of the results on paper (no money should actually change hands). Compute the mean earnings to the player from the game. Are the results close to what you expected? If not, why?

6 Compute the Variance and Standard Deviation of a Discrete Random Variable

We now introduce a method for computing the variance and standard deviation of a discrete random variable.

In Other Words

The variance of a discrete random variable is a weighted average of the squared deviations where the weights are the probabilities.

Variance and Standard Deviation of a Discrete Random Variable

The variance of a discrete random variable is given by

$$\sigma_X^2 = \sum[(x - \mu_X)^2 \cdot P(x)] \qquad \textbf{(2a)}$$

$$= \sum[x^2 \cdot P(x)] - \mu_X^2 \qquad \textbf{(2b)}$$

where x is the value of the random variable, μ_X is the mean of the random variable, and $P(x)$ is the probability of observing the random variable x.

To find the standard deviation of the discrete random variable, take the square root of the variance. That is, $\sigma_X = \sqrt{\sigma_X^2}$.

EXAMPLE 8 Computing the Variance and Standard Deviation of a Discrete Random Variable

Problem: Find the variance and standard deviation of the discrete random variable given in Table 1 from Example 2.

Approach: We will use Formula (2a) with the unrounded mean $\mu_X = 2.39$.

Solution: Refer to Table 6. The first two columns represent the discrete probability distribution. The third column represents $(x - \mu_X)^2 \cdot P(x)$. We sum the entries in the third column to get the variance.

Table 6

x	$P(x)$	$(x - \mu_X)^2 \cdot P(x)$
0	0.01	$(0 - 2.39)^2 \cdot 0.01 = 0.057121$
1	0.10	$(1 - 2.39)^2 \cdot 0.10 = 0.19321$
2	0.38	$(2 - 2.39)^2 \cdot 0.38 = 0.057798$
3	0.51	$(3 - 2.39)^2 \cdot 0.51 = 0.189771$
		$\sum(x - \mu_X)^2 \cdot P(x) = 0.4979$

The variance of the discrete random variable X is

$$\sigma_X^2 = \sum(x - \mu_X)^2 \cdot P(x) = 0.4979 \approx 0.5$$

Approach: We will use Formula (2b) with the unrounded mean $\mu_X = 2.39$.

Solution: Refer to Table 7. The first two columns represent the discrete probability distribution. The third column represents $x^2 \cdot P(x)$.

Table 7

x	$P(x)$	$x^2 \cdot P(x)$
0	0.01	$0^2 \cdot 0.01 = 0$
1	0.10	$1^2 \cdot 0.10 = 0.10$
2	0.38	$2^2 \cdot 0.38 = 1.52$
3	0.51	$3^2 \cdot 0.51 = 4.59$
		$\sum x^2 \cdot P(x) = 6.21$

The variance of the discrete random variable X is

$$\sigma_X^2 = \sum[x^2 \cdot P(x)] - \mu_X^2 = 6.21 - 2.39^2 = 0.4979 \approx 0.5$$

The standard deviation of the discrete random variable is found by taking the square root of the variance.

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{0.4979} \approx 0.7$$

Now Work Problems 19(d) and (e).

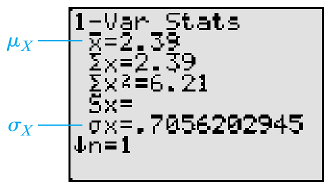
EXAMPLE 9 Obtaining the Mean and Standard Deviation of a Discrete Random Variable Using Technology

Problem: Use a statistical spreadsheet or calculator to determine the mean and the standard deviation of the random variable whose distribution is given in Table 1.

Approach: We will use a TI-84 Plus graphing calculator to obtain the mean and standard deviation. The steps for determining the mean and standard deviation using a TI-83 or TI-84 Plus graphing calculator are given in the Technology Step by Step on page 297.

Result: Figure 4 shows the results from a TI-84 Plus graphing calculator. **Note:** The TI does not find s_X when the sum of L_2 is one.

Figure 4



6.1 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. What is a random variable?
2. What is the difference between a discrete random variable and a continuous random variable? Provide your own examples of each.
3. What are the two requirements for a discrete probability distribution?
4. In your own words, provide an interpretation of the mean of a discrete random variable.
5. Suppose a baseball player historically hits 0.300. (This means that the player averages three hits in every 10 at-bats.) Suppose the player has zero hits in four at-bats in a game and enters the batter's box for the fifth time, whereupon the announcer declares that the player is "due for a hit." What is the flaw in the announcer's reasoning? If the player had four hits in the last four at-bats, is the player "due to make an out"?
6. A game is called a zero-sum game if the expected value of the game is zero. Explain what a game whose expected value is zero means.

Skill Building

In Problems 7–10, determine whether the random variable is discrete or continuous. In each case, state the possible values of the random variable.

7. (a) The number of lightbulbs that burn out in the next week in a room of with 20 bulbs.
NW
   (b) The time it takes to fly from New York City to Los Angeles.
   (c) The number of hits to a Web site in a day.
   (d) The amount of snow in Toronto during the winter.

8. (a) The time it takes for a lightbulb to burn out.
   (b) The weight of a T-bone steak.
   (c) The number of free-throw attempts before the first shot is made.
   (d) In a random sample of 20 people, the number who are blood type A.

9. (a) The amount of rain in Seattle during April.
   (b) The number of fish caught during a fishing tournament.
   (c) The number of customers arriving at a bank between noon and 1:00 P.M.
   (d) The time required to download a file from the Internet.

10. (a) The number of defects in a roll of carpet.
    (b) The distance a baseball travels in the air after being hit.
    (c) The number of points scored during a basketball game.
    (d) The square footage of a house.

In Problems 11–16, determine whether the distribution is a discrete probability distribution. If not, state why.

NW **11.**

x	$P(x)$
0	0.2
1	0.2
2	0.2
3	0.2
4	0.2

12.

x	$P(x)$
0	0.1
1	0.5
2	0.05
3	0.25
4	0.1

13.

x	$P(x)$
10	0.1
20	0.23
30	0.22
40	0.6
50	−0.15

14.

x	$P(x)$
1	0
2	0
3	0
4	0
5	1

15.

x	$P(x)$
100	0.1
200	0.25
300	0.2
400	0.3
500	0.1

16.

x	$P(x)$
100	0.25
200	0.25
300	0.25
400	0.25
500	0.25

In Problems 17 and 18, determine the required value of the missing probability to make the distribution a discrete probability distribution.

17.

x	$P(x)$
3	0.4
4	?
5	0.1
6	0.2

18.

x	$P(x)$
0	0.30
1	0.15
2	?
3	0.20
4	0.15
5	0.05

Applying the Concepts

19. Parental Involvement In the following probability distribution, the random variable X represents the number of activities a parent of a student in kindergarten through fifth grade is involved in.
NW

(a) Verify that this is a discrete probability distribution.
(b) Draw a probability histogram.
(c) Compute and interpret the mean of the random variable X.
(d) Compute the variance of the random variable X.
(e) Compute the standard deviation of the random variable X.
(f) What is the probability that a randomly selected student has a parent involved in three activities?
(g) What is the probability that a randomly selected student has a parent involved in three or four activities?

x	$P(x)$
0	0.035
1	0.074
2	0.197
3	0.320
4	0.374

Source: U.S. National Center for Education Statistics

20. Parental Involvement In the following probability distribution, the random variable X represents the number of activities a parent of a student in grades 6 through 8 is involved in.

(a) Verify that this is a discrete probability distribution.
(b) Draw a probability histogram.
(c) Compute and interpret the mean of the random variable X.
(d) Compute the variance of the random variable X.
(e) Compute the standard deviation of the random variable X.
(f) What is the probability that a randomly selected student has a parent involved in three activities?
(g) What is the probability that a randomly selected student has a parent involved in three or four activities?

x	$P(x)$
0	0.073
1	0.117
2	0.258
3	0.322
4	0.230

Source: U.S. National Center for Education Statistics

21. Ichiro's Hit Parade In the 2004 baseball season, Ichiro Suzuki of the Seattle Mariners set the record for most hits in a season with a total of 262 hits. In the following probability distribution, the random variable X represents the number of hits Ichiro obtained in a game.

x	$P(x)$
0	0.1677
1	0.3354
2	0.2857
3	0.1491
4	0.0373
5	0.0248

Source: *Chicago Tribune*

(a) Verify that this is a discrete probability distribution.
(b) Draw a probability histogram.
(c) Compute and interpret the mean of the random variable X.
(d) Compute the standard deviation of the random variable X.
(e) What is the probability that in a randomly selected game Ichiro got 2 hits?
(f) What is the probability that in a randomly selected game Ichiro got more than 1 hit?

22. Waiting in Line A Wendy's manager performed a study to determine a probability distribution for the number of people waiting in line X during lunch. The results were as follows:

x	$P(x)$	x	$P(x)$	x	$P(x)$
0	0.011	5	0.172	10	0.019
1	0.035	6	0.132	11	0.002
2	0.089	7	0.098	12	0.006
3	0.150	8	0.063	13	0.001
4	0.186	9	0.035	14	0.001

(a) Verify that this is a discrete probability distribution.
(b) Draw a probability histogram.
(c) Compute and interpret the mean of the random variable X.
(d) Compute the variance of the random variable X.
(e) Compute the standard deviation of the random variable X.
(f) What is the probability that there are eight people waiting in line for lunch?
(g) What is the probability that there are 10 or more people waiting in line for lunch? Would this be unusual?

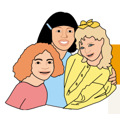
In Problems 23–26, (a) construct a discrete probability distribution for the random variable X [Hint: $P(x_i) = \frac{f_i}{N}$.*], (b) draw the probability histogram, (c) compute and interpret the mean of the random variable X, and (d) compute the standard deviation of the random variable X.*

23. The World Series The following data represent the number of games played in each World Series from 1923 to 2005.

x (games played)	Frequency
4	16
5	15
6	18
7	33

Source: *Information Please Almanac*

24. Number of 5- to 9-Year-Old Girls The following data represent (in thousands) the number of 5- to 9-year-old females in the United States in 2000.



x (age)	Frequency
5	1934
6	1961
7	2008
8	2041
9	2081

Source: U.S. Census Bureau

25. Grade School Enrollment The following data represent (in thousands) the enrollment levels in grades 1 to 8 in the United States in 2000.

x (grade level)	Frequency
1	3635
2	3633
3	3673
4	3708
5	3701
6	3658
7	3624
8	3532

Source: U.S. National Center for Education Statistics

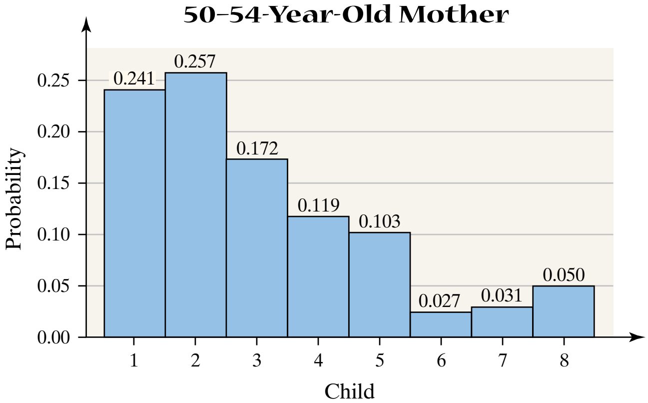
26. High School Enrollment The following data represent (in thousands) the enrollment levels in grades 9 to 12 in the United States in 2000.

x (grade level)	Frequency
9	3958
10	3487
11	3080
12	2799

Source: U.S. National Center for Education Statistics

27. Number of Births The probability histogram that follows represents the number of live births by a mother 50 to 54 years old who had a live birth in 2002. The data are from the *National Vital Statistics Report*, Vol. 52, No. 10, December 17, 2003.

Number of Live Births, 50–54-Year-Old Mother



(a) What is the probability that a randomly selected 50- to 54-year-old mother who had a live birth in 2002 has had her fourth live birth?

(b) What is the probability that a randomly selected 50- to 54-year-old mother who had a live birth in 2002 has had her fourth or fifth live birth?

(c) What is the probability that a randomly selected 50- to 54-year-old mother who had a live birth in 2002 has had her sixth or more live birth?

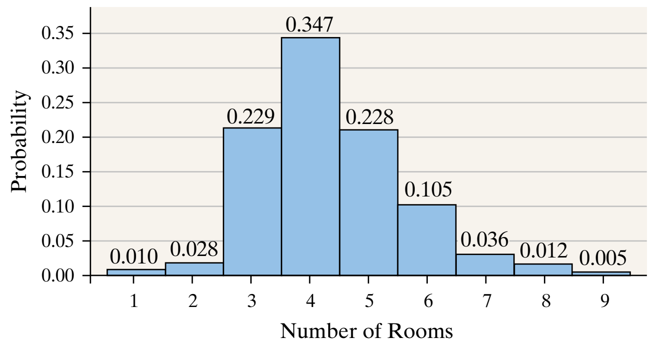
(d) If a 50- to 54-year-old mother who had a live birth in 2002 is randomly selected, how many live births would you expect the mother to have had?

28. Rental Units The probability histogram that follows represents the number of rooms in rented housing units in 2003. The data are from the U.S. Department of Housing and Urban Development.

(a) What is the probability that a randomly selected rental unit has five rooms?

(b) What is the probability that a randomly selected rental unit has five or six rooms?

Number of Rooms in Rental Unit



(c) What is the probability that a randomly selected rental unit has seven or more rooms?

(d) If a rental unit is randomly selected, how many rooms would you expect the unit to have?

29. Life Insurance NW Suppose a life insurance company sells a $250,000 one-year term life insurance policy to a 20-year-old female for $200. According to the *National Vital Statistics Report*, Vol. 53, No 6, the probability that the female survives the year is 0.999546. Compute and interpret the expected value of this policy to the insurance company.

30. Life Insurance Suppose a life insurance company sells a $250,000 one-year term life insurance policy to a 20-year-old male for $350. According to the *National Vital Statistics Report*, Vol. 53, No. 6, the probability that the male survives the year is 0.998611. Compute and interpret the expected value of this policy to the insurance company.

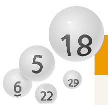
31. Investment An investment counselor calls with a hot stock tip. He believes that if the economy remains strong the investment will result in a profit of $50,000. If the economy grows at a moderate pace, the investment will result in a profit of $10,000. However, if the economy goes into recession, the investment will result in a loss of $50,000. You contact an economist who believes there is a 20% probability the economy will remain strong, a 70% probability the economy will grow at a moderate pace, and a 10% probability the economy will slip into recession. What is the expected profit from this investment?

32. Real Estate Investment Shawn and Maddie purchase a foreclosed property for $50,000 and spend an additional $27,000 fixing up the property. They feel that they can resell the property for $120,000 with probability 0.15, $100,000 with probability 0.45, $80,000 with probability 0.25, and $60,000 with probability 0.15. Compute and interpret the expected profit for reselling the property.

33. Roulette In the game of roulette, a player can place a $5 bet on the number 17 and have a $\frac{1}{38}$ probability of winning. If the metal ball lands on 17, the player wins $175. Otherwise, the casino takes the player's $5. What is the expected value of the game to the player? If you played the game 1000 times, how much would you expect to lose?

34. Connecticut Lottery In the Cash Five Lottery in Connecticut, a player pays \$1 for a single ticket with five numbers. Five Ping-Pong balls numbered 1 through 35 are randomly chosen from a bin without replacement. If all five numbers on a player's ticket match the five chosen, the player wins \$100,000. The probability of this occurring is $\frac{1}{324{,}632}$. If four numbers match, the player wins \$300. This occurs with probability $\frac{1}{2164}$. If three numbers match, the player wins \$10. This occurs with probability $\frac{1}{75}$. Compute and interpret the expected value of the game from the player's point of view.

35. Powerball Powerball is a multistate lottery. The following probability distribution represents the cash prizes of Powerball with their corresponding probabilities.



x (cash prize, \$)	$P(x)$
Grand prize	0.00000000684
200,000	0.00000028
10,000	0.000001711
100	0.000153996
7	0.004778961
4	0.007881463
3	0.01450116
0	0.9726824222

Source: www.powerball.com

(a) If the grand prize is \$15,000,000, find and interpret the expected cash prize. If a ticket costs \$1, what is your expected profit from one ticket?

(b) To the nearest million, how much should the grand prize be so that you can expect a profit? Assume nobody else wins so that you do not have to share the grand prize.

(c) Does the size of the grand prize affect your chance of winning? Explain.

36. SAT Test Penalty Some standardized tests, such as the SAT test, incorporate a penalty for wrong answers. For example, a multiple-choice question with five possible answers will have 1 point awarded for a correct answer and $\frac{1}{4}$ deducted point for an incorrect answer. Questions left blank are worth 0 points.

(a) Find the expected number of points received for a multiple-choice question with five possible answers when a student just guesses.

(b) Explain why there is a deduction for wrong answers.

37. Simulation Use the probability distribution from Problem 21 and a DISCRETE command for some statistical software to simulate 100 repetitions of the experiment (100 games). The number of hits is recorded. Approximate the mean and standard deviation of the random variable X based on the simulation. Repeat the simulation by performing 500 repetitions of the experiment. Approximate the mean and standard deviation of the random variable. Compare your results to the theoretical mean and standard deviation. What property is being illustrated?

38. Simulation Use the probability distribution from Problem 22 and a DISCRETE command for some statistical software to simulate 100 repetitions of the experiment. Approximate the mean and standard deviation of the random variable X based on the simulation. Repeat the simulation by performing 500 repetitions of the experiment. Approximate the mean and standard deviation of the random variable. Compare your results to the theoretical mean and standard deviation. What property is being illustrated?

Technology Step by Step

Finding the Mean and Standard Deviation of a Discrete Random Variable Using Technology

TI-83/84 Plus

***Step 1*:** Enter the values of the random variable in L1 and their corresponding probabilities in L2.

***Step 2*:** Press `STAT`, highlight CALC, and select `1: 1-Var Stats`.

***Step 3*:** With 1-Var Stats on the HOME screen, type L1 followed by a comma, followed by L2 as follows:

```
1-Var Stats L1, L2
```

Hit ENTER.

6.2 The Binomial Probability Distribution

Preparing for This Section Before getting started, review the following:

- Independence (Section 5.3, pp. 249–251)
- Combinations (Section 5.5, pp. 270–273)
- Multiplication Rule for Independent Events (Section 5.3, pp. 251–252)
- Addition Rule for Disjoint Events (Section 5.2, pp. 238–241)
- Complement Rule (Section 5.2, pp. 244–245)
- Empirical Rule (Section 3.2, pp. 131–132)

Objectives

1. **Determine whether a probability experiment is a binomial experiment**
2. **Compute probabilities of binomial experiments**
3. **Compute the mean and standard deviation of a binomial random variable**
4. **Construct binomial probability histograms**

1 Determine Whether a Probability Experiment Is a Binomial Experiment

In Section 6.1, we stated that probability distributions could be presented using tables, graphs, or mathematical formulas. In this section, we introduce a specific type of discrete probability distribution that can be presented using a formula, *the binomial probability distribution*.

The binomial probability distribution is a discrete probability distribution that describes probabilities for experiments in which there are two mutually exclusive (disjoint) outcomes. These two outcomes are generally referred to as *success* and *failure*. For example, a basketball player can either make a free throw (success) or miss (failure). A new surgical procedure can result in either life (success) or death (failure).

In Other Words
The prefix *bi* means "two." This should help remind you that binomial experiments deal with situations in which there are only two outcomes: success and failure.

Experiments in which there are only two possible outcomes are referred to as *binomial experiments*, provided that certain criteria are met.

Criteria for a Binomial Probability Experiment

An experiment is said to be a **binomial experiment** if

1. The experiment is performed a fixed number of times. Each repetition of the experiment is called a **trial**.
2. The trials are independent. This means the outcome of one trial will not affect the outcome of the other trials.
3. For each trial, there are two mutually exclusive (disjoint) outcomes: success or failure.
4. The probability of success is the same for each trial of the experiment.

Let the random variable X be the number of successes in n trials of a binomial experiment. Then X is called a **binomial random variable**. Before introducing the method for computing binomial probabilities, it is worthwhile to introduce some notation.

Notation Used in the Binomial Probability Distribution

- There are n independent trials of the experiment.
- Let p denote the probability of success so that $1 - p$ is the probability of failure.
- Let X denote the number of successes in n independent trials of the experiment. So $0 \le x \le n$.

EXAMPLE 1 Identifying Binomial Experiments

Problem: Determine which of the following probability experiments qualify as a binomial experiment. For those that are binomial experiments, identify the number of trials, probability of success, probability of failure, and possible values of the random variable X.

(a) An experiment in which a basketball player who historically makes 80% of his free throws is asked to shoot three free throws, and the number of made free throws is recorded.

(b) The number of people with blood type O-negative based on a simple random sample of size 10 is recorded. According to the *Information Please Almanac*, 6% of the human population is blood type O-negative.

(c) A probability experiment in which three cards are drawn from a deck without replacement and the number of aces is recorded.

Approach: We determine whether or not the four conditions for a binomial experiment are satisfied.

1. The experiment is performed a fixed number of times.
2. The trials are independent.
3. There are only two possible outcomes of the experiment.
4. The probability of success for each trial is constant.

Solution

(a) This is a binomial experiment because

1. There are $n = 3$ trials.
2. The trials are independent.
3. There are two possible outcomes: make or miss.
4. The probability of success (make) is 0.8 and the probability of failure (miss) is 0.2. The probabilities are the same for each trial.

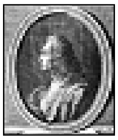
The random variable X is the number of free throws made with $x = 0, 1, 2,$ or 3.

(b) This is a binomial experiment because

1. There are 10 trials (the 10 randomly selected people).
2. The trials are independent.*
3. There are two possible outcomes: finding a person with blood type O-negative or not.
4. The probability of success is 0.06 and the probability of failure is 0.94.

The random variable X is the number of people with blood type O-negative with $x = 0, 1, 2, 3, \dots, 10$.

(c) This is not a binomial experiment because the trials are not independent. The probability of an ace on the first trial is $\frac{4}{52}$. Because we are sampling without replacement, if an ace is selected on the first trial, the probability of an ace on the second trial is $\frac{3}{51}$. If an ace is not selected on the first trial, the probability of an ace on the second trial is $\frac{4}{51}$.



Historical Note

Jacob Bernoulli was born on December 27, 1654, in Basel, Switzerland. He studied philosophy and theology at the urging of his parents. (He resented this.) In 1671, he graduated from the University of Basel with a master's degree in philosophy. In 1676, he received a licentiate in theology. After earning his philosophy degree, Bernoulli traveled to Geneva to tutor. From there, he went to France to study with the great mathematicians of the time. One of Bernoulli's greatest works is *Ars Conjectandi*, published 8 years after his death. In this publication, Bernoulli proved the binomial probability formula. To this day, each observed outcome in a binomial probability experiment is called a *Bernoulli trial*.

Now Work Problem 9.

*In sampling from large populations without replacement, the trials are assumed to be independent, provided that the sample size is small in relation to the size of the population. As a rule of thumb, if the sample size is less than 5% of the population size, the trials are assumed to be independent, although they are technically dependent. See Example 6 in Section 5.4.

CAUTION

The probability of success, p, is always associated with the random variable X, the number of successes. So if X represents the number of 18-year-olds involved in an accident, then p represents the probability of an 18-year-old being involved in an accident.

It is worth mentioning that the word *success* does not necessarily imply that something positive has occurred. Success means that an outcome has occurred that corresponds with p, the probability of success. For example, a probability experiment might be to randomly select ten 18-year-old male drivers. We might let X denote the number who have been involved in an accident within the last year. In this case, a success would mean obtaining an 18-year-old male who was involved in an accident. This outcome is certainly not positive, but still represents a success as far as the experiment goes.

2 Compute Probabilities of Binomial Experiments

We are now prepared to compute probabilities for a binomial random variable X. We present three methods for obtaining binomial probabilities: (1) the binomial probability distribution formula, (2) a table of binomial probabilities, and (3) technology. We develop the binomial probability formula in Example 2.

EXAMPLE 2 Constructing a Binomial Probability Distribution

Problem: According to the *Information Please Almanac*, 6% of the human population is blood type O-negative. A simple random sample of size 4 is obtained, and the number of people X with blood type O-negative is recorded. Construct a probability distribution for the random variable X.

Approach: This is a binomial experiment with $n = 4$ trials. We define a success as selecting an individual with blood type O-negative. The probability of success, p, is 0.06, and X is the random variable representing the number of successes with $x = 0, 1, 2, 3,$ or 4.

***Step 1*:** Construct a tree diagram listing the various outcomes of the experiment by listing each outcome as S (success) or F (failure).

***Step 2*:** Compute the probabilities for each value of the random variable X.

***Step 3*:** Construct the probability distribution.

Solution

***Step 1*:** Figure 5 contains a tree diagram listing the 16 possible outcomes of the experiment.

***Step 2*:** We now compute the probability for each possible value of the random variable X. We start with $P(0)$:

$$\begin{aligned}P(0) &= P(FFFF) = P(F)\cdot P(F)\cdot P(F)\cdot P(F) && \text{Multiplication Rule for Independent Events}\\ &= (0.94)(0.94)(0.94)(0.94)\\ &= (0.94)^4\\ &= 0.78075\end{aligned}$$

$$\begin{aligned}P(1) &= P(SFFF \text{ or } FSFF \text{ or } FFSF \text{ or } FFFS)\\ &= P(SFFF) + P(FSFF) + P(FFSF) + P(FFFS) && \text{Addition Rule for Disjoint Events}\\ &= (0.06)^1(0.94)^3 + (0.06)^1(0.94)^3 + (0.06)^1(0.94)^3 + (0.06)^1(0.94)^3 && \text{Multiplication Rule for Independent Events}\\ &= 4(0.06)^1(0.94)^3\\ &= 0.19934\end{aligned}$$

$$\begin{aligned} P(2) &= P(SSFF \text{ or } SFSF \text{ or } SFFS \text{ or } FSSF \text{ or } FSFS \text{ or } FFSS) \\ &= P(SSFF) + P(SFSF) + P(SFFS) + P(FSSF) + P(FSFS) + P(FFSS) \\ &= (0.06)^2(0.94)^2 + (0.06)^2(0.94)^2 + (0.06)^2(0.94)^2 + (0.06)^2(0.94)^2 + (0.06)^2(0.94)^2 + (0.06)^2(0.94)^2 \\ &= 6(0.06)^2(0.94)^2 \\ &= 0.01909 \end{aligned}$$

Figure 5

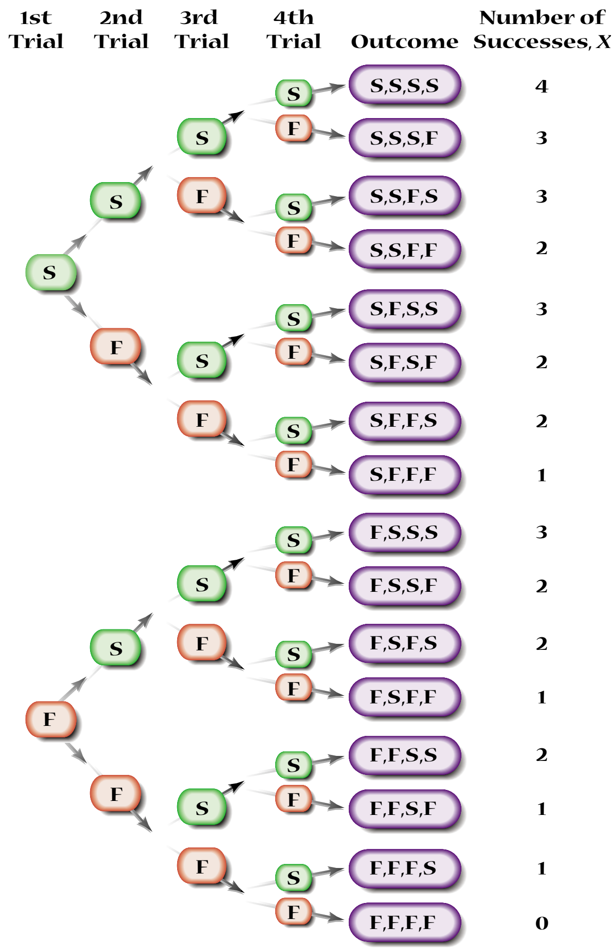


Table 8

x	$P(x)$
0	0.78075
1	0.19934
2	0.01909
3	0.00081
4	0.00001

We compute $P(3)$ and $P(4)$ similarly and obtain $P(3) = 0.00081$ and $P(4) = 0.00001$. You are encouraged to verify these probabilities.

***Step 3*:** We use these results and obtain the probability distribution in Table 8.

As we look back at the solution in Example 2, we note some interesting results. Consider the probability of obtaining $X = 1$ success:

$$P(1) = 4(0.06)^1(0.94)^3$$

"4 is the number of ways we obtain 1 success in 4 trials of the experiment. Here, it is ${}_4C_1$."

"0.06 is the probability of success and the exponent 1 is the number of successes."

"0.94 is the probability of failure and the exponent 3 is the number of failures."

The coefficient 4 is the number of ways of obtaining one success in four trials. In general, the coefficient will be ${}_nC_x$, the number of ways of obtaining x successes in n trials. The second factor in the formula, $(0.06)^1$, is the probability of success, p, raised to the number of successes, x. The third factor in the formula, $(0.94)^3$, is the probability of failure, $1 - p$, raised to the number of failures, $n - x$. This formula holds for all binomial experiments, and we have the *binomial probability distribution function (pdf)*.

CAUTION
Before using the binomial probability distribution function, be sure the requirements for a binomial experiment are satisfied.

Binomial Probability Distribution Function

The probability of obtaining x successes in n independent trials of a binomial experiment, where the probability of success is p, is given by

$$P(x) = {}_nC_x\, p^x(1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n \qquad \textbf{(1)}$$

While reading probability problems, pay special attention to key phrases that translate into mathematical symbols. Table 9 lists various phrases and their corresponding mathematical equivalent.

Table 9

Phrase	Math Symbol
"at least" or "no less than" or "greater than or equal to"	$\geq$
"more than" or "greater than"	$>$
"fewer than" or "less than"	$<$
"no more than" or "at most" or "less than or equal to"	$\leq$
"exactly or "equals" or "is"	$=$

EXAMPLE 3 Using the Binomial Probability Distribution Function

Problem: According to the Federal Communications Commission, 70% of all U.S. households have cable television.

(a) In a random sample of 15 households, what is the probability that exactly 10 have cable?

(b) In a random sample of 15 households, what is the probability that at least 13 have cable?

(c) In a random sample of 15 households, what is the probability that fewer than 13 have cable?

(d) In a random sample of 15 households, what is the probability that the number of households with cable is between 10 and 12, inclusive?

Approach: This is a binomial experiment with $n = 15$ independent trials with the probability of success, p, equal to 0.70. The possible values of the random variable X are $x = 0, 1, 2, \ldots, 15$. We use Formula (1) to compute the probabilities.

Solution

(a)
$$\begin{aligned} P(10) &= {}_{15}C_{10}(0.70)^{10}(1-0.70)^{15-10} && n = 15,\ x = 10,\ p = 0.70 \\ &= \frac{15!}{10!(15-10)!}(0.70)^{10}(0.30)^5 && {}_nC_x = \frac{n!}{x!(n-x)!} \\ &= 3003(0.02825)(0.00243) \\ &= 0.2061 \end{aligned}$$

Interpretation: The probability of getting exactly 10 households out of 15 with cable is 0.2061. In 100 trials of this experiment, we would expect about 21 trials to result in 10 households with cable.

(b) The phrase *at least* means "greater than or equal to." The values of the random variable X greater than or equal to 13 are $x = 13, 14,$ or 15.

$$\begin{aligned} P(X \geq 13) &= P(13 \text{ or } 14 \text{ or } 15) \\ &= P(13) + P(14) + P(15) \qquad \text{Addition Rule for Disjoint Events} \\ &= {}_{15}C_{13}(0.70)^{13}(1-0.70)^{15-13} + {}_{15}C_{14}(0.70)^{14}(1-0.70)^{15-14} + {}_{15}C_{15}(0.70)^{15}(1-0.70)^{15-15} \\ &= 0.0916 + 0.0305 + 0.0047 \\ &= 0.1268 \end{aligned}$$

Interpretation: There is a 0.1268 probability that in a random sample of 15 households at least 13 will have cable. In 100 trials of this experiment, we would expect about 13 trials to result in at least 13 households having cable.

(c) The values of the random variable X less than 13 are $x = 0, 1, 2, \ldots, 12$. Rather than compute $P(X \leq 12)$ directly by computing $P(0) + P(1) + \cdots + P(12)$, we can use the Complement Rule.

$$P(X < 13) = P(X \leq 12) = 1 - P(X \geq 13) = 1 - 0.1268 = 0.8732$$

Interpretation: There is a 0.8732 probability that in a random sample of 15 households, fewer than 13 will have cable. In 100 trials of this experiment, we expect about 87 trials to result in fewer than 13 households that have cable.

(d) The word *inclusive* means "including," so we want to determine the probability that 10, 11, or 12 households have cable.

$$\begin{aligned} P(10 \leq X \leq 12) &= P(10 \text{ or } 11 \text{ or } 12) \\ &= P(10) + P(11) + P(12) \qquad \text{Addition Rule for Disjoint Events} \\ &= {}_{15}C_{10}(0.70)^{10}(1-0.70)^{15-10} + {}_{15}C_{11}(0.70)^{11}(1-0.70)^{15-11} + {}_{15}C_{12}(0.70)^{12}(1-0.70)^{15-12} \\ &= 0.2061 + 0.2186 + 0.1700 \\ &= 0.5947 \end{aligned}$$

Interpretation: The probability that the number of households with cable is between 10 and 12, inclusive, is 0.5947. In 100 trials of this experiment, we expect about 59 trials to result in 10 to 12 households having cable.

Obtaining Binomial Probabilities from Tables

Another method for obtaining probabilities is the binomial probability table. Table II in Appendix A gives probabilities for a binomial random variable X taking on a specific value such as $P(10)$ for select values of n and p. Table III in Appendix A gives cumulative probabilities of a binomial random variable X. This means Table III gives "less than or equal to" binomial probabilities such as $P(X \leq 6)$. We illustrate how to use Tables II and III in Example 4.

EXAMPLE 4 Computing Binomial Probabilities Using the Binomial Table

Problem: According to the National Endowment for the Arts, 20% of U.S. women attended a musical play in 2002.

(a) In a random sample of 15 U.S. women, what is the probability that exactly 5 have attended a musical play in 2002?

(b) In a random sample of 15 U.S. women, what is the probability that fewer than 7 attended a musical play in 2002?

(c) In a random sample of 15 U.S. women, what is the probability that 7 or more attended a musical play in 2002?

Approach: We use Tables II and III in Appendix A to obtain the probabilities.

Solution

(a) We have $n = 15$, $p = 0.20$, and $x = 5$. In Table II, we go to the section that contains $n = 15$ and the column that contains $p = 0.20$. Within the $n = 15$ section, we look for the row $x = 5$. The value at which the $x = 5$ row intersects with the $p = 0.20$ column is the probability we seek. See Figure 6. So $P(5) = 0.1032$.

Figure 6

n	x	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.85	0.90	0.95
15	0	0.8601	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.1303	0.3658	0.3432	0.2312	0.1319	0.0668	0.0305	0.0126	0.0047	0.0015	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.0092	0.1348	0.2669	0.2856	0.2309	0.1559	0.0916	0.0476	0.0219	0.0090	0.0032	0.0010	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.0004	0.0307	0.1285	0.2184	0.2501	0.2252	0.1700	0.1110	0.0634	0.0318	0.0139	0.0052	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.0000	0.0049	0.0428	0.1156	0.1676	0.2252	0.2186	0.1782	0.1268	0.0780	0.0417	0.0191	0.0074	0.0024	0.0006	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.0000	0.0006	0.0105	0.0449	0.1032	0.1651	0.2061	0.2123	0.1859	0.1404	0.0916	0.0515	0.0245	0.0096	0.0030	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
	6	0.0000	0.0000	0.0019	0.0132	0.0430	0.0917	0.1472	0.1906	0.2056	0.1914	0.1527	0.1048	0.0612	0.0298	0.0116	0.0034	0.0007	0.0001	0.0000	0.0000	0.0000
	7	[illegible]	[illegible]	[illegible]	0.9994	[illegible]	[illegible]	[illegible]	[illegible]	0.7869	[illegible]	0.500	[illegible]	[illegible]	[illegible]	0.0500	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Interpretation: There is a 0.1032 probability that in a random sample of 15 U.S. women, exactly 5 have attended a musical play in 2002. In 100 trials of this experiment, we expect about 10 trials to result in exactly 5 women who have attended a musical play in 2002.

(b) The values of the random variable X that are fewer than 7 are 0, 1, 2, 3, 4, 5, or 6. So $P(X < 7) = P(X \leq 6)$. To compute $P(X \leq 6)$, we use the cumulative binomial table, Table III in Appendix A. The cumulative binomial table lists binomial probabilities less than or equal to a specified value. We have $n = 15$ and $p = 0.20$. In Table III, we go to the row that contains $n = 15$ and the column that contains $p = 0.20$. Within the $n = 15$ section, we look for the row $x = 6$. This row represents $P(X \leq 6)$. The value at which the $x = 6$ row intersects with the $p = 0.20$ column is the probability we seek. See Figure 7. So $P(X \leq 6) = 0.9819$.

Figure 7

n	x	0.01	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.85	0.90	0.95
15	0	0.8601	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.9904	0.8290	0.5490	0.3188	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.9996	0.9538	0.8159	0.6042	0.3980	0.2361	0.1288	0.0617	0.0271	0.0107	0.0037	0.0011	0.0003	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	1.0000	0.9945	0.9944	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0178	0.0083	0.0019	0.0005	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	1.0000	0.9994	0.9873	0.9388	0.8358	0.6886	0.5155	0.3519	0.2173	0.1204	0.0592	0.0255	0.0093	0.0028	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	5	1.0000	0.9999	0.9978	0.9964	0.9389	0.8516	0.7216	0.5843	0.4032	0.2608	0.1509	0.0789	0.0338	0.0124	0.0037	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000
	6	1.0000	1.0000	0.9997	0.9964	0.9819	0.9434	0.9689	0.7546	0.6098	0.4522	0.3036	0.1618	0.0950	0.0422	0.0152	0.0042	0.0008	0.0001	0.0000	0.0000	0.0000
	7	1.0000	1.0000	1.0000	0.9994	0.9958	0.9627	0.9500	0.8868	0.7869	0.6535	0.500	0.3465	0.2131	0.1132	0.0500	0.0173	0.0042	0.0006	0.0000	0.0000	0.0000
	8	[illegible]	[illegible]	[illegible]	0.9994	[illegible]	[illegible]	[illegible]	[illegible]	0.7869	[illegible]	0.500	[illegible]	[illegible]	[illegible]	0.0500	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Interpretation: There is a 0.9819 probability that in a random sample of 15 U.S. women, fewer than 7 have attended a musical play in 2002. In 100 trials of this experiment, we would expect about 98 trials to result in fewer than 7 women who have attended a musical play in 2002.

(c) To obtain $P(X \geq 7)$, we use the Complement Rule and the results of part (b) as follows:

$$
\begin{aligned}
P(X \geq 7) &= 1 - P(X < 7) \\
&= 1 - P(X \leq 6) \\
&= 1 - 0.9819 \\
&= 0.0181
\end{aligned}
$$

Interpretation: There is a 0.0181 probability that in a random sample of 15 U.S. women, at least 7 have attended a musical play in 2002. In 100 trials of this experiment, we expect about 2 trials to result in at least 7 women who have attended a musical play in 2002. Because this event only happens about 2 out of 100 times, we consider it to be unusual.

Obtaining Binomial Probabilities Using Technology

Statistical software and graphing calculators have the ability to compute binomial probabilities as well. We illustrate this approach to computing probabilities in the next example.

EXAMPLE 5 Obtaining Binomial Probabilities Using Technology

Problem: According to the National Endowment for the Arts, 20% of U.S. women attended a musical play in 2002.

(a) In a random sample of 15 U.S. women, what is the probability that exactly 5 have attended a musical play in 2002?

(b) In a random sample of 15 U.S. women, what is the probability that fewer than 7 attended a musical play in 2002?

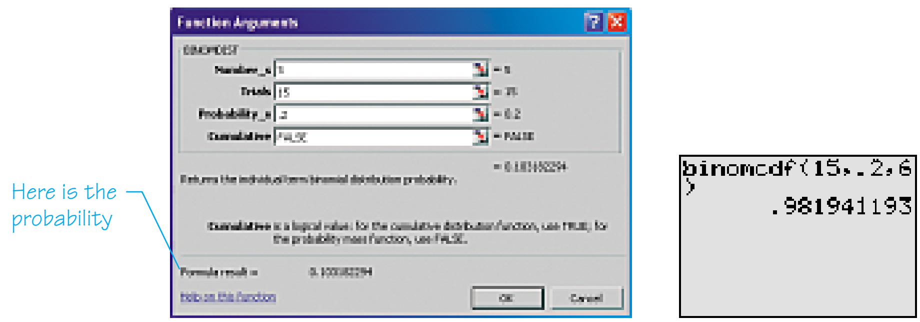
Approach: Statistical software or graphing calculators with advanced statistical features have the ability to determine binomial probabilities. The steps for determining binomial probabilities using MINITAB, Excel, and the TI-83/84 Plus graphing calculators can be found in the Technology Step by Step on page 313.

Result: We will use Excel to determine the probability for part (a) and a TI-84 Plus to determine the probability for part (b).

(a) Using Excel's formula wizard, we obtain the results in Figure 8(a). So $P(5) = 0.1032$. This agrees with the results of Example 4(a).

(b) To compute probabilities such as $P(X < 7) = P(X \leq 6)$, it is best to use the **cumulative distribution function** (or **cdf**), which computes probabilities less than or equal to a specified value. Using a TI-84 Plus graphing calculator to compute $P(X \leq 6)$ with $n = 15$ and $p = 0.2$, we find $P(X \leq 6) = 0.9819$. See Figure 8(b).

Figure 8



(a) (b)

Now Work Problem 35.

3 Compute the Mean and Standard Deviation of a Binomial Random Variable

We discussed finding the mean (or expected value) and standard deviation of a discrete random variable in Section 6.1. These formulas can be used to find the mean (or expected value) and standard deviation of a binomial random variable as well. However, there is a faster method.

In Other Words
The mean of a binomial random variable equals the number of trials of the experiment times the probability of success. It can be interpreted as the expected number of successes in n trials of the experiment.

Mean (or Expected Value) and Standard Deviation of a Binomial Random Variable

A binomial experiment with n independent trials and probability of success p has a mean and standard deviation given by the formulas

$$\mu_X = np \qquad \text{and} \qquad \sigma_X = \sqrt{np(1-p)} \qquad \textbf{(2)}$$

EXAMPLE 6 Finding the Mean and Standard Deviation of a Binomial Random Variable

Problem: According to the Federal Communications Commission, 70% of all U.S. households had cable television in 2002. In a simple random sample of 300 households, determine the mean and standard deviation number of households that will have cable television.

Approach: This is a binomial experiment with $n = 300$ and $p = 0.70$. We can use Formula (2) to find the mean and standard deviation, respectively.

Solution: $\mu_X = np = 300(0.70) = 210$
and

$$\sigma_X = \sqrt{np(1-p)} = \sqrt{300(0.70)(1-0.70)} = \sqrt{63} = 7.9$$

Interpretation: We expect that in a random sample of 300 households 210 will have cable.

Now Work Problems 29(a), (b), and (c).

4 Construct Binomial Probability Histograms

Constructing binomial probability histograms is no different from constructing other probability histograms.

EXAMPLE 7 Constructing Binomial Probability Histograms

Problem

(a) Construct a binomial probability histogram with $n = 10$ and $p = 0.2$. Comment on the shape of the distribution.
(b) Construct a binomial probability histogram with $n = 10$ and $p = 0.5$. Comment on the shape of the distribution.
(c) Construct a binomial probability histogram with $n = 10$ and $p = 0.8$. Comment on the shape of the distribution.

Approach: To construct a binomial probability histogram, we will first obtain the probability distribution. We then construct the probability histogram of the probability distribution.

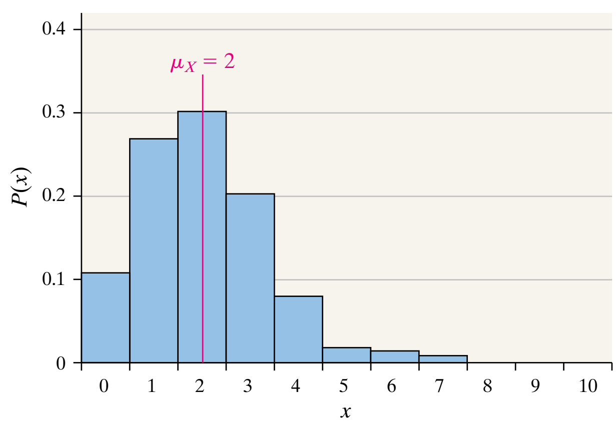
Solution

(a) We obtain the probability distribution with $n = 10$ and $p = 0.2$. See Table 10. Note in Table 10, $P(9) = 0.0000$. The probability is actually 0.000004096 but is written as 0.0000 to four significant digits. The same idea applies to $P(10)$. Figure 9 shows the corresponding probability histogram with the mean $\mu_X = 10(0.2) = 2$ labeled. The distribution is skewed right.

Table 10

x	$P(x)$
0	0.1074
1	0.2684
2	0.3020
3	0.2013
4	0.0881
5	0.0264
6	0.0055
7	0.0008
8	0.0001
9	0.0000
10	0.0000

Figure 9

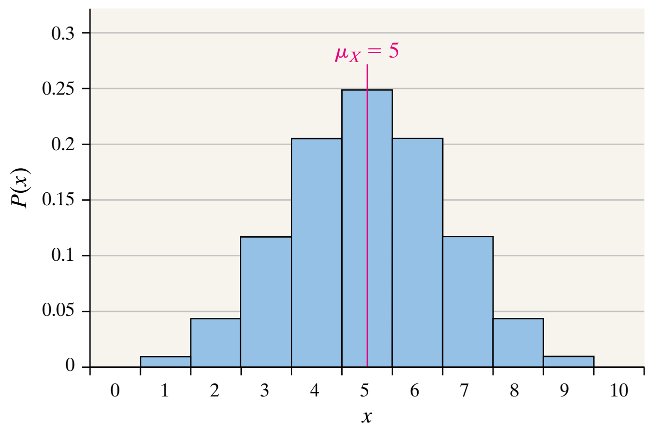


(b) We obtain the probability distribution with $n = 10$ and $p = 0.5$. See Table 11. Figure 10 shows the corresponding probability histogram with the mean $\mu_X = 10(0.5) = 5$ labeled. The distribution is symmetric and approximately bell shaped.

Table 11

x	$P(x)$
0	0.0010
1	0.0098
2	0.0439
3	0.1172
4	0.2051
5	0.2461
6	0.2051
7	0.1172
8	0.0439
9	0.0098
10	0.0010

Figure 10

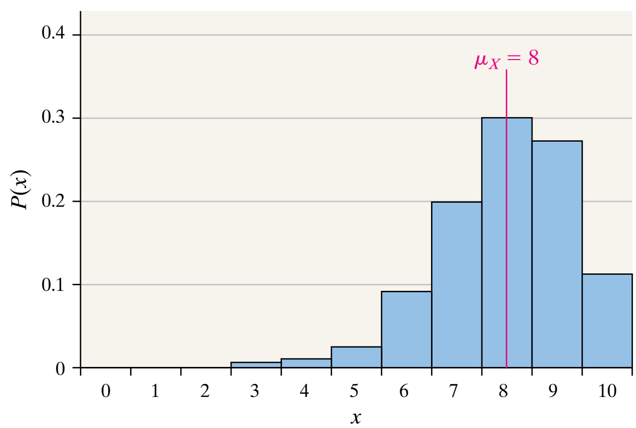


(c) We obtain the probability distribution with $n = 10$ and $p = 0.8$. See Table 12. Figure 11 shows the corresponding probability histogram with the mean $\mu_X = 10(0.8) = 8$ labeled. The distribution is skewed left.

Table 12

x	$P(x)$
0	0.0000
1	0.0000
2	0.0001
3	0.0008
4	0.0055
5	0.0264
6	0.0881
7	0.2013
8	0.3020
9	0.2684
10	0.1074

Figure 11



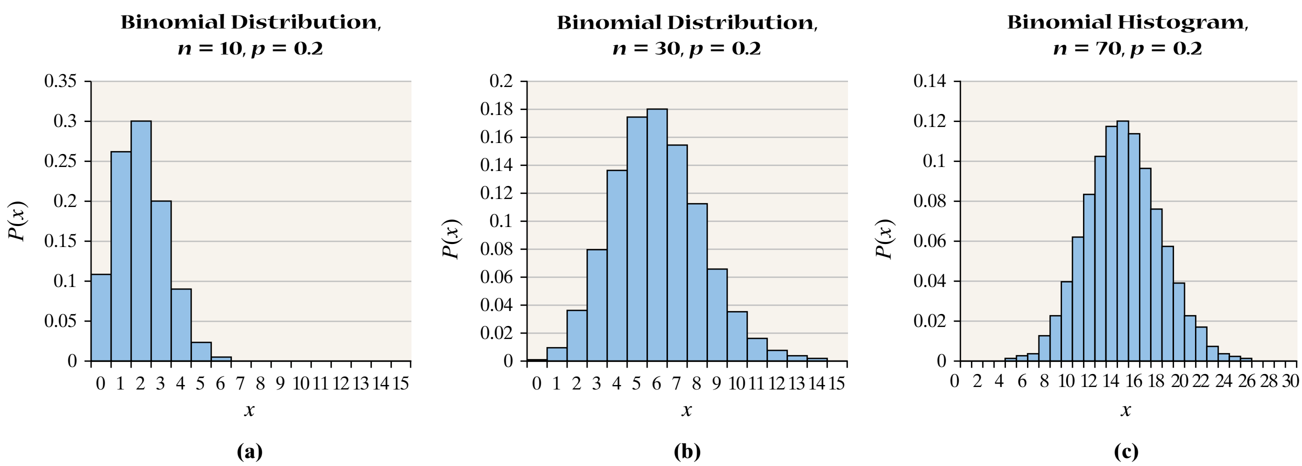
Now Work Problem 29(d).

Based on the results of Example 7, we might conclude that the binomial probability distribution is skewed right if $p < 0.5$, symmetric and approximately bell shaped if $p = 0.5$, and skewed left if $p > 0.5$. Notice that Figure 9 $(p = 0.2)$ and Figure 11 $(p = 0.8)$ are mirror images.

The binomial probability distribution depends on the parameter p, and n, the number of trials. What role does n play in the shape of the distribution? To answer this question we compare the binomial probability histogram with $n = 10$ and $p = 0.2$. [see Figure 12(a)] to the binomial probability histogram with $n = 30$ and $p = 0.2$. [Figure 12(b)] and the binomial probability histogram with $n = 70$ and $p = 0.2$ [Figure 12(c)].

Figure 12(a) is skewed right. Figure 12(b) is slightly skewed right, and Figure 12(c) appears bell shaped.

Figure 12



We conclude the following:

As the number of trials n in a binomial experiment increases, the probability distribution of the random variable X becomes bell shaped. As a rule of thumb, if $np(1 - p) \geq 10$,* the probability distribution will be approximately bell shaped.

This result allows us to use the Empirical Rule to identify unusual observations in a binomial experiment. Recall that the Empirical Rule states that in a bell-shaped distribution about 95% of all observations lie within two standard deviations of the mean. That is, about 95% of the observations lie between $\mu - 2\sigma$ and $\mu + 2\sigma$. Any observation that lies outside this interval may be considered unusual because the observation occurs less than 5% of the time.

EXAMPLE 8 Using the Mean, Standard Deviation, and Empirical Rule to Check for Unusual Results in a Binomial Experiment

Problem: According to the Federal Communications Commission, in 2002, 70% of all U.S. households had cable television. In a simple random sample of 300 households, 230 had cable. Is this result unusual?

Approach: Because $np(1 - p) = 300(0.70)(0.30) = 63 \geq 10$, the binomial probability distribution is approximately bell shaped. Therefore, we can use the Empirical Rule to check for unusual observations. If the observation is less than $\mu - 2\sigma$ or greater than $\mu + 2\sigma$, we say it is unusual.

Solution: From Example 6, we have $\mu = 210$ and $\sigma = 7.9$.

$$\mu - 2\sigma = 210 - 2(7.9) = 210 - 15.8 = 194.2$$

and

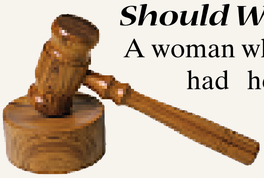
$$\mu + 2\sigma = 210 + 2(7.9) = 210 + 15.8 = 225.8$$

*Ramsey, P. P., and P. H. Ramsey, Evaluating the Normal Approximation to the Binomial Test, *Journal of Educational Statistics* 13 (1998): 173–182.

Interpretation: Since any value less than 194.2 or greater than 225.8 is unusual, 230 is an unusual result. We should attempt to identify reasons for its value. It may be that the percentage of households that have cable has increased since 2002.

Now Work Problem 43.

MAKING AN INFORMED DECISION



Should We Convict?

A woman who was shopping in Los Angeles had her purse stolen by a young, blonde female who was wearing a ponytail. The blonde female got into a yellow car that was driven by a black male who had a mustache and a beard. The police located a blonde female named Janet Collins who wore her hair in a ponytail and had a friend who was a black male who had a mustache and beard and also drove a yellow car. The police arrested the two subjects.

Because there were no eyewitnesses and no real evidence, the prosecution used probability to make its case against the defendants. The following probabilities were presented by the prosecution for the known characteristics of the thieves.

Characteristic	Probability
Yellow car	$\frac{1}{10}$
Man with a mustache	$\frac{1}{4}$
Woman with a ponytail	$\frac{1}{10}$
Woman with blonde hair	$\frac{1}{3}$
Black man with beard	$\frac{1}{10}$
Interracial couple in car	$\frac{1}{1000}$

(a) Assuming that the characteristics listed are independent of each other, what is the probability that a randomly selected couple would have all these characteristics? That is, what is P ("yellow car" and "man with a mustache" and ... and "interracial couple in a car")?

(b) Would you convict the defendants based on this probability? Why or why not?

(c) Now let n represent the number of couples in the Los Angeles area who could have committed the crime. Let p represent the probability a randomly selected couple has all six characteristics listed. Let the random variable X represent the number of couples who have all the characteristics listed in the table. Assuming that the random variable X follows the binomial probability function, we have

$$P(x) = {}_nC_x \cdot p^x(1-p)^{n-x}, \qquad x = 0, 1, 2, \dots, n$$

Assuming that there were $n = 1{,}000{,}000$ couples in the Los Angeles area, what is the probability that more than one of them have the characteristics listed in the table? Does this result cause you to change your mind regarding the defendants' guilt?

(d) Now let's look at this case from a different point of view. We will compute the probability that more than one couple has the characteristics described, given that at least one couple has the characteristics.

$$P(X > 1 \mid X \geq 1) = \frac{P(X > 1 \text{ and } X \geq 1)}{P(X \geq 1)} = \frac{P(X > 1)}{P(X \geq 1)}$$

Conditional Probability Rule

Compute this probability, assuming $n = 1{,}000{,}000$. Compute this probability again, but this time assume that $n = 2{,}000{,}000$. Do you think that the couple should be convicted "beyond all reasonable doubt"? Why or why not?

6.2 ASSESS YOUR UNDERSTANDING

Concepts and Vocabulary

1. State the criteria for a binomial probability experiment.
2. What role does ${}_nC_x$ play in the binomial probability distribution function?
3. How can the Empirical Rule be used to identify unusual results in a binomial experiment? When can the Empirical Rule be used?
4. Describe how the value of n affects the shape of the binomial probability histogram.
5. Describe how the value of p affects the shape of the binomial probability histogram.
6. Explain what "success" means in a binomial experiment.

Skill Building

In Problems 7–16, determine which of the following probability experiments represents a binomial experiment. If the probability experiment is not a binomial experiment, state why.

7. A random sample of 15 college seniors is obtained, and the individuals selected are asked to state their ages.
8. A random sample of 30 cars in a used car lot is obtained, and their mileage recorded.
9. NW An experimental drug is administered to 100 randomly selected individuals, with the number of individuals responding favorably recorded.
10. A poll of 1200 registered voters is conducted in which the respondents are asked whether they believe Congress should reform Social Security.
11. Three cards are selected from a standard 52-card deck without replacement. The number of aces selected is recorded.
12. Three cards are selected from a standard 52-card deck with replacement. The number of aces selected is recorded.
13. A basketball player who makes 80% of her free throws is asked to shoot free throws until she misses. The number of free-throw attempts is recorded.
14. A baseball player who reaches base safely 30% of the time is allowed to bat until he reaches base safely for the third time. The number of at-bats required is recorded.
15. An investor randomly purchases 10 stocks listed on the New York Stock Exchange. Historically, the probability that a stock listed on the NYSE will increase in value over the course of a year is 48%. The number of stocks that increase in value is recorded.
16. According to Nielsen Media Research, 70% of all U.S. households have cable television. In a small town of 40 households, a random sample of 10 households is asked whether they have cable television. The number of households with cable television is recorded.

In Problems 17–28, a binomial probability experiment is conducted with the given parameters. Compute the probability of x successes in the n independent trials of the experiment.

17. $n = 10, p = 0.4, x = 3$
18. $n = 15, p = 0.85, x = 12$
19. $n = 40, p = 0.99, x = 38$
20. $n = 50, p = 0.02, x = 3$
21. $n = 8, p = 0.35, x = 3$
22. $n = 20, p = 0.6, x = 17$
23. $n = 9, p = 0.2, x \leq 3$
24. $n = 10, p = 0.65, x < 5$
25. $n = 7, p = 0.5, x > 3$
26. $n = 20, p = 0.7, x \geq 12$
27. $n = 12, p = 0.35, x \leq 4$
28. $n = 11, p = 0.75, x \geq 8$

In Problems 29–34, (a) construct a binomial probability distribution with the given parameters; (b) compute the mean and standard deviation of the random variable using the methods of Section 6.1; (c) compute the mean and standard deviation, using the methods of this section; and (d) draw the probability histogram, comment on its shape, and label the mean on the histogram.

29. NW $n = 6, p = 0.3$
30. $n = 8, p = 0.5$
31. $n = 9, p = 0.75$
32. $n = 10, p = 0.2$
33. $n = 10, p = 0.5$
34. $n = 9, p = 0.8$

Applying the Concepts

35. On-Time Flights According to American Airlines, its NW flight 215 from Orlando to Los Angeles is on time 90% of the time. Suppose 15 flights are randomly selected and the number of on-time flights is recorded.

(a) Explain why this is a binomial experiment.
(b) Find the probability that exactly 14 flights are on time.
(c) Find the probability that at least 14 flights are on time.
(d) Find the probability that fewer than 14 flights are on time.
(e) Find the probability that between 12 and 14 flights, inclusive, are on time.

36. Smokers According to the *Information Please Almanac*, 80% of adult smokers started smoking before turning 18 years old. Suppose 10 smokers 18 years old or older are randomly selected and the number of smokers who started smoking before 18 is recorded.

(a) Explain why this is a binomial experiment.
(b) Find the probability that exactly 8 of them started smoking before 18 years of age.
(c) Find the probability that at least 8 of them started smoking before 18 years of age.
(d) Find the probability that fewer than 8 of them started smoking before 18 years of age.
(e) Find the probability that between 7 and 9 of them, inclusive, started smoking before 18 years of age.

37. High-Speed Internet According to a report by the Commerce Department in the fall of 2004, 20% of U.S. households had some type of high-speed Internet connection. Suppose 20 U.S. households are selected at random and the number of households with high-speed Internet is recorded.

(a) Find the probability that exactly 5 households have high-speed Internet.
(b) Find the probability that at least 10 households have high-speed Internet. Would this be unusual?
(c) Find the probability that fewer than 4 households have high-speed Internet.
(d) Find the probability that between 2 and 5 households, inclusive, have high-speed Internet.

38. Allergy Sufferers Clarinex-D is a medication whose purpose is to reduce the symptoms associated with a variety of allergies. In clinical trials of Clarinex-D, 5% of the patients in the study experienced insomnia as a side effect. Suppose a random sample of 20 Clarinex-D users is obtained and the number of patients who experienced insomnia is recorded.

(a) Find the probability that exactly 3 experienced insomnia as a side effect.
(b) Find the probability that 3 or fewer experienced insomnia as a side effect.
(c) Find the probability that between 1 and 4 patients, inclusive, experienced insomnia as a side effect.
(d) Would it be unusual to find 4 or more patients who experienced insomnia as a side effect? Why?

39. Murder by Firearms According to the *Uniform Crime Report, 2003*, 66.9% of murders are committed with a firearm. Suppose that 25 murders were randomly selected and the number of murders committed with a firearm is recorded.

(a) Find the probability that exactly 22 murders were committed using a firearm.
(b) Find the probability that between 14 and 16 murders, inclusive, were committed using a firearm.
(c) Would it be unusual if 22 or more murders were committed using a firearm? Why?

40. Migraine Sufferers Depakote is a medication whose purpose is to reduce the pain associated with migraine headaches. In clinical trials of Depakote, 2% of the patients in the study experienced weight gain as a side effect. Suppose a random sample of 30 Depakote users is obtained and the number of patients who experienced weight gain is recorded.

(*Source*:Abbott Laboratories)

(a) Find the probability that exactly 3 experienced weight gain as a side effect.
(b) Find the probability that 3 or fewer experienced weight gain as a side effect.
(c) Find the probability that 4 or more patients experienced weight gain as a side effect.
(d) Find the probability that between 1 and 4 patients, inclusive, experienced weight gain as a side effect.

41. Airline Satisfaction A CNN/*USA Today*/Gallup poll in April 2005 reported that 75% of adult Americans were satisfied with the job the nation's major airlines were doing. Suppose 10 adult Americans are selected at random and the number who are satisfied is recorded.

(a) Find the probability that exactly 6 are satisfied with the airlines.
(b) Find the probability that fewer than 7 are satisfied with the airlines.
(c) Find the probability that 5 or more are satisfied with the airlines.
(d) Find the probability that between 5 and 8, inclusive, are satisfied with the airlines.

42. College Freshmen According to the Higher Education Research Institute, 55% of college freshmen in 4-year colleges and universities during 2003 were female. Suppose 12 freshmen are randomly selected and the number of females is recorded.

(a) Find the probability that exactly 7 of them are female.
(b) Find the probability that 5 or more are female.
(c) Find the probability that 8 or fewer are female.
(d) Find the probability that between 7 and 10, inclusive, are female.

43. On-Time Flights According to American Airlines, its NW flight 215 from Orlando to Los Angeles is on time 90% of the time. Suppose 100 flights are randomly selected.

(a) Compute the mean and standard deviation of the random variable X, the number of on-time flights in 100 trials of the probability experiment.
(b) Interpret the mean.
(c) Would it be unusual to observe 80 on-time flights in a random sample of 100 flights from Orlando to Los Angeles? Why?

44. Smokers According to the *Information Please Almanac*, NW 80% of adult smokers started smoking before turning 18 years old.

(a) Compute the mean and standard deviation of the random variable X, the number of smokers who started before 18 in 200 trials of the probability experiment.
(b) Interpret the mean.
(c) Would it be unusual to observe 180 smokers who started smoking before turning 18 years old in a random sample of 200 adult smokers? Why?

45. High-Speed Internet According to a report by the Commerce Department in the fall of 2004, 20% of U.S. households had some type of high-speed Internet connection.

(a) Compute the mean and standard deviation of the random variable X, the number of U.S. households with a high-speed Internet connection in 100 households.
(b) Interpret the mean.
(c) Would it be unusual to observe 18 U.S. households that have a high-speed Internet connection in 100 households? Why?

46. Allergy Sufferers Clarinex-D is a medication whose purpose is to reduce the symptoms associated with a variety of allergies. In clinical trials of Clarinex-D, 5% of the patients in the study experienced insomnia as a side effect.

(a) If 240 users of Clarinex-D are randomly selected, how many would we expect to experience insomnia as a side effect?
(b) Would it be unusual to observe 20 patients experiencing insomnia as a side effect in 240 trials of the probability experiment? Why?

47. Murder by Firearms According to the *Uniform Crime Report, 2003*, 66.9% of murders are committed with a firearm.

(a) If 100 murders are randomly selected, how many would we expect to be committed with a firearm?
(b) Would it be unusual to observe 75 murders by firearm in a random sample of 100 murders? Why?

48. Migraine Sufferers Depakote is a medication whose purpose is to reduce the pain associated with migraine headaches. In clinical trials and extended studies of Depakote, 2% of the patients in the study experienced weight gain as a side effect. Would it be unusual to observe 16 patients who experience weight gain in a random sample of 600 patients who take the medication? Why?

49. Asthma Control Singulair is a medication whose purpose is to control asthma attacks. In clinical trials of Singulair, 18.4% of the patients in the study experienced headaches as a side effect. Would it be unusual to observe 86 patients who experience headaches in a random sample of 400 patients who use this medication? Why?

50. Simulation According to the U.S. National Center for Health Statistics, there is a 98% probability that a 20-year-old male will survive to age 30.

(a) Using statistical software, simulate taking 100 random samples of size 30 from this population.
(b) Using the results of the simulation, compute the probability that exactly 29 of the 30 males survive to age 30.
(c) Compute the probability that exactly 29 of the 30 males survive to age 30, using the binomial probability distribution. Compare the results with part (b).
(d) Using the results of the simulation, compute the probability that at most 27 of the 30 males survive to age 30.
(e) Compute the probability that at most 27 of the 30 males survive to age 30, using the binomial probability distribution. Compare the results with part (d).
(f) Compute the mean number of male survivors in the 100 simulations of the probability experiment. Is it close to the expected value?
(g) Compute the standard deviation of the number of male survivors in the 100 simulations of the probability experiment. Compare the result to the theoretical standard deviation of the probability distribution.
(h) Did the simulation yield any unusual results?

51. Probability Applet Load the binomial applet on your APPLET computer.

(a) Set the probability of success, p, to 0.8 and the number of trials of the binomial experiment, n, to 10. Simulate shooting 10 free throws for $N = 1$. How many were made?
(b) Set the probability of success to 0.8 and the number of trials of the binomial experiment to 10. Simulate shooting 10 free throws $N = 1000$ times. Use the results of the simulation to estimate the probability of making 10 out of 10 free throws.
(c) Use the binomial probability formula to compute the probability of making 10 out of 10 free throws if the probability of success is 0.8.
(d) Use the results of the simulation to estimate the probability of making at least 8 out of 10 free throws.
(e) Use the binomial probability formula to compute the probability of making at least 8 out of 10 free throws.
(f) Determine the mean number of free throws made for the 1000 repetitions of the experiment. Is it close to the expected value?

52. Leisure Activity According to a 2002 survey by the National Endowment for the Arts, 60% of U.S. residents 18 and older attended a movie at least once in the previous year. Suppose you are performing a study and would like at least 12 people in the study to have attended a movie at least once in the past year.

(a) How many residents of the United States 18 years old or older do you expect to have to randomly select?
(b) How many residents of the United States 18 years old or older do you have to randomly select to have a 99% probability that the sample contains at least 12 who have attended a movie in the past year?

53. Educational Attainment According to the U.S. Census Bureau, in 2003 about 27% of residents of the United States 25 years old or older had earned at least a bachelor's degree. Suppose you are performing a study and would like at least 10 people in the study to have earned at least a bachelor's degree.

(a) How many residents of the United States 25 years old or older do you expect to randomly select?
(b) How many residents of the United States 25 years old or older do you have to randomly select to have probability 0.99 that the sample contains at least 10 who have earned at least a bachelor's degree?

Technology Step by Step **Computing Binomial Probabilities via Technology**

TI-83/84 Plus

Computing $P(x)$

***Step 1*: ** Press 2nd `VARS` to access the probability distribution menu.

***Step 2*: ** Highlight `Ø: binompdf(` for the TI-83 and `A: binompdf(` for the TI-84 and hit ENTER.

***Step 3*: ** With `binompdf(` on the `HOME` screen, type the number of trials n, the probability of success, p, and the number of successes, x. For example, with $n = 10$, $p = 0.2$, and $x = 4$, type

```
binompdf(10, 0.2, 4)
```

Then hit ENTER.

Computing $P(X \leq x)$

***Step 1*: ** Press 2nd `VARS` to access the probability distribution menu.

***Step 2*: ** Highlight `A: binomcdf(` for the TI-83 and `B: binomcdf(` for the TI-84 and hit ENTER.

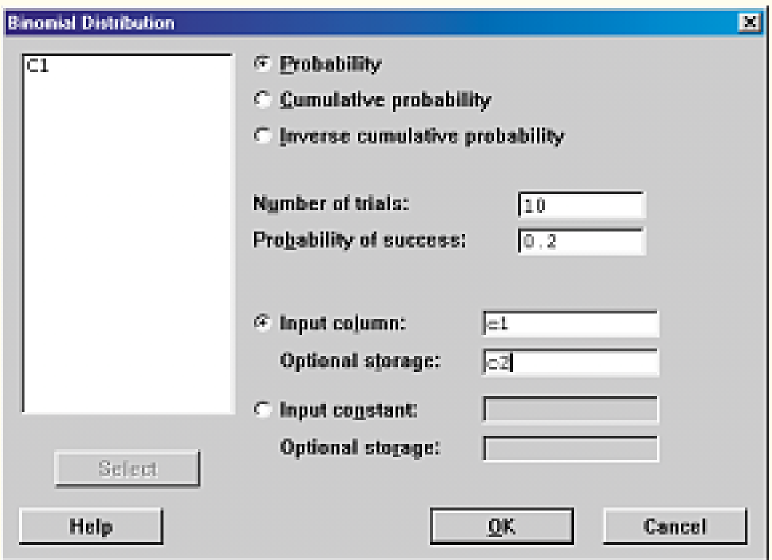
***Step 3*: ** With `binomcdf(` on the `HOME` screen, type the number of trials n, the probability of success, p, and the number of successes, x. For example, with $n = 10$, $p = 0.2$, and $x = 4$, type

```
binomcdf(10, 0.2, 4)
```

Then hit ENTER.

MINITAB

Computing $P(x)$



***Step 1*: ** Enter the possible values of the random variable x in C1. For example, with $n = 10$, $p = 0.2$, enter $0, 1, 2, \ldots, 10$ into C1.

***Step 2*: ** Select the **CALC** menu, highlight **Probability Distributions**, then highlight **Binomial** . . . .

***Step 3*: ** Fill in the window as shown to the left. Click OK.

Computing $P(X \leq x)$

Follow the same steps as those for computing $P(x)$. In the window that comes up after selecting Binomial Distribution, select Cumulative probability instead of Probability.

Excel

Computing $P(x)$

***Step 1*: ** Select the *fx* icon. Highlight Statistical in the Function category window. Highlight BINOMDIST in the Function name window.

***Step 2*: ** Fill in the window with the appropriate values. For example, if $x = 5$, $n = 10$, and $p = 0.2$, fill in the window as shown in Figure 8(a). Click OK.

Computing $P(X \leq x)$

Follow the same steps as those presented for computing $P(x)$. In the BINOMDIST window, type "TRUE" in the cumulative cell.

Consumer Reports® Quality Assurance in Customer Relations

The Customer Relations Department at Consumers Union (CU) receives thousands of letters and e-mails from customers each month. Some people write in asking how well a product performed during CU's testing, some people write in sharing their own experiences with their household products, and the remaining people write in for an array of other reasons.

To respond to each letter and e-mail that is received, Customer Relations recently upgraded its customer contact database. Although much of the process has been automated, it still requires employees to manually draft the responses. Given the current size of the department, each Customer Relations representative is required to draft approximately 300 responses each month.

As part of a quality assurance program, the Customer Relations manager would like to develop a plan that allows him to evaluate the performance of his employees. From past experience, he knows that the probability a new employee will write an initial draft of a response that contains errors is approximately 10%. The manager would like to know how many of the 300 responses he should sample to have a cost-effective quality assurance program.

(a) Let X be a discrete random variable that represents the number of the $n = 300$ draft responses that contain errors. Describe the probability distribution for X. Be sure to include the name of the probability distribution, possible values for the random variable X, and values of the parameters.

(b) To be effective, suppose the manager would like to have a 95% probability of finding at least one draft document that contains an error. Assuming that the probability that a draft document will have errors is known to be 10%, determine the appropriate sample size to satisfy the manager's requirements. *Hint*: We are required to find the number of draft documents that must be sampled so that the probability of finding at least one document containing an error is 95%. In other words, we have to determine n by solving: $P(X \geq 1) = 0.95$.

(c) Suppose the error rate is really 20%. What sample size will the manager have to review to have a 95% probability of finding one or more documents containing an error?

Note to Readers: *In many cases, our test protocol and analytical methods are more complicated than described in these examples. The data and discussions have been modified to make the material more appropriate for the audience.*

CHAPTER 6 Review

Summary

In this chapter, we discussed discrete probability distributions. A random variable represents the numerical measurement of the outcome from a probability experiment. Discrete random variables have either a finite or a countable number of outcomes. The term *countable* means that the values result from counting. Probability distributions must satisfy the following two criteria: (1) All probabilities must be between 0 and 1, inclusive, and (2) the sum of all probabilities must equal 1. Discrete probability distributions can be presented in a table, graph, or mathematical formula.

The mean and standard deviation of a random variable describe the center and spread of the distribution. The mean of a random variable is also called its expected value.

We discussed two discrete probability distributions in particular, the binomial and Poisson. A probability experiment is considered a binomial experiment if there are n independent trials of the experiment with only two outcomes. The probability of success, p, is the same for each trial of the experiment. Special formulas exist for computing the mean and standard deviation of a binomial random variable.

Formulas

Mean (or Expected Value) of a Discrete Random Variable

$\mu_X = E(X) = \sum xP(x)$

Variance of a Discrete Random Variable

$\sigma_X^2 = \sum (x - \mu_X)^2 \cdot P(x) = \sum [x^2 P(x)] - \mu_X^2$

Binomial Probability Distribution Function

$P(x) = {}_nC_x p^x (1 - p)^{n-x} \quad x = 0, 1, 2, \ldots$

Mean of a Binomial Random Variable

$\mu_X = np$

Standard Deviation of a Binomial Random Variable

$\sigma_X = \sqrt{np(1 - p)}$

Vocabulary

Random variable (p. 285)
Discrete random variable (p. 285)
Continuous random variable (p. 285)
Probability distribution (p. 286)
Probability histogram (p. 287)
Expected value (p. 291)
Binomial experiment (p. 298)
Trial (p. 298)
Binomial random variable (p. 298)

Objectives

Section	You should be able to . . .	Examples	Review Exercises
6.1	1 Distinguish between discrete and continuous random variables (p. 285)	1	1, 2
	2 Identify discrete probability distributions (p. 286)	2 and 3	3, 4, 5(a), 6(a)
	3 Construct probability histograms (p. 287)	4	5(b), 6(b)
	4 Compute and interpret the mean of a discrete random variable (p. 288)	5, 6 and 9	5(c), 6(c), 15(b), 16(b)
	5 Interpret the mean of a discrete random variable as an expected value (p. 291)	7	7, 8
	6 Compute the variance and standard deviation of a discrete random variable (p. 292)	8 and 9	5(d), 6(d), 15(b), 16(b)
6.2	1 Determine whether a probability experiment is a binomial experiment (p. 298)	1	9, 10
	2 Compute probabilities of binomial experiments (p. 300)	2 through 5	11(a)–(d), 12(a)–(d), 13(a)–(d), 14(a)–(d)
	3 Compute the mean and standard deviation of a binomial random variable (p. 305)	6	11(e), 12(e), 13(e), 14(e)
	4 Construct binomial probability histograms (p. 306)	7	15(d), 16(d)

Review Exercises

In Problems 1 and 2, determine whether the random variable is discrete or continuous. In each case, state the possible values of the random variable.

1. (a) The number of inches of snow that falls in Buffalo during the winter season.
(b) The number of days snow accumulates in Buffalo during the winter season.
(c) The number of golf balls hit into the ocean on the famous 18th hole at Pebble Beach on a randomly selected Sunday.

2. (a) The miles per gallon of gasoline in a 2005 Toyota Sienna.
(b) The number of children a randomly selected family has.
(c) The number of goals scored by the Edmonton Oilers in a season.

In Problems 3 and 4, determine whether the distribution is a discrete probability distribution. If not, state why.

3. No,

x	$P(x)$
0	0.34
1	0.21
2	0.13
3	0.04
4	0.01

4.

x	$P(x)$
0	0.40
1	0.31
2	0.23
3	0.04
4	0.02

5. Stanley Cup The Stanley Cup is a best-of-seven series to determine the champion of the National Hockey League. The following data represent the number of games played, X, in the Stanley Cup before a champion was determined from 1939 to 2004.

x	Frequency
4	20
5	16
6	17
7	13

Source: Information Please Almanac

(a) Construct a probability model for the random variable X, the number of games in the Stanley Cup.
(b) Draw a probability histogram.
(c) Compute and interpret the mean of the random variable X.
(d) Compute the standard deviation of the random variable X.

6. Property Crime In 2003, 77% of crime was property crime, according to the National Crime Victimization Survey. Suppose that four crimes are randomly selected. Let the random variable X represent the number of property crimes.
(a) Construct a probability model for the random variable X by constructing a tree diagram.
(b) Draw a probability histogram.

(c) Compute and interpret the mean of the random variable X.
(d) Compute the standard deviation of the random variable X.

7. **Life Insurance** Suppose a life insurance company sells a \$100,000 one-year term life insurance policy to a 35-year-old male for \$200. According to the *National Vital Statistics Report*, Vol. 53, No. 6, the probability that the male survives the year is 0.998592. Compute and interpret the expected value of this policy to the life insurance company.

8. **The Carnival** A carnival game is played as follows: You pay \$2 to draw a card from an ordinary deck. If you draw an ace, you win \$5. You win \$3 for a face card and \$10 for the seven of spades. If you pick anything else, you lose \$2. On average, how much money can the operator expect to make per customer?

In Problems 9 and 10, determine which of the following probability experiments represents a binomial experiment. If the probability experiment is not a binomial experiment, state why.

9. According to the *Chronicle of Higher Education*, there is a 54% probability that a randomly selected incoming freshman will graduate from college within 6 years. Suppose 10 incoming male freshmen are randomly selected. After 6 years, each student is asked whether he or she graduated.

10. An experiment is conducted in which a single die is cast until a 3 comes up. The number of throws required is recorded.

11. **High Cholesterol** According to the National Center for Health Statistics, 8% of 20- to 34-year-old females have high serum cholesterol.
(a) In a random sample of 10 females 20 to 34 years old, find the probability that exactly 0 have high serum cholesterol. Interpret this result.
(b) In a random sample of 10 females 20 to 34 years old, find the probability that exactly 2 have high serum cholesterol. Interpret this result.
(c) In a random sample of 10 females 20 to 34 years old, find the probability that at least 2 have high serum cholesterol. Interpret this result.
(d) In a random sample of 10 females 20 to 34 years old, find the probability that exactly 9 will not have high serum cholesterol. Interpret this result.
(e) In a random sample of 250 females 20 to 34 years old, what is the expected number with high serum cholesterol? What is the standard deviation?
(f) If a random sample of 250 females 20 to 34 years old resulted in 12 of them having high serum cholesterol, would this be unusual? Why?

12. **Driving Age** According to a Gallup poll conducted December 17 to 19, 2004, 60% of U.S. women 18 years old or older stated that the minimum driving age should be 18 years or older.
(a) In a random sample of 15 U.S. women 18 years old or older, find the probability that exactly 10 believe the minimum driving age should be 18 years or older.
(b) In a random sample of 15 U.S. women 18 years old or older, find the probability that fewer than 5 believe the minimum driving age should be 18 years or older.
(c) In a random sample of 15 U.S. women 18 years old or older, find the probability that at least 5 believe the minimum driving age should be 18 years or older.
(d) In a random sample of 15 U.S. women 18 years old or older, find the probability that exactly 12 *do not* believe the minimum driving age should be 18 years or older.
(e) In a random sample of 200 U.S. women 18 years old or older, what is the expected number who believe the minimum driving age should be 18 years old or older? What is the standard deviation?
(f) If a random sample of 200 U.S. women 18 years old or older resulted in 110 who believed that the minimum driving age should be 18 years or older, would this be unusual? Why?

13. **Nielsen Ratings** Nielsen Media Research determines ratings for television programs by placing meters on 5000 televisions throughout the United States. The 2005 NCAA Basketball Championship broadcast resulted in a rating of 15.0, which means 15% of households were tuned into the game.
(a) In a random sample of 20 households, find the probability that exactly 6 were tuned into the 2005 NCAA championship.
(b) In a random sample of 20 households, find the probability that fewer than 4 were tuned into the 2005 NCAA championship.
(c) In a random sample of 20 households, find the probability that at least 2 were tuned into the 2005 NCAA championship.
(d) In a random sample of 20 households, find the probability that exactly 17 were *not* tuned into the 2005 NCAA championship.
(e) In a random sample of 500 households, what is the expected number who were tuned into the 2005 NCAA championship?
(f) If a random sample of 500 households resulted in 95 that were tuned into the game, would this be unusual? Why?

14. **Quit Smoking** The drug Zyban is meant to suppress the urge to smoke. In clinical trials, 35% of the study's participants experienced insomnia when taking 300 mg of Zyban per day.
(*Source*: GlaxoSmithKline)
(a) In a random sample of 25 users of Zyban, find the probability that exactly 8 will experience insomnia.
(b) In a random sample of 25 users of Zyban, find the probability that fewer than 4 will experience insomnia.
(c) In a random sample of 25 users of Zyban, find the probability that at least 5 will experience insomnia.
(d) In a random sample of 25 users of Zyban, find the probability that exactly 20 will not experience insomnia.
(e) In a random sample of 1000 users of Zyban, what is the expected number who experience insomnia? What is the standard deviation?
(f) If a random sample of 1000 users of Zyban results in 330 who experience insomnia, would this be unusual? Why?

In Problems 15 and 16, (a) construct a binomial probability distribution with the given parameters, (b) compute the mean and standard deviation of the random variable by using the methods of Section 6.1, (c) compute the mean and standard deviation by using the methods of Section 6.2, and (d) draw the probability histogram, comment on its shape, and label the mean on the histogram.

15. $n = 5, p = 0.2$

16. $n = 8, p = 0.75$

17. State the condition required to use the Empirical Rule to check for unusual observations in a binomial experiment.

18. In sampling without replacement, the assumption of independence required for a binomial experiment is violated. Under what circumstances can we sample without replacement and still use the binomial probability formula to approximate probabilities?

THE CHAPTER 6 CASE STUDY IS LOCATED ON THE CD THAT ACCOMPANIES THIS TEXT.