CHAPTER 5

Discrete Random Variables

His sacred majesty, chance, decides everything.

Voltaire

CONTENTS

Introductory Statistics, DOI: 10.1016/B978-0-12-374388-6.00005-3

We continue our study of probability by introducing random variables —quantities whose values are determined by the outcome of the experiment. The expected value of a random variable is defined, and its properties are explored. The concept of *variance* is introduced. An important special type of random variable, known as the *binomial*, is studied.

5.1 INTRODUCTION

The National Basketball Association (NBA) draft lottery involves the 11 teams that had the worst won–lost records during the preceding year. Sixty-six Ping-Pong balls are placed in an urn. Each of these balls is inscribed with the name of a team; 11 have the name of the team with the worst record, 10 have the name of the team with the second-worst record, 9 have the name of the team with the third-worst record, and so on (with 1 ball having the name of the team with the 11th-worst record). A ball is then chosen at random, and the team whose name is on the ball is given the first pick in the draft of players about to enter the league. All the other balls belonging to this team are then removed, and another ball is chosen. The team to which this ball "belongs" receives the second draft pick. Finally, another ball is chosen, and the team named on this ball receives the third draft pick. The remaining draft picks, 4 through 11, are then awarded to the 8 teams that did not "win the lottery," in inverse order of their won–lost records. For instance, if the team with the worst record did not receive any of the 3 lottery picks, then that team would receive the fourth draft pick.

The outcome of this draft lottery is the order in which the 11 teams get to select players. However, rather than being concerned mainly about the actual outcome, we are sometimes more interested in the values of certain specified quantities. For instance, we may be primarily interested in finding out which team gets the first choice or in learning the draft number of our home team. These quantities of interest are known as *random variables*, and a special type, called *discrete*, will be studied in this chapter.

Random variables are introduced in Sec. 5.2. In Sec. 5.3 we consider the notion of the expected value of a random variable. We see that this represents, in a sense made precise, the average value of the random variable. Properties of the expected value are presented in Sec. 5.3.

Section 5.4 is concerned with the variance of a random variable, which is a measure of the amount by which a random variable tends to differ from its expected value. The concept of independent random variables is introduced in this section.

Section 5.5 deals with a very important type of discrete random variable that is called *binomial*. We see how such random variables arise and study their properties.

Sections 5.6 and 5.7 introduce the hypergeometric and the Poisson random variable. We explain how these discrete random variables arise and study their properties.

The first ball drawn in the 1993 NBA draft lottery belonged to the Orlando Magic, even though the Magic had finished the season with the 11th-worst record and so had only 1 of the 66 balls!

5.2 RANDOM VARIABLES

When a probability experiment is performed, often we are not interested in all the details of the experimental result, but rather are interested in the value of some numerical quantity determined by the result. For instance, in tossing dice, often we care about only their sum and are not concerned about the values on the individual dice. Also, an investor might not be interested in all the changes in the price of a stock on a given day, but rather might care about only the price at the end of the day. These quantities of interest that are determined by the result of the experiment are known as *random variables*.

Since the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to its possible values.

■ Example 5.1

The outcome of the NBA draft lottery experiment, which was discussed in Sec. 5.1, is the specification of the teams that are to receive the first, second, and third picks in the draft. For instance, outcome (3, 1, 4) could mean that the team with the third-worst record received pick number 1, the team with the worst record received pick number 2, and the team with the fourth-worst record received pick number 3. If we let X denote the team that received draft pick 1, then X would equal 3 if the outcome of the experiment were (3, 1, 4).

Clearly, X can take on any integral value between 1 and 11 inclusive. It will equal 1 if the first ball chosen is one of the 11 balls that belong to the team with the worst record, it will equal 2 if the first ball is one of the 10 balls that belong to the team with the second-worst record, and so on. Since each of the 66 balls is equally likely to be the first ball chosen, it follows that

$$P\{X = 1\} = \frac{11}{66} \qquad P\{X = 7\} = \frac{5}{66}$$
$$P\{X = 2\} = \frac{10}{66} \qquad P\{X = 8\} = \frac{4}{66}$$

$$P\{X = 3\} = \frac{9}{66} \qquad P\{X = 9\} = \frac{3}{66}$$
$$P\{X = 4\} = \frac{8}{66} \qquad P\{X = 10\} = \frac{2}{66}$$
$$P\{X = 5\} = \frac{7}{66} \qquad P\{X = 11\} = \frac{1}{66}$$
$$P\{X = 6\} = \frac{6}{66}$$

■

■ Example 5.2

Suppose we are about to learn the sexes of the three children of a certain family. The sample space of this experiment consists of the following 8 outcomes:

$$\{(b, b, b), (b, b, g), (b, g, b), (b, g, g), (g, b, b), (g, b, g), (g, g, b), (g, g, g)\}$$

The outcome (g, b, b) means, for instance, that the youngest child is a girl, the next youngest is a boy, and the oldest is a boy. Suppose that each of these 8 possible outcomes is equally likely, and so each has probability 1/8.

If we let X denote the number of female children in this family, then the value of X is determined by the outcome of the experiment. That is, X is a random variable whose value will be 0, 1, 2, or 3. We now determine the probabilities that X will equal each of these four values.

Since X will equal 0 if the outcome is (b, b, b), we see that

$$P\{X = 0\} = P\{(b, b, b)\} = \frac{1}{8}$$

Since X will equal 1 if the outcome is (b, b, g) or (b, g, b) or (g, b, b), we have

$$P\{X = 1\} = P(\{(b, b, g), (b, g, b), (g, b, b)\}) = \frac{3}{8}$$

Similarly,

$$P\{X = 2\} = P(\{(b, g, g), (g, b, g), (g, g, b)\}) = \frac{3}{8}$$
$$P\{X = 3\} = P(\{(g, g, g)\}) = \frac{1}{8}$$

■

A random variable is said to be *discrete* if its possible values constitute a sequence of separated points on the number line. Thus, for instance, any random variable that can take on only a finite number of different values is discrete.

In this chapter we will study discrete random variables. Let X be such a quantity, and suppose that it has n possible values, which we will label $x_1, x_2, \ldots, x_n$. As in Examples 5.1 and 5.2, we will use the notation $P\{X = x_i\}$ to represent the probability that X is equal to x_i. The collection of these probabilities is called the *probability distribution* of X. Since X must take on one of these n values, we know that

$$\sum_{i=1}^{n} P\{X = x_i\} = 1$$

■ Example 5.3

Suppose that X is a random variable that takes on one of the values 1, 2, or 3. If

$$P\{X = 1\} = 0.4 \quad \text{and} \quad P\{X = 2\} = 0.1$$

what is $P\{X = 3\}$?

Solution

Since the probabilities must sum to 1, we have

$$1 = P\{X = 1\} + P\{X = 2\} + P\{X = 3\}$$

or

$$1 = 0.4 + 0.1 + P\{X = 3\}$$

Therefore,

$$P\{X = 3\} = 1 - 0.5 = 0.5$$

A graph of $P\{X = i\}$ is shown in Fig. 5.1. ■

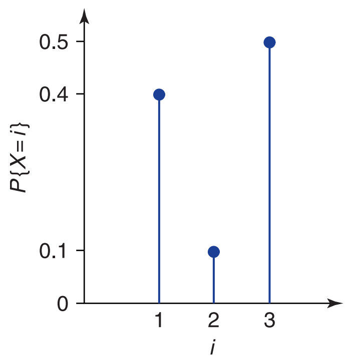

FIGURE 5.1
A graph of $P\{X = i\}$.

■ Example 5.4

A saleswoman has scheduled two appointments to sell encyclopedias. She feels that her first appointment will lead to a sale with probability 0.3. She also feels that the second will lead to a sale with probability 0.6 and that the results from the two appointments are independent. What is the probability distribution of X, the number of sales made?

Solution

The random variable X can take on any of the values 0, 1, or 2. It will equal 0 if neither appointment leads to a sale, and so

$$\begin{aligned} P\{X = 0\} &= P\{\text{no sale on first, no sale on second}\} \\ &= P\{\text{no sale on first}\}P\{\text{no sale on second}\} \quad \text{by independence} \\ &= (1 - 0.3)(1 - 0.6) = 0.28 \end{aligned}$$

The random variable X will equal 1 either if there is a sale on the first and not on the second appointment or if there is no sale on the first and one sale on the second appointment. Since these two events are disjoint, we have

$$\begin{aligned} P\{X = 1\} &= P\{\text{sale on first, no sale on second}\} \\ &\quad + P\{\text{no sale on first, sale on second}\} \\ &= P\{\text{sale on first}\}P\{\text{no sale on second}\} \\ &\quad + P\{\text{no sale on first}\}P\{\text{sale on second}\} \\ &= 0.3(1 - 0.6) + (1 - 0.3)0.6 = 0.54 \end{aligned}$$

Finally the random variable X will equal 2 if both appointments result in sales; thus

$$\begin{aligned} P\{X = 2\} &= P\{\text{sale on first, sale on second}\} \\ &= P\{\text{sale on first}\}P\{\text{sale on second}\} \\ &= (0.3)\,(0.6) = 0.18 \end{aligned}$$

As a check on this result, we note that

$$P\{X = 0\} + P\{X = 1\} + P\{X = 2\} = 0.28 + 0.54 + 0.18 = 1 \qquad ■$$

PROBLEMS

1. In Example 5.2, let the random variable Y equal 1 if the family has at least one child of each sex, and let it equal 0 otherwise. Find $P\{Y = 0\}$ and $P\{Y = 1\}$.
2. In Example 5.2, let the random variable W equal the number of girls that came before the first boy. (If the outcome is (g, g, g), take W equal to 3.) Give the possible values of W along with their probabilities. That is, give the probability distribution of W.
3. The following table presents the total number of tornadoes (violent, rotating columns of air with wind speeds over 100 miles per hour) in the United States between 1980 and 1991.

Year	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
Tornadoes	866	783	1046	931	907	684	764	656	702	856	1133	1132

Source: U.S. National Oceanic and Atmospheric Administration.

Suppose that one of these years is randomly selected, and let X denote the number of tornadoes in that year. Find

(a) $P\{X > 900\}$

(b) $P\{X \leq 800\}$

(c) $P\{X = 852\}$

(d) $P\{700 < X < 850\}$

4. Suppose a pair of dice is rolled. Let X denote their sum. What are the possible values of X? Assuming that each of the 36 possible outcomes of the experiment is equally likely, what is the probability distribution of X?
5. In Prob. 4, let Y denote the smaller of the two numbers appearing on the two dice. (If both dice show the same number, take that as the value of Y.) Determine the probability distribution of Y.
6. Two people are to meet in the park. Each person is equally likely to arrive, independent of the other, at 2:00, 2:30, or 3:00 p.m. Let X equal the time that the first person to arrive has to wait, where X is taken to equal 0 if both people arrive at the same time.

   (a) What are the possible values of X?

   (b) What are the probabilities that X assumes each of these values?

7. Two volleyball teams are to play a 2-out-of-3 series, in which they continue to play until one has won 2 games. Suppose that the home team wins each game played, independently, with probability 0.7. Let X denote the number of games played.

(a) What are the possible values of X?
(b) What is the probability distribution of X?

8. Suppose that 2 batteries are randomly chosen from a bin containing 10 batteries, of which 7 are good and 3 are defective. Let X denote the number of defective batteries chosen. Give the possible values of X along with their probabilities.

9. A shipment of parts contains 120 items of which 10 are defective. Two of these items are randomly chosen and inspected. Let X denote the number that are defective. Find the probability distribution of X.

10. A contractor will bid for two jobs in sequence. She has a 0.5 probability of winning the first job. If she wins the first job, then she has a 0.2 chance of winning the second job; if she loses the first job, then she has a 0.4 chance of winning the second job. (In the latter case, her bid will be lower.) Let X denote the number of jobs that she wins. Find the probability distribution of X.

11. Whenever a certain college basketball player goes to the foul line for two shots, he makes his first shot with probability 0.75. If he makes the first shot, then he makes the second shot with probability 0.80; if he misses the first shot, then he makes the second one with probability 0.70. Let X denote the number of shots he makes when he goes to the foul line for two shots. Find the probability distribution of X.

In Probs. 12, 13, and 14, tell whether the set of numbers $p(i)$, $i = 1, 2, 3, 4, 5$, can represent the probabilities $P\{X = i\}$ of a random variable whose set of values is 1, 2, 3, 4, or 5. If your answer is no, explain why.

12.

i	$p(i)$
1	0.4
2	0.1
3	0.2
4	0.1
5	0.3

13.

i	$p(i)$
1	0.2
2	0.3
3	0.4
4	−0.1
5	0.2

14.

i	$p(i)$
1	0.3
2	0.1
3	0.2
4	0.4
5	0.0

15. In a study of 223 households in a small rural town in Iowa, a sociologist has collected data about the number of children in each household. The data showed that there are 348 children in the town, with the breakdown of the number of children in each household as follows: 38 households have 0 children, 82 have 1 child, 57 have 2 children, 34 have 3 children, 10 have 4 children, and 2 have 5 children. Suppose that one of these households is randomly selected for a more detailed interview. Let X denote the number of children in the household selected. Give the probability distribution of X.

16. Suppose that, in Prob. 15, one of the 348 children of the town is randomly selected. Let Y denote the number of children in the family of the selected child. Find the probability distribution of Y.

17. Suppose that X takes on one of the values 1, 2, 3, 4, or 5. If $P\{X < 3\} = 0.4$ and $P\{X > 3\} = 0.5$, find

(a) $P\{X = 3\}$

(b) $P\{X < 4\}$

18. An insurance agent has two clients, each of whom has a life insurance policy that pays $100,000 upon death. Their probabilities of dying this year are 0.05 and 0.10. Let X denote the total amount of money that will be paid this year to the clients' beneficiaries. Assuming that the event that client 1 dies is independent of the event that client 2 dies, determine the probability distribution of X.

19. A bakery has 3 special cakes at the beginning of the day. The daily demand for this type of cake is

0	with probability 0.15
1	with probability 0.20
2	with probability 0.35
3	with probability 0.15
4	with probability 0.10
5 or more	with probability 0.05

Let X denote the number of cakes that remain unsold at the end of the day. Determine the probability distribution of X.

5.3 EXPECTED VALUE

A key concept in probability is the expected value of a random variable. If X is a discrete random variable that takes on one of the possible values $x_1, x_2, \ldots, x_n$, then the *expected value* of X, denoted by $E[X]$, is defined by

$$E[X] = \sum_{i=1}^{n} x_i P\{X = x_i\}$$

The expected value of X is a weighted average of the possible values of X, with each value weighted by the probability that X assumes it. For instance, suppose X is equally likely to be either 0 or 1, and so

$$P\{X = 0\} = P\{X = 1\} = \frac{1}{2}$$

then

$$E[X] = 0\left(\frac{1}{2}\right) + 1\left(\frac{1}{2}\right) = \frac{1}{2}$$

is equal to the ordinary average of the two possible values 0 and 1 that X can assume. On the other hand, if

$$P\{X = 0\} = \frac{2}{3} \quad \text{and} \quad P\{X = 1\} = \frac{1}{3}$$

then

$$E[X] = 0\left(\frac{2}{3}\right) + 1\left(\frac{1}{3}\right) = \frac{1}{3}$$

is a weighted average of the two possible values 0 and 1, where the value 0 is given twice as much weight as the value 1, since it is twice as likely that X will equal 0 as it is that X will equal 1.

Definition and Terminology

The *expected value* of a discrete random variable X whose possible values are $x_1, x_2, \ldots, x_n$, is denoted by $E[X]$ and is defined by

$$E[X] = \sum_{i=1}^{n} x_i P\{X = x_i\}$$

Other names used for $E[X]$ are the *expectation* of X and the *mean* of X.

Another motivation for the definition of the expected value relies on the frequency interpretation of probabilities. This interpretation assumes that if a very large number (in theory, an infinite number) of independent replications of an experiment are performed, then the proportion of time that event A occurs will equal $P(A)$. Now consider a random variable X that takes on one of the possible values

$x_1, x_2, \ldots, x_n$, with respective probabilities $p(x_1), p(x_2), \ldots, p(x_n)$; and think of X as representing our winnings in a single game of chance. We will now argue that if we play a large number of such games, then our average winning per game will be $E[X]$. To see this, suppose that we play N games, where N is a very large number. Since, by the frequency interpretation of probability, the proportion of games in which we win x_i will approximately equal $p(x_i)$, it follows that we will win x_i in approximately $Np(x_i)$ of the N games. Since this is true for each x_i, it follows that our total winnings in the N games will be approximately equal to

$$\sum_{i=1}^{n} x_i(\text{number of games we win } x_i) = \sum_{i=1}^{n} x_i N p(x_i)$$

Therefore, our average winning per game will be

$$\frac{\sum_{i=1}^{n} x_i N p(x_i)}{N} = \sum_{i=1}^{n} x_i p(x_i) = E[X]$$

In other words, if X is a random variable associated with some experiment, then the average value of X over a large number of replications of the experiment is approximately $E[X]$.

■ Example 5.5

Suppose we roll a die that is equally likely to have any of its 6 sides appear face up. Find $E[X]$, where X is the side facing up.

Solution

Since

$$P[X = i] = \frac{1}{6} \quad \text{for } i = 1, 2, 3, 4, 5, 6$$

we see that

$$\begin{aligned} E[X] &= 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) \\ &= \frac{21}{6} = 3.5 \end{aligned}$$

Note that the expected value of X is not one of the possible values of X. Even though we call $E[X]$ the expected value of X, it should be interpreted not as the value that we *expect* X to have, but rather as the average value of X in a large number of repetitions of the experiment. That is, if we continually roll a die, then after a large number of rolls the average of all the outcomes will be approximately 3.5. ■

■ Example 5.6

Consider a random variable X that takes on either the value 1 or 0 with respective probabilities p and $1 - p$. That is,

$$P[X = 1] = p \quad \text{and} \quad P\{X = 0\} = 1 - p$$

Find $E[X]$.

Solution

The expected value of this random variable is

$$E[X] = 1(p) + 0(1 - p) = p$$ ■

■ Example 5.7

An insurance company sets its annual premium on its life insurance policies so that it makes an expected profit of 1 percent of the amount it would have to pay out upon death. Find the annual premium on a \$200,000 life insurance policy for an individual who will die during the year with probability 0.02.

Solution

In units of \$1000, the insurance company will set its premium so that its expected profit is 1 percent of 200, or 2. If we let A denote the annual premium, then the profit of the insurance company will be either

$$A \quad \text{if policyholder lives}$$

or

$$A - 200 \quad \text{if policyholder dies}$$

Therefore, the expected profit is given by

$$\begin{aligned} E[\text{profit}] &= AP\{\text{policyholder lives}\} + (A - 200)P\{\text{policyholder dies}\} \\ &= A(1 - 0.02) + (A - 200)(0.02) \\ &= A - 200(0.02) \\ &= A - 4 \end{aligned}$$

So the company will have an expected profit of \$2000 if it charges an annual premium of \$6000. ■

As seen in Example 5.7, $E[X]$ is always measured in the same units (dollars in that example) as the random variable X.

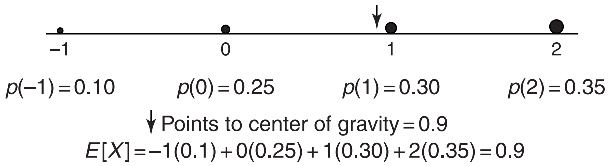

FIGURE 5.2
Center of gravity $= E[X]$.

The concept of expected value is analogous to the physical concept of the center of gravity of a distribution of mass. Consider a discrete random variable with probabilities given by $p(x_i), i \geq 1$. If we imagine a rod on which weights having masses $p(x_i)$ are placed at points $x_i, i \geq 1$ (Fig. 5.2), then the point at which the rod would be in balance is known as the *center of gravity*. It can be shown by the laws of mechanics that this point is

$$\sum_i x_i p(x_i) = E[X]$$

5.3.1 Properties of Expected Values

Let X be a random variable with expected value $E[X]$. If c is a constant, then the quantities cX and $X + c$ are also random variables and so have expected values. The following useful results can be shown:

$$E[cX] = cE[X]$$

$$E[X + c] = E[X] + c$$

That is, the expected value of a constant times a random variable is equal to the constant times the expected value of the random variable; and the expected value of a constant plus a random variable is equal to the constant plus the expected value of the random variable.

■ Example 5.8

A married couple works for the same employer. The wife's Christmas bonus is a random variable whose expected value is \$1500.

(a) If the husband's bonus is set to equal 80 percent of his wife's, find the expected value of the husband's bonus.
(b) If the husband's bonus is set to equal \$1000 more than his wife's, find its expected value.

Solution

Let X denote the bonus (in dollars) to be paid to the wife.

(a) Since the bonus paid to the husband is equal to $0.8X$, we have

$$E[\text{bonus to husband}] = E[0.8X] = 0.8E[X] = \$1200$$

(b) In this case the bonus to be paid to the husband is $X + 1000$, and so

$$E[\text{bonus to husband}] = E[X + 1000] = E[X] + 1000 = \$2500$$ ■

A very useful property is that the expected value of the sum of random variables is equal to the sum of the individual expected values.

For any random variables X and Y,

$$E[X + Y] = E[X] + E[Y]$$

■ Example 5.9

The following are the annual incomes of 7 men and 7 women who are residents of a certain community.

Annual Income (in $1000)	
Men	**Women**
33.5	24.2
25.0	19.5
28.6	27.4
41.0	28.6
30.5	32.2
29.6	22.4
32.8	21.6

Suppose that a woman and a man are randomly chosen. Find the expected value of the sum of their incomes.

Solution

Let X be the man's income and Y the woman's income. Since X is equally likely to be any of the 7 values in the men's column, we see that

$$E[X] = \frac{1}{7}(33.5 + 25 + 28.6 + 41 + 30.5 + 29.6 + 32.8)$$
$$= \frac{221}{7} \approx 31.571$$

Similarly,

$$E[Y] = \frac{1}{7}(24.2 + 19.5 + 27.4 + 28.6 + 32.2 + 22.4 + 21.6)$$
$$= \frac{175.9}{7} \approx 25.129$$

Therefore, the expected value of the sum of their incomes is

$$E[X + Y] = E[X] + E[Y]$$
$$\approx 56.700$$

That is, the expected value of the sum of their incomes is approximately \$56,700. ■

■ Example 5.10

The following table lists the number of civilian full-time law enforcement employees in eight cities in 1990.

City	Civilian law enforcement employees
Minneapolis, MN	105
Newark, NJ	155
Omaha, NE	149
Portland, OR	195
San Antonio, TX	290
San Jose, CA	357
Tucson, AZ	246
Tulsa, OK	178

Source: Department of Justice, *Uniform Crime Reports for the United States, 1990.*

Suppose that two of the cities are to be randomly chosen and all the civilian law enforcement employees of these cities are to be interviewed. Find the expected number of people who will be interviewed.

Solution

Let X be the number of civilian employees in the first city chosen, and let Y be the number in the second city chosen. Since the selection of the cities is random, each of the 8 cities has the same chance to be the first city selected; similarly, each of the 8 cities has the same chance to be the second selection. Therefore, both X and Y are equally likely to be any of the 8 values in the given table, and so

$$\begin{aligned} E[X] = E[Y] &= \frac{1}{8}(105 + 155 + 149 + 195 + 290 + 357 + 246 + 178) \\ &= \frac{1675}{8} \end{aligned}$$

and so

$$E[X + Y] = E[X] + E[Y] = \frac{1675}{4} = 418.75$$

That is, the expected number of interviews that will be needed is 418.75. ■

By using the frequency interpretation of expected value as being the average value of a random variable over a large number of replications of the experiment, it is easy to see intuitively why the expected value of a sum is equal to the sum of the expected values. For instance, suppose we always make the same two bets on each spin of a roulette wheel, one bet concerning the color of the slot where the ball lands and the other concerning the number on that slot. Let X and Y be the amounts (in dollars) that we lose on the color bet and on the number bet, respectively, in a single spin of the wheel. Then, $X + Y$ is our total loss in a single spin. Now, if in the long run we lose an average of 1 per spin on the color bet (so $E[X] = 1$) and we lose an average of 2 per spin on the number bet (so $E[Y] = 2$), then our average total loss per spin (equal to $E[X + Y]$) will clearly be $1 + 2 = 3$.

The result that the expected value of the sum of random variables is equal to the sum of the expected values holds for not only two but any number of random variables.

Useful Result

For any positive integer k and random variables $X_1, \ldots, X_k$,

$$E\left[\sum_{i=1}^{k} X_i\right] = \sum_{i=1}^{k} E[X_i]$$

■ Example 5.11

A building contractor has sent in bids for three jobs. If the contractor obtains these jobs, they will yield respective profits of 20, 25, and 40 (in units of \$1000). On the other hand, for each job the contractor does not win, he will incur a loss (due to time and money already spent in making the bid) of 2. If the probabilities that the contractor will get these jobs are, respectively, 0.3, 0.6, and 0.2, what is the expected total profit?

Solution

Let X_i denote the profit from job $i, i = 1, 2, 3$. Now by interpreting a loss as a negative profit, we have

$$P\{X_1 = 20\} = 0.3 \quad P\{X_1 = -2\} = 1 - 0.3 = 0.7$$

Therefore,

$$E[X_1] = 20(0.3) - 2(0.7) = 4.6$$

Similarly,

$$E[X_2] = 25(0.6) - 2(0.4) = 14.2$$

and

$$E[X_3] = 40(0.2) - 2(0.8) = 6.4$$

The total profit is $X_1 + X_2 + X_3$, and so

$$\begin{aligned} E[\text{total profit}] &= E[X_1 + X_2 + X_3] \\ &= E[X_1] + E[X_2] + E[X_3] \\ &= 4.6 + 14.2 + 6.4 \\ &= 25.2 \end{aligned}$$

Therefore, the expected total profit is \$25,200. ■

PROBLEMS

In the following problems, $p(i)$ stands for $P\{X = i\}$.

1. Find the expected value of X when
   (a) $p(1) = 1/3, p(2) = 1/3, p(3) = 1/3$
   (b) $p(1) = 1/2, p(2) = 1/3, p(3) = 1/6$
   (c) $p(1) = 1/6, p(2) = 1/3, p(3) = 1/2$
2. Find $E[X]$ when
   (a) $p(1) = 0.1, p(2) = 0.3, p(3) = 0.3, p(4) = 0.2, p(5) = 0.1$
   (b) $p(1) = 0.3, p(2) = 0.1, p(3) = 0.2, p(4) = 0.1, p(5) = 0.3$
   (c) $p(1) = 0.2, p(2) = 0, p(3) = 0.6, p(4) = 0, p(5) = 0.2$
   (d) $p(3) = 1$

3. A distributor makes a profit of \$30 on each item that is received in perfect condition and suffers a loss of \$6 on each item that is received in less-than-perfect condition. If each item received is in perfect condition with probability 0.4, what is the distributor's expected profit per item?
4. In a certain liability suit, a lawyer has to decide whether to charge a straight fee of \$1200 or to take the case on a contingency basis, in which case she will receive a fee of \$5000 only if her client wins the case. Determine whether the straight fee or the contingency arrangement will result in a higher expected fee when the probability that the client will win the case is
   (a) 1/2
   (b) 1/3
   (c) 1/4
   (d) 1/5
5. Suppose X can take on any of the values 1, 2, and 3. Find $E[X]$ if

$$p(1) = 0.3 \quad \text{and} \quad p(2) = 0.5$$

6. Let X be a random variable that is equally likely to take on any of the values $1, 2, \ldots, n$. That is,

$$P\{X = i\} = \frac{1}{n} \qquad i = 1, \ldots, n$$

   (a) If $n = 2$, find $E[X]$.
   (b) If $n = 3$, find $E[X]$.
   (c) If $n = 4$, find $E[X]$.
   (d) For general n, what is the value of $E[X]$?
   (e) Verify your answer in part (d) by making use of the algebraic identity

$$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

7. A pair of fair dice is rolled. Find the expected value of the
   (a) Smaller
   (b) Larger
   of the two upturned faces. (If both dice show the same number, then take this to be the value of both the smaller and the larger of the upturned faces.)
8. A computer software firm has been told by its local utility company that there is a 25 percent chance that the electricity will be shut off at some time during the next working day. The company estimates

that it will cost $400 in lost revenues if employees do not use their computers tomorrow, and it will cost $1200 if the employees suffer a cutoff in power while using them. If the company wants to minimize the expected value of its loss, should it risk using the computers?

9. An engineering firm must decide whether to prepare a bid for a construction project. It will cost $800 to prepare a bid. If it does prepare a bid, then the firm will make a gross profit (excluding the preparation cost) of $0 if it does not get the contract, $3000 if it gets the contract and the weather is bad, or $6000 if it gets the contract and the weather is not bad. If the probability of getting the contract is 0.4 and the probability that the weather will be bad is 0.6, what is the company's expected net profit if it prepares a bid?

10. All blood donated to a blood bank is tested before it is used. To reduce the total number of tests, the bank takes small samples of the blood of four separate donors and pools these samples. The pooled blood is analyzed. If it is deemed acceptable, then the bank stores the blood of these four people for future use. If it is deemed unacceptable, then the blood from each of the four donors is separately tested. Therefore, either one test or five tests are needed to handle the blood of four donors. Find the expected number of tests needed if each donor's blood is independently unacceptable with probability 0.1.

11. Two people are randomly chosen from a group of 10 men and 20 women. Let X denote the number of men chosen, and let Y denote the number of women.
    (a) Find $E[X]$.
    (b) Find $E[Y]$.
    (c) Find $E[X + Y]$.

12. If the two teams in a World Series have the same chance of winning each game, independent of the results of previously played games, then the probabilities that the series will end in 4, 5, 6, or 7 games are, respectively, 1/8, 1/4, 5/16, and 5/16. What is the expected number of games played in such a series?

13. A company that operates a chain of hardware stores is planning to open a new store in one of two locations. If it chooses the first location, the company thinks it will make a first-year profit of $40,000 if the store is successful and will have a first-year loss of $10,000 if the store is unsuccessful. At the second location, the company thinks it will make a first-year profit of $60,000 if the store is successful and a first-year loss of $25,000 if the store is unsuccessful.
    (a) If the probability of success is 1/2 for both locations, which location will result in a larger expected first-year profit?
    (b) Repeat part (a), this time assuming that the probability that the store is successful is 1/3.

14. If it rains tomorrow, you will earn \$200 by doing some tutoring; if it is dry, you will earn \$300 by doing construction work. If the probability of rain is 1/4, what is the expected amount that you will earn tomorrow?
15. If you have a 1/10 chance of gaining \$400 and a 9/10 chance of gaining −\$50 (that is, of losing \$50), what is your expected gain?
16. If an investment has a 0.4 probability of making a \$30,000 profit and a 0.6 probability of losing \$15,000, does this investment have a positive expected gain?
17. It costs \$40 to test a certain component of a machine. If a defective component is installed, it costs \$950 to repair the damage that results to the machine. From the point of view of minimizing the expected cost, determine whether the component should be installed without testing if it is known that its probability of being defective is
    (a) 0.1
    (b) 0.05
    (c) 0.01
    (d) What would the probability of a defective component be if one were indifferent between testing and installing the component untested?
18. A fair bet is one in which the expected gain is equal to 0. If you bet 1 unit on a number in roulette, then you will gain 35 units if the number appears and will lose 1 unit if it does not. If the roulette wheel is perfectly balanced, then the probability that your number will appear is 1/38. What is the expected gain on a 1-unit bet? Is it a fair bet?
19. A school holding a raffle will sell each ticket for \$1. The school will give out seven prizes −1 for \$100, 2 for \$50, and 4 for \$25. Suppose you purchase one ticket. If a total of 500 tickets is sold, what is your expected gain? (*Hint*: Your gain is −1 (if you do not win a prize), 24 (if you win a \$25 prize), 49 (if you win a \$50 prize), or 99 (if you win a \$100 prize).)
20. A roulette wheel has 18 numbers colored red, 18 colored black, and 2 (zero and double zero) that are uncolored. If you bet 1 unit on the outcome red, then either you win 1 if a red number appears or you lose 1 if a red number does not appear. What is your expected gain?
21. The first player to win 2 sets is the winner of a tennis match. Suppose that whatever happened in the previous sets, each player has probability 1/2 of winning the next set. Determine the expected number of sets played.
22. Suppose in Prob. 21 that the players are not of equal ability and that player 1 wins each set, independent of the results of earlier sets, with probability 1/3.
    (a) Find the expected number of sets played.
    (b) What is the probability that player 1 wins?

23. An insurance company sells a life insurance policy that pays $250,000 if the insured dies for an annual premium of $1400. If the probability that the policyholder dies in the course of the year is 0.005, what is the company's expected annual profit from that policyholder?

24. In Example 5.8, find in both (a) and (b) the expected value of the sum of the bonuses earned by the wife and husband.

25. If $E[X] = \mu$, what is $E[X - \mu]$?

26. Four buses carrying 148 students from the same school arrive at a football stadium. The buses carry, respectively, 40, 33, 50, and 25 students. One of the students is randomly selected. Let X be the number of students who were on the bus carrying the selected student. One of 4 bus drivers is also randomly chosen. Let Y be the number of students who were on his or her bus.

(a) Calculate $E[X]$ and $E[Y]$.

(b) Can you give an intuitive reason why $E[X]$ is larger than $E[Y]$?

27. A small nursery must decide on the number of Christmas trees to stock. The trees cost $6 each and are to be sold for $20. Unsold trees are worthless. The nursery estimates that the probability distribution for the demand on trees is as follows:

Amount demanded	1200	1500	1800
Probability	0.5	0.2	0.3

Determine the nursery's expected profit if it purchases

(a) 1200 trees

(b) 1500 trees

(c) 1800 trees

28. Repeat Prob. 27, this time assuming that any unsold tree must be disposed of at a cost of $2 per tree.

29. The daily demand at a bakery for a certain cake is as follows:

Daily demand	0	1	2	3	4
Probability	0.15	0.25	0.30	0.15	0.15

It costs the bakery $4 to bake each cake, which sells for $20. Any cakes left unsold at the end of the day are thrown away. Would the bakery have a higher expected profit if it baked 2 or 3 or 4 cakes daily?

30. If $E[X] = 5$ and $E[Y] = 12$, find

(a) $E[3X + 4Y]$

(b) $E[2 + 5Y + X]$

(c) $E[4 + Y]$

31. Determine the expected sum of a pair of fair dice by
   (a) Using the probability distribution of the sum
   (b) Using Example 5.5 along with the fact that the expected value of the sum of random variables is equal to the sum of their expected values
32. A husband's year-end bonus will be

0	with probability 0.3
\$1000	with probability 0.6
\$2000	with probability 0.1

His wife's bonus will be

\$1000	with probability 0.7
\$2000	with probability 0.3

Let S be the sum of their bonuses, and find $E[S]$.
33. The following data give the numbers of U.S. bank failures in the years 1995 to 2002.

Year	Closed or assisted
1995	8
1996	6
1997	1
1998	3
1999	8
2000	7
2001	4
2002	11

Suppose that a congressional committee has decided to randomly choose 2 of these years and then document each of the incidents that occurred in either year. Determine the expected number of such incidents.
34. Repeat Prob. 33, this time supposing that the committee is to randomly choose 3 of the years.
35. A small taxi company has 4 taxis. In a month's time, each taxi will get 0 traffic tickets with probability 0.3, 1 traffic ticket with probability 0.5, or 2 traffic tickets with probability 0.2. What is the expected number of tickets per month amassed by the fleet of 4 taxis?
36. Suppose that 2 batteries are randomly selected from a drawer containing 8 good and 2 defective batteries. Let W denote the number of defective batteries selected.

(a) Find $E[W]$ by first determining the probability distribution of W. Let X equal 1 if the first battery chosen is defective, and let X equal 0 otherwise. Also let Y equal 1 if the second battery is defective and equal 0 otherwise.
(b) Give an equation relating X, Y, and W.
(c) Use the equation in (b) to obtain $E[W]$.

5.4 VARIANCE OF RANDOM VARIABLES

It is useful to be able to summarize the properties of a random variable by a few suitably chosen measures. One such measure is the expected value. However, while the expected value gives the weighted average of the possible values of the random variable, it does not tell us anything about the variation, or spread, of these values. For instance, consider random variables U, V, and W, whose values and probabilities are as follows:

$$U = 0 \quad \text{with probability } 1$$

$$V = \begin{cases} -1 & \text{with probability } 1/2 \\ 1 & \text{with probability } 1/2 \end{cases}$$

$$W = \begin{cases} -10 & \text{with probability } 1/2 \\ 10 & \text{with probability } 1/2 \end{cases}$$

Whereas all three random variables have expected value 0, there is clearly less spread in the values of U than in V and less spread in the values of V than in W.

Since we expect a random variable X to take on values around its mean $E[X]$, a reasonable way of measuring the variation of X is to consider how far X tends to be from its mean on the average. That is, we could consider $E[|X - \mu|]$, where $\mu = E[X]$ and $|X - \mu|$ is the absolute value of the difference between X and μ. However, it turns out to be more convenient to consider not the absolute value but the square of the difference.

Definition *If X is a random variable with expected value μ, then the* variance *of X, denoted by $\text{Var}(X)$, is defined by*

$$\text{Var}(X) = E[(X - \mu)^2]$$

Upon expanding $(X - \mu)^2$ to obtain $X^2 - 2\mu X + \mu^2$ and then taking the expected value of each term, we obtain after a little algebra the following useful computational formula for $\text{Var}(X)$:

$$\text{Var}(X) = E[X^2] - \mu^2 \tag{5.1}$$

where

$$\mu = E[X]$$

Using Eq. (5.1) is usually the easiest way to compute the variance of X.

■ Example 5.12

Find Var(X) when the random variable X is such that

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

Solution

In Example 5.6 we showed that $E[X] = p$. Therefore, using the computational formula for the variance, we have

$$\mathrm{Var}(X) = E[X^2] - p^2$$

Now,

$$X^2 = \begin{cases} 1^2 & \text{if } X = 1 \\ 0^2 & \text{if } X = 0 \end{cases}$$

Since $1^2 = 1$ and $0^2 = 0$, we see that

$$\begin{aligned} E[X^2] &= 1 \cdot P\{X = 1\} + 0 \cdot P\{X = 0\} \\ &= 1 \cdot p = p \end{aligned}$$

Hence,

$$\mathrm{Var}(X) = p - p^2 = p(1 - p)$$ ■

■ Example 5.13

The return from a certain investment (in units of \$1000) is a random variable X with probability distribution

$$P\{X = -1\} = 0.7 \quad P\{X = 4\} = 0.2 \quad P\{X = 8\} = 0.1$$

Find Var(X), the variance of the return.

Solution

Let us first compute the expected return as follows:

$$\begin{aligned}\mu = E[X] &= -1(0.7) + 4(0.2) + 8(0.1)\\ &= 0.9\end{aligned}$$

That is, the expected return is \$900. To compute $\text{Var}(X)$, we use the formula

$$\text{Var}(X) = E[X^2] - \mu^2$$

Now, since X^2 will equal $(-1)^2$, 4^2, or 8^2 with respective probabilities of 0.7, 0.2, and 0.1, we have

$$\begin{aligned}E[X^2] &= 1(0.7) + 16(0.2) + 64(0.1)\\ &= 10.3\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Var}(X) &= 10.3 - (0.9)^2\\ &= 9.49\end{aligned}$$

■

5.4.1 Properties of Variances

For any random variable X and constant c, it can be shown that

$$\begin{aligned}\text{Var}(cX) &= c^2\text{Var}(X)\\ \text{Var}(X + c) &= \text{Var}(X)\end{aligned}$$

That is, the variance of the product of a constant and a random variable is equal to the constant squared times the variance of the random variable; and the variance of the sum of a constant and a random variable is equal to the variance of the random variable.

Whereas the expected value of the sum of random variables is always equal to the sum of the expectations, the corresponding result for variances is generally not true. For instance, consider the following.

$$\begin{aligned}\text{Var}(X + X) &= \text{Var}(2X)\\ &= 2^2\text{Var}(X)\\ &\neq \text{Var}(X) + \text{Var}(X)\end{aligned}$$

However, there is an important case in which the variance of the sum of random variables is equal to the sum of the variances, and this occurs when the random variables are independent. Before presenting this result, we must introduce the concept of independent random variables.

We say that X and Y are independent if knowing the value of one of them does not change the probabilities of the other. That is, if X takes on one of the values x_i, $i \geq 1$, and Y takes on one of the values y_j, $j \geq 1$, then X and Y are independent if the events that X is equal to x_i and Y is equal to y_j are independent events for all x_i and y_j.

Definition *Random variables X and Y are* independent *if knowing the value of one of them does not change the probabilities of the other.*

It turns out that the variance of the sum of independent random variables is equal to the sum of their variances.

Useful Result

If X and Y are independent random variables, then

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

More generally, if $X_1, X_2, \ldots, X_k$ are independent random variables, then

$$\mathrm{Var}\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} \mathrm{Var}(X_i)$$

■ Example 5.14

Determine the variance of the sum obtained when a pair of fair dice is rolled.

Solution

Number the dice, and let X be the value of the first die and Y the value of the second die. Then the desired quantity is $\mathrm{Var}(X + Y)$. Since the outcomes of the two dice are independent, we know that

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

To compute $\mathrm{Var}(X)$, the variance of the face of the first die, recall that it was shown in Example 5.5 that

$$E[X] = \frac{7}{2}$$

Since X^2 is equally likely to be any of the values 1^2, 2^2, 3^2, 4^2, 5^2, and 6^2, we have

$$E[X^2] = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6}$$

Therefore,

$$\begin{aligned}\text{Var}(X) &= E[X^2] - \left(\frac{7}{2}\right)^2 \\ &= \frac{91}{6} - \frac{49}{4} \\ &= \frac{35}{12}\end{aligned}$$

Since Y has the same probability distribution as X, it also has variance $35/12$, and so

$$\text{Var}(X + Y) = \frac{35}{12} + \frac{35}{12} = \frac{35}{6}$$

■

The positive square root of the variance is called the *standard deviation* (SD).

Definition *The quantity $SD(X)$, defined by*

$$SD(X) = \sqrt{Var(X)}$$

is called the standard deviation *of X.*

The standard deviation, like the expected value, is measured in the same units as is the random variable. That is, if the value of X is given in terms of miles, then so will the expected value and the standard deviation, too. To compute the standard deviation of a random variable, compute the variance and then take its square root.

■ Example 5.15

The annual gross earnings of a certain rock singer are a random variable with an expected value of \$400,000 and a standard deviation of \$80,000. The singer's manager receives 15 percent of this amount. Determine the expected value and standard deviation of the amount received by the manager.

Solution

If we let X denote the earnings (in units of \$1000) of the singer, then the manager earns $0.15X$. Its expected value is obtained as follows:

$$E[0.15X] = 0.15E[X] = 60$$

To compute the standard deviation, first determine the variance:

$$\mathrm{Var}(0.15X) = (0.15)^2 \mathrm{Var}(X)$$

Taking the square root of both sides of the preceding gives

$$\mathrm{SD}(0.15X) = 0.15\ \mathrm{SD}(X) = 12$$

Therefore, the amount received by the manager is a random variable with an expected value of \$60,000 and a standard deviation of \$12,000. ■

PROBLEMS

1. Determine the variances of random variables U, V, and W, defined at the beginning of Sec. 5.4.
2. Let $p(i) = P\{X = i\}$. Consider
   (a) $p(0) = 0.50, p(1) = 0.50$
   (b) $p(0) = 0.60, p(1) = 0.40$
   (c) $p(0) = 0.90, p(1) = 0.10$
   In which case do you think $\mathrm{Var}(X)$ would be largest? And in which case would it be smallest? Determine the actual variances and check your answers.
3. Suppose that, for some constant c, $P\{X = c\} = 1$. Find $\mathrm{Var}(X)$.
4. Find the variances of the random variables specified in Prob. 1 of Sec. 5.3.
5. Find $\mathrm{Var}(X)$ for the X given in Prob. 5 of Sec. 5.3.
6. If the probability that you earn \$300 is 1/3 and the probability that you earn \$600 is 2/3, what is the variance of the amount that you earn?
7. Find the variance of the number of sets played in the situation described in Prob. 21 of Sec. 5.3.
8. A small electronics company that started up 4 years ago has 60 employees. The following is a frequency table relating the number of years (rounded up) that these employees have been with the company.

Number of years	Frequency
1	12
2	25
3	16
4	7

Suppose one of these workers is randomly chosen. Let X denote the number of years he or she has been with the company. Find

(a) $E[X]$

(b) $\mathrm{Var}(X)$

9. The vacation time received by a worker of a certain company depends on the economic performance of the company. Suppose that Fong, an employee of this company, will receive

0 weeks' vacation	with probability 0.4
1 week's vacation	with probability 0.2
2 weeks' vacation	with probability 0.4

Suppose also that Fontanez, another employee, will receive

0 weeks' vacation	with probability 0.3
1 week's vacation	with probability 0.4
2 weeks' vacation	with probability 0.3

Let X denote the number of weeks of vacation for Fong and Y denote the number of weeks for Fontanez.

(a) Which do you think is larger, $\mathrm{Var}(X)$ or $\mathrm{Var}(Y)$?

(b) Find $\mathrm{Var}(X)$.

(c) Find $\mathrm{Var}(Y)$.

10. Find the variance of the profit earned by the nursery in Prob. 27(b) of Sec. 5.3.

11. Two fair coins are tossed. Determine $\mathrm{Var}(X)$ when X is the number of heads that appear.

(a) Use the definition of the variance.

(b) Use the fact that the variance of the sum of independent random variables is equal to the sum of the variances.

12. Find the variance of the number of tickets obtained by the fleet of taxis, as described in Prob. 35 of Sec. 5.3. Assume that the numbers of tickets received by each of the taxis are independent.

13. A lawyer must decide whether to charge a fixed fee of \$2000 or to take a contingency fee of \$8000 if she wins the case (and \$0 if she loses). She estimates that her probability of winning is 0.3. Determine the standard deviation of her fee if

(a) She takes the fixed fee.

(b) She takes the contingency fee.

14. Find the standard deviation of the amount of money you will earn in Prob. 14 of Sec. 5.3.

15. The following is a frequency table giving the number of courses being taken by 210 first-year students at a certain college.

Number of classes	Frequency
1	2
2	15
3	37
4	90
5	49
6	14
7	3

Let X denote the number of courses taken by a randomly chosen student. Find

(a) $E[X]$

(b) $\mathrm{SD}(X)$

16. The amount of money that Robert earns has expected value \$30,000 and standard deviation \$3000. The amount of money that his wife Sandra earns has expected value \$32,000 and standard deviation \$5000. Determine the

(a) Expected value

(b) Standard deviation

of the total earnings of this family. In answering part (b), assume that Robert's earnings and Sandra's earnings are independent. (*Hint*: In answering part (b), first find the variance of the family's total earnings.)

17. If $\mathrm{Var}(X) = 4$, what is $\mathrm{SD}(3X)$? (*Hint*: First find $\mathrm{Var}(3X)$.)

18. If $\mathrm{Var}(2X + 3) = 16$, what is $\mathrm{SD}(X)$?

19. If X and Y are independent random variables, both having variance 1, find

(a) $\mathrm{Var}(X + Y)$

(b) $\mathrm{Var}(X - Y)$

5.5 BINOMIAL RANDOM VARIABLES

One of the most important types of random variables is the binomial, which arises as follows. Suppose that n independent subexperiments (or *trials*) are performed, each of which results in either a "success" with probability p or a "failure" with probability $1 - p$. If X is the total number of successes that occur in n trials, then X is said to be a *binomial* random variable with parameters n and p.

Before presenting the general formula for the probability that a binomial random variable X takes on each of its possible values $0, 1, \ldots, n$, we consider a special case. Suppose that $n = 3$ and that we are interested in the probability that X is equal to 2. That is, we are interested in the probability that 3 independent trials, each of which is a success with probability p, will result in a total of 2 successes. To determine this probability, consider all the outcomes that give rise to exactly 2 successes:

$$(s, s, f), (s, f, s), (f, s, s)$$

The outcome (s, f, s) means, for instance, that the first trial is a success, the second a failure, and the third a success. Now, by the assumed independence of the trials, it follows that each of these outcomes has probability $p^2(1-p)$. For instance, if S_i is the event that trial i is a success and F_i is the event that trial i is a failure, then

$$\begin{aligned} P(s, f, s) &= P(S_1 \cap F_2 \cap S_3) \\ &= P(S_1)P(F_2)P(S_3) \quad \text{by independence} \\ &= p(1-p)p \end{aligned}$$

Since each of the 3 outcomes that result in a total of 2 successes consists of 2 successes and 1 failure, it follows in a similar fashion that each occurs with probability $p^2(1-p)$. Therefore, the probability of a total of 2 successes in the 3 trials is $3p^2(1-p)$.

Consider now the general case in which we have n independent trials. Let X denote the number of successes. To determine $P\{X = i\}$, consider any outcome that results in a total of i successes. Since this outcome will have a total of i successes and $n-i$ failures, it follows from the independence of the trials that its probability will be $p^i(1-p)^{n-i}$. That is, each outcome that results in $X = i$ will have the same probability $p^i(1-p)^{n-i}$. Therefore, $P\{X = i\}$ is equal to this common probability multiplied by the number of different outcomes that result in i successes. Now, it can be shown that there are $n!/[i!(n-i)!]$ different outcomes that result in a total of i successes and $n-i$ failures, where $n!$ (read "n factorial") is equal to 1 when $n = 0$ and is equal to the product of the natural numbers from 1 to n otherwise. That is,

$$\begin{aligned} 0! &= 1 \\ n! &= n \cdot (n-1) \cdots 3 \cdot 2 \cdot 1 \quad \text{if } n > 0 \end{aligned}$$

A binomial random variable with parameters n and p represents the number of successes in n independent trials, when each trial is a success with probability p. If X is such a random variable, then for $i = 0, \ldots, n$,

$$P\{X = i\} = \frac{n!}{i!\,(n-i)!} p^i (1-p)^{n-i}$$

As a check of the preceding equation, note that it states that the probability that there are no successes in n trials is

$$\begin{aligned} p\{X=0\} &= \frac{n!}{0!\,(n-0)!}p^0(1-p)^{n-0} \\ &= (1-p)^n \quad \text{since } 0! = p^0 = 1 \end{aligned}$$

However, the foregoing is clearly correct since the probability that there are 0 successes, and so all the trials are failures, is, by independence, $(1-p)(1-p)\cdots(1-p) = (1-p)^n$.

The probabilities of three binomial random variables with respective parameters $n = 10$, $p = 0.5$, $n = 10$, $p = 0.3$, and $n = 10$, $p = 0.6$ are presented in Fig. 5.3.

■ Example 5.16

Three fair coins are flipped. If the outcomes are independent, determine the probability that there are a total of i heads, for $i = 0, 1, 2, 3$.

Solution

If we let X denote the number of heads ("successes"), then X is a binomial random variable with parameters $n = 3$, $p = 0.5$. By the preceding we have

$$P\{X=0\} = \frac{3!}{0!\,3!}\left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^3 = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

$$P\{X=1\} = \frac{3!}{1!\,2!}\left(\frac{1}{2}\right)^1\left(\frac{1}{2}\right)^2 = 3\left(\frac{1}{2}\right)^3 = \frac{3}{8}$$

$$P\{X=2\} = \frac{3!}{2!\,1!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = 3\left(\frac{1}{2}\right)^3 = \frac{3}{8}$$

$$P\{X=3\} = \frac{3!}{3!\,0!}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^0 = \left(\frac{1}{2}\right)^3 = \frac{1}{8}$$

■

■ Example 5.17

Suppose that a particular trait (such as eye color or handedness) is determined by a single pair of genes, and suppose that d represents a dominant gene and r a recessive gene. A person with the pair of genes (d, d) is said to be *pure dominant*, one with the pair (r, r) is said to be *pure recessive*, and one with the pair (d, r) is said to be *hybrid*. The pure dominant and the hybrid are alike in appearance. When two individuals mate, the resulting offspring receives one gene from each parent, and this gene is equally likely to be either of the parent's two genes.

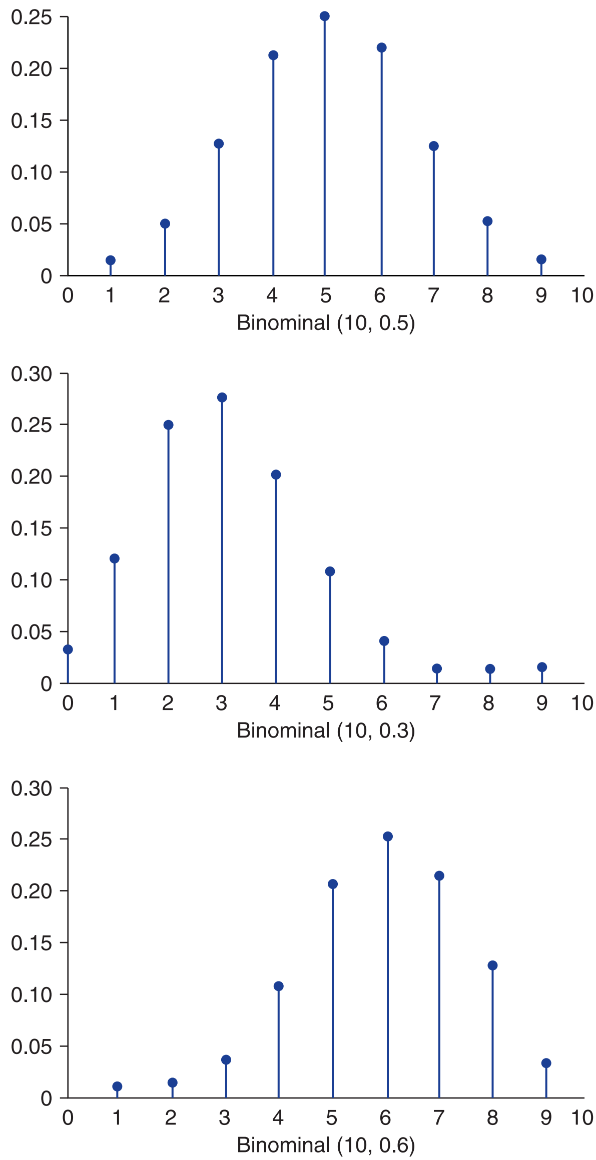

FIGURE 5.3
Binomial probabilities.

(a) What is the probability that the offspring of two hybrid parents has the opposite (recessive) appearance?

(b) Suppose two hybrid parents have 4 offsprings. What is the probability 1 of the 4 offspring has the recessive appearance?

Solution

(a) The offspring will have the recessive appearance if it receives a recessive gene from each parent. By independence, the probability of this is $(1/2)(1/2) = 1/4$.

(b) Assuming the genes obtained by the different offspring are independent (which is the common assumption in genetics), it follows from part (a) that the number of offspring having the recessive appearance is a binomial random variable with parameters $n = 4$ and $p = 1/4$. Therefore, if X is the number of offspring that have the recessive appearance, then

$$
\begin{aligned}
P\{X = 1\} &= \frac{4!}{1!\,3!}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^3 \\
&= 4\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^3 \\
&= \frac{27}{64}
\end{aligned}
$$

Suppose that X is a binomial random variable with parameters n and p, and suppose we want to calculate the probability that X is less than or equal to some value j. In principle, we could compute this as follows:

$$
P\{X \le j\} = \sum_{i=0}^{j} P\{X = i\} = \sum_{i=0}^{j} \frac{n!}{i!\,(n-i)!} p^i (1-p)^{n-i}
$$

The amount of computation called for in the preceding equation can be rather large. To relieve this, Table D.5 (in App. D) gives the values of $P\{X \le j\}$ for $n \le 20$ and for various values of p. In addition, you can use Program 5-1. In this program you enter the binomial parameters and the desired value of j, and you get as output the probability that the binomial is less than or equal to j, the probability that the binomial is equal to j, and the probability that the binomial is greater than or equal to j. ■

■ Example 5.18

(a) Determine $P\{X \le 12\}$ when X is a binomial random variable with parameters 20 and 0.4.

(b) Determine $P\{Y \le 10\}$ when Y is a binomial random variable with parameters 16 and 0.5.

Solution

From Table D.5, we see that

(a) $P\{X \leq 12\} = 0.9790$
(b) $P\{Y \leq 10\} = 1 - P\{Y < 10\} = 1 - P\{Y \leq 9\} = 1 - 0.7728 = 0.2272$

We could also have run Program 5-1 to obtain the following:
The probability that a binomial (20, 0.4) is less than or equal to 12 is 0.978969.
The probability that a binomial (16, 0.5) is greater than or equal to 10 is 0.2272506. ■

5.5.1 Expected Value and Variance of a Binomial Random Variable

A binomial (n, p) random variable X is equal to the number of successes in n independent trials when each trial is a success with probability p. As a result, we can represent X as

$$X = \sum_{i=1}^{n} X_i$$

where X_i is equal to 1 if trial i is a success and is equal to 0 if trial i is a failure. Since

$$P\{X_i = 1\} = p \quad \text{and} \quad P\{X_i = 0\} = 1 - p$$

it follows from the results of Examples 5.6 and 5.12 that

$$E[X_i] = p \quad \text{and} \quad \text{Var}(X_i) = p(1 - p)$$

Therefore, using the fact that the expectation of the sum of random variables is equal to the sum of their expectations, we see that

$$E[X] = np$$

Also, since the variance of the sum of independent random variables is equal to the sum of their variances, we have

$$\text{Var}(X) = np(1 - p)$$

Let us summarize.

If X is binomial with parameters n and p, then

$$E[X] = np$$

$$\text{Var}(X) = np(1 - p)$$

■ Example 5.19

Suppose that each screw produced is independently defective with probability 0.01. Find the expected value and variance of the number of defective screws in a shipment of size 1000.

Solution

The number of defective screws in the shipment of size 1000 is a binomial random variable with parameters $n = 1000$, $p = 0.01$. Hence, the expected number of defective screws is

$$E[\text{number of defectives}] = 1000(0.01) = 10$$

and the variance of the number of defective screws is

$$\text{Var}(\text{number of defectives}) = 1000(0.01)(0.99) = 9.9$$ ■


Bettmann

Jacques Bernoulli

Historical Perspective

Independent trials having a common success probability p were first studied by the Swiss mathematician Jacques Bernoulli (1654–1705). In his book *Ars Conjectandi* (The Art of Conjecturing), published by his nephew Nicholas eight years after his death in 1713, Bernoulli showed that if the number of such trials were large, then the proportion of them that were successes would be close to p with a probability near 1.

Jacques Bernoulli was from the first generation of the most famous mathematical family of all time. Altogether there were anywhere between 8 and 12 Bernoullis, spread over three generations, who made fundamental contributions to probability, statistics, and mathematics. One difficulty in knowing their exact number is the fact that several had the same name. (For example, two of the sons of Jacques' brother Jean were named Jacques and Jean.) Another difficulty is that several of the Bernoullis were known by different names in different places. Our Jacques (sometimes written Jaques), for instance, was also known as Jakob (sometimes written Jacob) and as James Bernoulli. But whatever their number, their influence and output were prodigious. Like the Bachs of music, the Bernoullis of mathematics were a family for the ages!

PROBLEMS

1. Find (a) $4!$ (b) $5!$ (c) $7!$
2. Find (a) $\dfrac{8!}{3!\,5!}$ (b) $\dfrac{7!}{3!\,4!}$ (c) $\dfrac{9!}{4!\,5!}$

3. Given that $9! = 362{,}880$, find $10!$.
4. Use the probability distribution of a binomial random variable with parameters n and p to show that

$$P\{X = n\} = p^n$$

Then argue directly why this is valid.
5. If X is a binomial random variable with parameters $n = 8$ and $p = 0.4$, find
   (a) $P\{X = 3\}$
   (b) $P\{X = 5\}$
   (c) $P\{X = 7\}$
6. Each ball bearing produced is independently defective with probability 0.05. If a sample of 5 is inspected, find the probability that
   (a) None are defective.
   (b) Two or more are defective.
7. Suppose you will be attending 6 hockey games. If each game independently will go to overtime with probability 0.10, find the probability that
   (a) At least 1 of the games will go into overtime.
   (b) At most 1 of the games will go into overtime.
8. A satellite system consists of 4 components and can function if at least 2 of them are working. If each component independently works with probability 0.8, what is the probability the system will function?
9. A communications channel transmits the digits 0 and 1. Because of static, each digit transmitted is independently incorrectly received with probability 0.1. Suppose an important single-digit message is to be transmitted. To reduce the chance of error, the string of digits 0 0 0 0 0 is to be transmitted if the message is 0 and the string 1 1 1 1 1 is to be transmitted if the message is 1. The receiver of the message uses "majority rule" to decode; that is, she decodes the message as 0 if there are at least 3 zeros in the message received and as 1 otherwise.
   (a) For the message to be incorrectly decoded, how many of the 5 digits received would have to be incorrect?
   (b) What is the probability that the message is incorrectly decoded?
10. A multiple-choice examination has 3 possible answers for each of 5 questions. What is the probability that a student will get 4 or more correct answers just by guessing?
11. A man claims to have extrasensory perception (ESP). As a test, a fair coin is to be flipped 8 times, and he is asked to predict the outcomes in advance. Suppose he gets 6 correct answers. What is the probability that he would have got at least this number of correct answers if he had no ESP but had just guessed?

12. Each diskette produced by a certain company will be defective with probability 0.05 independent of the others. The company sells the diskettes in packages of 10 and offers a money-back guarantee that all the diskettes in a package will be nondefective. Suppose that this offer is always taken up.
    (a) What is the probability that a package is returned?
    (b) If someone buys 3 packages, what is the probability that exactly 1 of them is returned?
13. Four fair dice are to be rolled. Find the probability that
    (a) 6 appears at least once.
    (b) 6 appears exactly once.
    (c) 6 appears at least twice.
14. Statistics indicate that alcohol is a factor in 55 percent of fatal automobile accidents. Of the next 3 fatal automobile accidents, find the probability that alcohol is a factor in
    (a) All 3
    (b) Exactly 2
    (c) At least 1
15. Individuals who have two sickle cell genes will develop the disease called *sickle cell anemia*, while individuals having none or one sickle cell gene will not be harmed. If two people, both of whom have one sickle cell gene, have a child, then that child will receive two sickle cell genes with probability 1/4. Suppose that both members of each of three different couples have exactly one sickle cell gene. If each of these couples has a child, find the probability that
    (a) None of the children receives two sickle cell genes.
    (b) Exactly one of the children receives two sickle cell genes.
    (c) Exactly two of the children receive two sickle cell genes.
    (d) All three children receive two sickle cell genes.
16. Let X be a binomial random variable with parameters $n = 20$ and $p = 0.6$. Find
    (a) $P\{X \le 14\}$
    (b) $P\{X < 10\}$
    (c) $P\{X \ge 13\}$
    (d) $P\{X > 10\}$
    (e) $P\{9 \le X \le 16\}$
    (f) $P\{7 < X < 15\}$
17. A fair die is to be rolled 20 times. Find the expected value of the number of times
    (a) 6 appears.
    (b) 5 or 6 appears.
    (c) An even number appears.
    (d) Anything else but 6 appears.
18. Find the variances of the random variables in Prob. 17.

19. The probability that a fluorescent bulb burns for at least 500 hours is 0.90. Of 8 such bulbs, find the probability that
    (a) All 8 burn for at least 500 hours.
    (b) Exactly 7 burn for at least 500 hours.
    (c) What is the expected value of the number of bulbs that burn for at least 500 hours?
    (d) What is the variance of the number of bulbs that burn for at least 500 hours?
20. If a fair coin is flipped 500 times, what is the standard deviation of the number of times that a head appears?
21. The FBI has reported that 44 percent of murder victims are killed with handguns. If 4 murder victims are randomly selected, find
    (a) The probability that they were all killed by handguns
    (b) The probability that none were killed by handguns
    (c) The probability that at least two were killed by handguns
    (d) The expected number killed by handguns
    (e) The standard deviation of the number killed by handguns
22. The expected number of heads in a series of 10 flips of a coin is 6. What is the probability there are 8 heads?
23. If X is a binomial random variable with expected value 4 and variance 2.4, find
    (a) $P\{X = 0\}$
    (b) $P\{X = 12\}$
24. If X is a binomial random variable with expected value 4.5 and variance 0.45, find
    (a) $P\{X = 3\}$
    (b) $P\{X \geq 4\}$
25. Find the mean and standard deviation of a binomial random variable with parameters
    (a) $n = 100, p = 0.5$
    (b) $n = 100, p = 0.4$
    (c) $n = 100, p = 0.6$
    (d) $n = 50, p = 0.5$
    (e) $n = 150, p = 0.5$
    (f) $n = 200, p = 0.25$
26. The National Basketball Association championship series is a best-of-seven series, meaning that the first team to win four games is declared the champion. In its history, no team has ever come back to win the championship after being behind three games to one. Assuming that each of the games played in this year's series is equally likely to be won by either team, independent of the results of earlier games, what is the probability that the upcoming championship series will be the first time that a team comes back from a three-game-to-one deficit to win the series?

*5.6 HYPERGEOMETRIC RANDOM VARIABLES

Suppose that n batteries are to be randomly selected from a bin of N batteries, of which Np are functional and the other $N(1-p)$ are defective. The random variable X, equal to the number of functional batteries in the sample, is then said to be a hypergeometric random variable with parameters n, N, p.

We can interpret the preceding experiment as consisting of n trials, where trial i is considered a success if the ith battery withdrawn is a functional battery. Since each of the N batteries is equally likely to be the ith one withdrawn, it follows that trial i is a success with probability $Np/N = p$. Therefore, X can be thought of as representing the number of successes in n trials where each trial is a success with probability p. What distinguishes X from a binomial random variable is that these trials are not independent. For instance, suppose that two batteries are to be withdrawn from a bin of five batteries, of which one is functional and the others defective. (That is, $n = 2$, $N = 5$, $p = 1/5$.) Then the probability that the second battery withdrawn is functional is $1/5$. However, if the first one withdrawn is functional, then the conditional probability that the second one is functional is 0 (since when the second battery is chosen all four remaining batteries in the bin are defective). That is, when the selections of the batteries are made without replacing the previously chosen ones, the trials are not independent, so X is not a binomial random variable.

By using the result that each of the n trials is a success with probability p, it can be shown that the expected number of successes is np. That is,

$$E[X] = np$$

In addition, it can be shown that the variance of the hypergeometric random variable is given by

$$\text{Var}(X) = \frac{N-n}{N-1} np(1-p)$$

Thus, whereas the expected value of the hypergeometric random variable with parameters n, N, p is the same as that of the binomial random variable with parameters n, p, its variance is smaller than that of the binomial by the factor $(N-n)/(N-1)$.

■ Example 5.20

If 6 people are randomly selected from a group consisting of 12 men and 8 women, then the number of women chosen is a hypergeometric random variable with parameters $n = 6$, $N = 20$, $p = 8/20 = 0.4$. Its mean and variance are

$$E[X] = 6(0.4) = 2.4 \quad \text{Var}(X) = \frac{14}{19} 6(0.4)(0.6) \approx 1.061$$

Similarly, the number of men chosen is a hypergeometric random variable with parameters $n = 6$, $N = 20$, $p = 0.6$. ■

Suppose now that N, the number of batteries in the urn, is large in comparison to n, the number to be selected. For instance, suppose that 20 batteries are to be randomly chosen from a bin containing 10,000 batteries of which 90 percent are functional. In this case, no matter which batteries were previously chosen each new selection will be defective with a probability that is approximately equal to 0.9. For instance, the first battery selected will be functional with probability 0.9. If the first battery is functional then the next one will also be functional with probability $8999/9999 \approx .89999$, whereas if the first battery is defective then the second one will be functional with probability $9000/9999 \approx .90009$. A similar argument holds for the other selections, and thus we may conclude that when N is large in relation to n, then the n trials are approximately independent, which means that X is approximately a binomial random variable.

When N is large in relation to n, a hypergeometric random variable with parameters n, N, p approximately has a binomial distribution with parameters n and p.

PROBLEMS

In the following problems, state whether the random variable X is binomial or hypergeometric. Also give its parameters (n and p if it is binomial or n, N, and p if it is hypergeometric).

1. A lot of 200 items contains 18 defectives. Let X denote the number of defectives in a sample of 20 items.
2. A restaurant knows from past experience that 15 percent of all reservations do not show. Twenty reservations are expected tonight. Let X denote the number that show.
3. In one version of the game of lotto each player selects six of the numbers from 1 to 54. The organizers also randomly select six of these numbers. These latter six are called the winning numbers. Let X denote how many of a given player's six selections are winning numbers.
4. Each new fuse produced is independently defective with probability 0.05. Let X denote the number of defective fuses in the last 100 produced.
5. Suppose that a collection of 100 fuses contains 5 that are defective. Let X denote the number of defectives discovered when 20 of them are randomly chosen and inspected.
6. A deck of cards is shuffled and the cards are successively turned over. Let X denote the number of aces in the first 10 cards.

7. A deck of cards is shuffled and the top card is turned over. The card is then returned to the deck and the operation repeated. This continues until a total of 10 cards have been turned over. Let X denote the number of aces that have appeared.

*5.7 POISSON RANDOM VARIABLES

A random variable X that takes on one of the values 0, 1, 2, ... is said to be a Poisson random variable with parameter λ if for some positive value λ its probabilities are given by

$$P\{X = i\} = c\lambda^i / i!, \quad i = 0, 1, \ldots$$

In the preceding, c is a constant that depends on λ. Its explicit value is given by $c = e^{-\lambda}$, where e is a famous mathematical constant that is approximately equal to 2.718.

A random variable X is called a *Poisson* random variable with parameter λ if

$$P\{X = i\} = \frac{e^{-\lambda}\lambda^i}{i!}, \quad i = 0, 1, \ldots$$

A graph of the probabilities of a Poisson random variable having parameter $\lambda = 4$ is presented in Fig. 5.4.

■ Example 5.21

If X is a Poisson random variable with parameter $\lambda = 2$, find $P\{X = 0\}$.

Solution

$$P\{X = 0\} = \frac{e^{-2}2^0}{0!}$$

Using the facts that $2^0 = 1$ and $0! = 1$, we obtain

$$P\{X = 0\} = e^{-2} = 0.1353$$

In the preceding, the value of e^{-2} was obtained from a table of exponentials. Alternatively, it could have been obtained from a scientific hand calculator or a personal computer. ■

Poisson random variables arise as approximations to binomial random variables. Consider n independent trials, each of which results in either a success with probability p or a failure with probability $1 - p$. If the number of trials is large and the

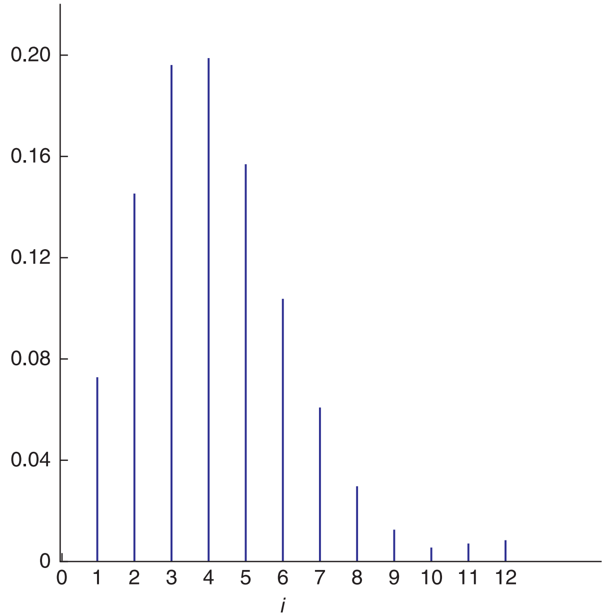

FIGURE 5.4
Probabilities of a Poisson random variable with $\lambda = 4$.

probability of a success on a trial is small, then the total number of successes will be approximately a Poisson random variable with parameter $\lambda = np$.

Some examples of random variables whose probabilities are approximately given, for some λ, by Poisson probabilities are the following:

1. The number of misprints on a page of a book
2. The number of people in a community who are at least 100 years old
3. The number of people entering a post office on a given day

Each of these is approximately Poisson because of the Poisson approximation to the binomial. For instance, we can suppose that each letter typed on a page has a small probability of being a misprint, and so the number of misprints on a page will be approximately a Poisson random variable with parameter $\lambda = np$, where n is the large number of letters on a page and p is the small probability that any given letter is a misprint.

■ Example 5.22

Suppose that items produced by a certain machine are independently defective with probability 0.1. What is the probability that a sample of 10 items will contain at most 1 defective item? What is the Poisson approximation for this probability?

Solution

If we let X denote the number of defective items, then X is a binomial random variable with parameters $n = 10$, $p = 0.1$. Thus the desired probability is

$$P\{X = 0\} + P\{X = 1\} = \binom{10}{0}(0.1)^0(0.9)^{10} + \binom{10}{1}(0.1)^1(0.9)^9$$
$$= 0.7361$$

Since $np = 10(0.1) = 1$, the Poisson approximation yields the value

$$P\{X = 0\} + P\{X = 1\} = e^{-1} + e^{-1} = 0.7358$$

Thus, even in this case, where n is equal to 10 (which is not that large) and p is equal to 0.1 (which is not that small), the Poisson approximation to the binomial probability is quite accurate. ■

Both the expected value and the variance of a Poisson random variable are equal to λ. That is, we have the following.

If X is a Poisson random variable with parameter $\lambda, \lambda > 0$, then

$$E[X] = \lambda$$
$$\mathrm{Var}(X) = \lambda$$

■ Example 5.23

Suppose the average number of accidents occurring weekly on a particular highway is equal to 1.2. Approximate the probability that there is at least one accident this week.

Solution

Let X denote the number of accidents. Because it is reasonable to suppose that there are a large number of cars passing along the highway, each having a small probability of being involved in an accident, the number of such accidents should be approximately a Poisson random variable. That is, if X denotes the number of accidents that will occur this week, then X is approximately a Poisson random variable with mean value $\lambda = 1.2$. The desired probability is now obtained as follows:

$$\begin{aligned}
P\{X > 0\} &= 1 - P\{X = 0\} \\
&= 1 - \frac{e^{-1.2}(1.2)^0}{0!} \\
&= 1 - e^{-1.2} \\
&= 1 - 0.3012 \\
&= 0.6988
\end{aligned}$$

Therefore, there is approximately a 70 percent chance that there will be at least one accident this week. ■

PROBLEMS

The following will be needed for the problems. The values given are correct to four decimal places.

$$e^{-1/2} = 0.6065, \quad e^{-4} = 0.0183, \quad e^{-1} = 0.3679, \quad e^{-0.3} = 0.7408$$

1. If X is Poisson with mean $\lambda = 4$, find
   (a) $P\{X = 1\}$
   (b) $P\{X = 2\}$
   (c) $P\{X > 2\}$
2. Compare the Poisson approximation with the true binomial probability in the following cases:
   (a) $P\{X = 2\}$ when $n = 10, p = 0.1$
   (b) $P\{X = 2\}$ when $n = 10, p = 0.05$
   (c) $P\{X = 2\}$ when $n = 10, p = 0.01$
   (d) $P\{X = 2\}$ when $n = 10, p = 0.3$
3. You buy a lottery ticket in 500 lotteries. In each lottery your chance of winning a prize is 1/1000. What is the approximate probability for the following?
   (a) You win 0 prizes.
   (b) You win exactly 1 prize.
   (c) You win at least 2 prizes.
4. If X is Poisson with mean $\lambda = 144$, find
   (a) $E[X]$
   (b) $\mathrm{SD}(X)$

5. A particular insurance company pays out an average of 4 major medical claims in a month.
   (a) Approximate the probability that it pays no major medical claims in the coming month?
   (b) Approximate the probability that it pays at most 2 major medical claims in the coming month?
   (c) Approximate the probability that it pays at least 4 major medical claims in the coming month?

KEY TERMS

Random variable: A quantity whose value is determined by the outcome of a probability experiment.

Discrete random variable: A random variable whose possible values constitute a sequence of disjoint points on the number line.

Expected value of a random variable: A weighted average of the possible values of a random variable; the weight given to a value is the probability that the random variable is equal to that value. Also called the **expectation** or the **mean** of the random variable.

Variance of a random variable: The expected value of the square of the difference between the random variable and its expected value.

Standard deviation of a random variable: The square root of the variance.

Independent random variables: A set of random variables having the property that knowing the values of any subset of them does not affect the probabilities of the remaining ones.

Binomial random variable with parameters n and p: A random variable equal to the number of successes in n independent trials when each trial is a success with probability p.

SUMMARY

A *random variable* is a quantity whose value is determined by the outcome of a probability experiment. If its possible values can be written as a sequence of distinct numbers, then the random variable is called *discrete*.

Let X be a random variable whose possible values are x_i, $i = 1, \ldots, n$; and suppose X takes on the value x_i with probability $P\{X = x_i\}$. The *expected value* of X, also referred to as the *mean* of X or the *expectation* of X, is denoted by $E[X]$ and is defined as

$$E[X] = \sum_{i=1}^{n} x_i P\{X = x_i\}$$

If X is a random variable and c is a constant, then

$$E[cX] = cE[X]$$
$$E[X + c] = E[X] + c$$

For any random variables $X_1, \ldots, X_k$,

$$E[X_1 + X_2 + \cdots + X_k] = E[X_1] + E[X_2] + \cdots + E[X_k]$$

The random variables X and Y are *independent* if knowing the value of one of them does not change the probabilities for the other.

The *variance* of a random variable measures the average squared distance of the random variable from its mean. Specifically, if X has mean $\mu = E[X]$, then the variance of X, denoted by $\mathrm{Var}(X)$, is defined as

$$\mathrm{Var}(X) = E[(X - \mu)^2]$$

A property of the variance is that for any constant c and random variable X,

$$\mathrm{Var}(cX) = c^2\mathrm{Var}(X)$$
$$\mathrm{Var}(X + c) = \mathrm{Var}(X)$$

Whereas the variance of the sum of random variables in general is not equal to the sum of their variances, it is true in the special case where the random variables are independent. That is,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

if X and Y are independent.

The square root of the variance is called the *standard deviation* and is denoted by $\mathrm{SD}(X)$. That is,

$$\mathrm{SD}(X) + \sqrt{\mathrm{Var}(X)}$$

Consider n independent trials in which each trial results in a success with probability p. If X is the total number of successes, then X is said to be a *binomial* random variable with parameters n and p. Its probabilities are given by

$$P\{X = i\} = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} \quad i = 0, \ldots, n$$

In the above, $n!$ (called *n factorial*) is defined by

$$0! = 1 \quad n! = n(n-1) \ldots 3 \cdot 2 \cdot 1$$

The mean and variance of a binomial random variable with parameters n and p are

$$E[X] = np \quad \text{and} \quad \text{Var}(X) = np(1-p)$$

Binomial random variables with large values of n and small values of p can be approximated by a *Poisson* random variable, whose probabilities are given by

$$P\{X = i\} = e^{-\lambda}\frac{\lambda^i}{i!}, \quad i = 0, 1, \ldots$$

where $\lambda = np$. The mean and variance of this random variable are both equal to λ.

REVIEW PROBLEMS

1. If $P\{X \leq 4\} = 0.8$ and $P\{X = 4\} = 0.2$, find
   (a) $P\{X \geq 4\}$
   (b) $P\{X < 4\}$
2. If $P\{X \leq 6\} = 0.7$ and $P\{X < 6\} = 0.5$, find
   (a) $P\{X = 6\}$
   (b) $P\{X > 6\}$
3. A graduating law student is not certain whether he actually wants to practice law or go into business with his family. He has decided to base his decision on whether he can pass the bar examination. He has decided to give himself at most 4 attempts at the examination; he will practice law if he passes the examination or go into the family business if he fails on all 4 tries. Suppose that each time he takes the bar examination he is successful, independent of his previous results, with probability 0.3. Let X denote the number of times he takes the bar examination.
   (a) What are the possible values that X can assume?
   (b) What is the probability distribution of X?
   (c) What is the probability that he passes the bar examination?
   (d) Find $E[X]$.
   (e) Find $\text{Var}(X)$.
4. Suppose that X is either 1 or 2. If $E[X] = 1.6$, find $P\{X = 1\}$.
5. A gambling book recommends the following "winning strategy" for the game of roulette. It recommends that a gambler bet \$1 on red. If red appears (which has probability 18/38 of occurring), then the gambler should take her \$1 profit and quit. If the gambler loses this bet (which has probability 20/38 of occurring) and so is behind \$1, then she should make a \$2 bet on red and then quit. Let X denote the gambler's final winnings.

(a) Find $P\{X > 0\}$.

(b) Are you convinced that the strategy is a winning strategy? Why or why not?

(c) Find $E[X]$.

6. Two people are to meet in the park. Each person is equally likely to arrive, independent of the other, at 3:00, 4:00, or 5:00 p.m. Let X equal the time that the first person to arrive has to wait, where X is taken to equal 0 if they both arrive at the same time. Find $E[X]$.

7. There is a 0.3 probability that a used-car salesman will sell a car to his next customer. If he does, then the car that is purchased is equally likely to cost \$4000 or \$6000. Let X denote the amount of money that the customer spends.

(a) Find the probability distribution of X.

(b) Find $E[X]$.

(c) Find $\text{Var}\,(X)$.

(d) Find $\text{SD}\,(X)$.

8. Suppose that 2 batteries are randomly chosen from a bin containing 12 batteries, of which 8 are good and 4 are defective. What is the expected number of defective batteries chosen?

9. A company is preparing a bid for a contract to supply a city's schools with notebook supplies. The cost to the company of supplying the material is \$140,000. It is considering two alternate bids: to bid high (25 percent above cost) or to bid low (10 percent above cost). From past experience the company knows that if it bids high, then the probability of winning the contract is 0.15, whereas if it bids low, then the probability of winning the contract is 0.40. Which bid will maximize the company's expected profit?

10. If $E[3X + 10] = 70$, what is $E[X]$?

11. The probability that a vacuum cleaner saleswoman makes no sales today is 1/3, the probability she makes 1 sale is 1/2, and the probability she makes 2 sales is 1/6. Each sale made is independent and equally likely to be either a standard cleaner, which costs \$500, or a deluxe cleaner, which costs \$1000. Let X denote the total dollar value of all sales.

(a) Find $P\{X = 0\}$.

(b) Find $P\{X = 500\}$.

(c) Find $P\{X = 1000\}$.

(d) Find $P\{X = 1500\}$.

(e) Find $P\{X = 2000\}$.

(f) Find $E[X]$.

(g) Suppose that the saleswoman receives a 20 percent commission on the sales that she makes. Let Y denote the amount of money she earns. Find $E[Y]$.

12. The 5 families living on a certain block have a total of 12 children. One of the families has 4 children, one has 3, two have 2, and one has 1. Let X denote the number of children in a randomly selected family, and let Y denote the number of children in the family of a randomly selected child. That is, X refers to an experiment in which each of the five families is equally likely to be selected, whereas Y refers to one in which each of the 12 children is equally likely to be selected.
    (a) Which do you think has the larger expected value, X or Y?
    (b) Calculate $E[X]$ and $E[Y]$.

13. A financier is evaluating two investment possibilities. Investment A will result in

\$200,000 profit	with probability 1/4
\$100,000 profit	with probability 1/4
\$150,000 loss	with probability 1/2

Investment B will result in

\$300,000 profit	with probability 1/8
\$200,000 profit	with probability 1/4
\$150,000 loss	with probability 3/8
\$400,000 loss	with probability 1/4

    (a) What is the expected profit of investment A?
    (b) What is the expected profit of investment B?
    (c) What is the investor's expected profit if she or he invests in both A and B?

14. If $\text{Var}(X) = 4$, find
    (a) $\text{Var}(2X + 14)$
    (b) $\text{SD}(2X)$
    (c) $\text{SD}(2X + 14)$

15. Suppose $E[X] = \mu$ and $\text{SD}(X) = \sigma$. Let

$$Y = \frac{X - \mu}{\sigma}$$

    (a) Show that $E[Y] = 0$.
    (b) Show that $\text{Var}(Y) = 1$.

The random variable Y is called the *standardized* version of X. That is, given a random variable, if we subtract its expected value and divide the result by its standard deviation, then the resulting random variable is said to be standardized. The standardized variable has expected value 0 and variance 1.

16. A manager has two clients. The gross annual earnings of his first client are a random variable with expected value \$200,000 and standard deviation \$60,000. The gross annual earnings of his second client are a random variable with expected value \$140,000 and standard deviation \$50,000. If the manager's fee is 15 percent of his first client's gross earnings and 20 percent of his second client's gross earnings, find the
    (a) Expected value of the manager's fee
    (b) Standard deviation of the manager's total fee

    In part (b) assume that the earnings of the two clients are independent.

17. A weighted coin that comes up heads with probability 0.6 is flipped n times. Find the probability that the total number of heads in these flips exceeds the total number of tails when
    (a) $n = 1$ (b) $n = 3$ (c) $n = 5$ (d) $n = 7$ (e) $n = 9$ (f) $n = 19$

18. Each customer who enters a television store will buy a normal-size television with probability 0.3, buy an extra-large television with probability 0.1, or not buy any television with probability 0.6. Find the probability that the next 5 customers
    (a) Purchase a total of 3 normal-size sets
    (b) Do not purchase any extra-large sets
    (c) Purchase a total of 2 sets

19. A saleswoman has a 60 percent chance of making a sale each time she visits a computer store. She visits 3 stores each month. Assume that the outcomes of successive visits are independent.
    (a) What is the probability she makes no sales next month?
    (b) What is the probability she makes 2 sales next month?
    (c) What is the probability that she makes at least 1 sale in each of the next 3 months?

20. Let X be a binomial random variable such that

$$E[X] = 6 \quad \text{and} \quad \text{Var}\,(X) = 2.4$$

    Find
    (a) $P\{X > 2\}$ (b) $P\{X \leq 9\}$ (c) $P\{X = 12\}$

21. A coin that comes up heads with probability 1/3 is to be flipped 3 times. Which is more likely: that heads appears exactly once, or that it does not appear exactly once?