

### > TRY IT 4.15

During a certain NBA season, a player for the Los Angeles Clippers had the highest field goal completion rate in the league. This player scored with 61.3% of his shots. Suppose you choose a random sample of 80 shots made by this player during the season. Let  $X$  = the number of shots that scored points.

- What is the probability distribution for  $X$ ?
- Using the formulas, calculate the (i) mean and (ii) standard deviation of  $X$ .
- Use your calculator to find the probability that this player scored with 60 of these shots.
- Find the probability that this player scored with more than 50 of these shots.

### EXAMPLE 4.16

The following example illustrates a problem that is **not** binomial. It violates the condition of independence. ABC College has a student advisory committee made up of ten staff members and six students. The committee wishes to choose a chairperson and a recorder. What is the probability that the chairperson and recorder are both students? The names of all committee members are put into a box, and two names are drawn **without replacement**. The first name drawn determines the chairperson and the second name the recorder. There are two trials. However, the trials are not independent because the outcome of the first trial affects the outcome of the second trial. The probability of a student on the first draw is  $\frac{6}{16}$ . The probability of a student on the second draw is  $\frac{5}{15}$ , when the first draw selects a student. The probability is  $\frac{6}{15}$ , when the first draw selects a staff member. The probability of drawing a student's name changes for each of the trials and, therefore, violates the condition of independence.

### > TRY IT 4.16

A lacrosse team is selecting a captain. The names of all the seniors are put into a hat, and the first three that are drawn will be the captains. The names are not replaced once they are drawn (one person cannot be two captains). You want to see if the captains all play the same position. State whether this is binomial or not and state why.

## 4.4 Geometric Distribution

There are four main characteristics of a geometric experiment.

- A trial is repeated until a success occurs. Think of this as one or more Bernoulli trials with all failures except the last one, which is a success. In other words, you keep repeating what you are doing until the first success. Then you stop. For example, you throw a dart at a bullseye until you hit the bullseye. The first time you hit the bullseye is a "success," so you stop throwing the dart. It might take six tries until you hit the bullseye. You can think of the trials as failure, failure, failure, failure, failure, success, STOP. In theory, the number of trials could go on forever.
- The repeated trials are independent of each other.
- The probability,  $p$ , of a success and the probability,  $q$ , of a failure is the same for each trial.  $p + q = 1$  and  $q = 1 - p$ . For example, the probability of rolling a three when you throw one fair die is  $1/6$ . This is true no matter how many times you roll the die. Suppose you want to know the probability of getting the first three on the fifth roll. On rolls one through four, you do not get a face with a three. The probability for each of the rolls is  $q = 5/6$ , the probability of a failure. The probability of getting a first three on the fifth roll is  $(5/6)(5/6)(5/6)(5/6)(1/6) = 0.0804$ .
- The random variable  $X$  represents the number of the trial in which the first success occurs. That is,  $X$  = the number of independent trials until the first success.

The following are additional attributes of the geometric distribution:

- The random variable is discrete. The random variable may be defined in two ways depending upon the analyst's interest. In the example above for throwing a die, the question was "What is the probability that the first success will be on the fifth throw?" Alternatively, the question could be asked as "What is the probability that it takes four failures before a success?" We will see that each way of asking the question will alter the form of the geometric probability density function slightly and will change the mean and standard deviation of the geometric pdf.

6. Implicit in the random variable is that the probability of a success is constant and therefore so is the probability of a failure. For flipping a coin this is obvious, but in experiments that require skill, such as hitting a baseball or throwing a dart, one might consider that learning during the experiment would alter the probability of a success. The geometric distribution cannot capture "learning" thus the historical probability of success is assumed to be constant.
7. The geometric distribution is "memoryless." There are very few probability density functions that are what is known as "memoryless," and the Geometric distribution is the only one with a discrete random variable that is memoryless. As an example, historically Major League Baseball player Jones has a record of hitting the ball for at least an advance to first base with a probability of 0.20. Jones has not had a hit in his last 10 times at bat. What is the probability that Jones will get a hit in his third time at bat? The answer ignores his 10 previous failures. All events prior to the events in current time are irrelevant and thus are considered "memoryless."

Formally:  $P(x = n + k | x \geq k + 1) = P(x = n)$ , where  $k$  = number of previous failures

Jones's probability of a hit begins anew each time he comes to bat. This feature of the geometric distribution results in a curious result: Drawing parts from a manufacturing process to test for parts that are defective, the geometric distribution begins with a clean slate each time the tests begin with no consideration of previous test results. More on this when we get to the exponential probability density function.

#### EXAMPLE 4.17

You play a game of chance that you can either win or lose (there are no other possibilities) **until** you lose. Your probability of losing is  $p = 0.57$ . What is the probability that it takes five games until you lose? Let  $X$  = the number of games you play until you lose (includes the losing game). Then  $X$  takes on the values 1, 2, 3, ... (could go on indefinitely). The probability question is  $P(x = 5)$ .

#### > TRY IT 4.17

You throw darts at a board until you hit the center area. Your probability of hitting the center area is  $p = 0.17$ . You want to find the probability that it takes eight throws until you hit the center. What values does  $X$  take on?

#### EXAMPLE 4.18

##### ? Problem

A safety engineer feels that 35% of all industrial accidents in the plant are caused by failure of employees to follow instructions. They decide to look at the accident reports (selected randomly and replaced in the pile after reading) **until** they find one that shows an accident caused by failure of employees to follow instructions. On average, how many reports would the safety engineer **expect** to look at until they find a report showing an accident caused by employee failure to follow instructions? What is the probability that the safety engineer will have to examine at least three reports until they find a report showing an accident caused by employee failure to follow instructions?

Let  $X$  = the number of accidents the safety engineer must examine **until** they find a report showing an accident caused by employee failure to follow instructions.  $X$  takes on the values 1, 2, 3, .... The first question asks you to find the **expected value** or the mean. The second question asks you to find  $P(x \geq 3)$ . ("At least" translates to a "greater than or equal to" symbol).

##### ✓ Solution

- a. On average, how many reports would the safety engineer **expect** to review until they find a report showing an accident caused by something other than failure to follow instructions? This question asks you to find the **expected value** or the mean. This is the average number of failures from something other than following instructions. The formula for the mean of this geometric distribution is:

$$\mu = E(x) = \frac{1}{p} \approx 2.8$$

Note that the mean does not need to be a whole number although the random variable is discrete and must be a counting number

- b. What is the probability that the safety engineer will have to examine at least three reports until they find a report

showing an accident caused by employee failure to follow instructions?

This question is answered by first defining the random variable. Let  $X$  = the number of accidents the safety engineer must examine **until** they find a report showing an accident caused by employee failure to follow instructions.  $X$  can take on the values 1, 2, 3, 4, ...  $\infty$ . Unlike the binomial distribution with a fixed number of trials, the geometric distribution may have an infinite number of failed trials before a success occurs.

This second question asks you to find  $P(x \geq 3)$ . ("At least" translates to a "greater than or equal to" symbol.)

In a binomial distribution, we saw the solution to questions of "more than" or "less than" would be to calculate each probability individually and add from zero to three. If the probability of interest is "greater than or equal to" ( $\geq$ ) the individual probabilities are added from zero to three and then subtracted from one.

$$P(x \geq 3) = 1 - P(x < 3)$$

Because we cannot add the full range of the geometric distribution random variable because it goes through infinity, an alternative solution is developed. An alternative geometric distribution pair of formulas provides solutions for questions asking for probabilities "more than" and "less than."

If the question is:

What is the probability it takes MORE THAN  $n$  events to get first success?

$$P(x > n) = (1 - p)^n$$

What is the probability it takes LESS THAN  $n$  event for success?

$$P(x < n) = 1 - (1 - p)^n$$

#### TRY IT 4.18

An instructor feels that 15% of students get below a C on their final exam. They decide to look at final exams (selected randomly and replaced in the pile after reading) until they find one that shows a grade below a C. We want to know the probability that the instructor will have to examine at least ten exams until they find one with a grade below a C. What is the probability question stated mathematically?

#### EXAMPLE 4.19

Suppose that you are looking for a student at your college who lives within five miles of you. You know that 55% of the 25,000 students do live within five miles of you. You randomly contact students from the college **until** one says they live within five miles of you. What is the probability that you need to contact four people?

This is a geometric problem because you may have a number of failures before you have the one success you desire. Also, the probability of a success stays the same each time you ask a student if they live within five miles of you. There is no definite number of trials (number of times you ask a student).

#### Problem

- Let  $X$  = the number of \_\_\_\_\_ you must ask \_\_\_\_\_ one says yes.
- What values does  $X$  take on?
- What are  $p$  and  $q$ ?
- The probability question is  $P(\text{_____})$ .

#### Solution

- Let  $X$  = the number of **students** you must ask **until** one says yes.
- 1, 2, 3, ..., (total number of students)
- $p = 0.55$ ;  $q = 0.45$
- $P(x = 4)$

### > TRY IT 4.19

You need to find a store that carries a special printer ink. You know that of the stores that carry printer ink, 10% of them carry the special ink. You randomly call each store until one has the ink you need. What are  $p$  and  $q$ ?

## Notation for the Geometric: $G = \text{Geometric Probability Distribution Function}$

$$X \sim G(p)$$

Read this as "X is a random variable with a **geometric distribution**." The parameter is  $p$ ;  $p$  = the probability of a success for each trial.

### CASE I: Random Variable X Is Event of First Success

In this case we ask, "What is the probability that we will have some number  $x$  of events of interest to us of failures before a success?"

The geometric pdf tells us the probability that the first occurrence of success requires  $x$  number of failure independent trials, each with probability  $(1 - p)$ . If the probability of success on each trial is  $p$ , then the probability that the  $x$ th trial (out of  $x$  trials) is the first success is:

$$P(X = x) = (1 - p)^{x-1} p$$

for  $x = 1, 2, 3, \dots$

Like the binomial distribution, the geometric distribution has the parameters of the mean and standard deviation. The expected value of  $X$ , the mean for Case I, is  $\mu = \frac{1}{p}$ . This tells us how many failed trials to expect until we get the first success. This count includes in the count of trials the trial that results in success. The above form of the geometric distribution is used for modeling the number of trials until the first success. The number of trials includes the one that is a success:  $x$  = all trials including the one that is a success. This can be seen in the form of the formula. If  $X$  = number of trials including the success, then we must multiply the probability of failure,  $(1 - p)$ , times the number of failures, that is  $x - 1$ . The standard deviation of Case I of the geometric distribution is:

$$\sigma = \sqrt{\left(\frac{1}{p}\right)\left(\frac{1}{p} - 1\right)}$$

### CASE II: Random Variable X Is Number of Failures BEFORE a Success

By contrast to Case I, the following form of the geometric distribution used for modeling number of failures until the first success is:

$$P(X = x) = (1 - p)^x p$$

for  $x = 0, 1, 2, 3, \dots$

In this case the trial that is the success is not counted as a trial in the formula:  $x$  = number of failures. The expected value, the mean, of this distribution is  $\mu = \frac{1-p}{p}$ . This tells us how many failures to expect before we have a success. In either case, the sequence of probabilities is a geometric sequence. In Case II, the standard deviation parameter is:

$$\sigma = \sqrt{\left(\frac{1-p}{p}\right)}.$$

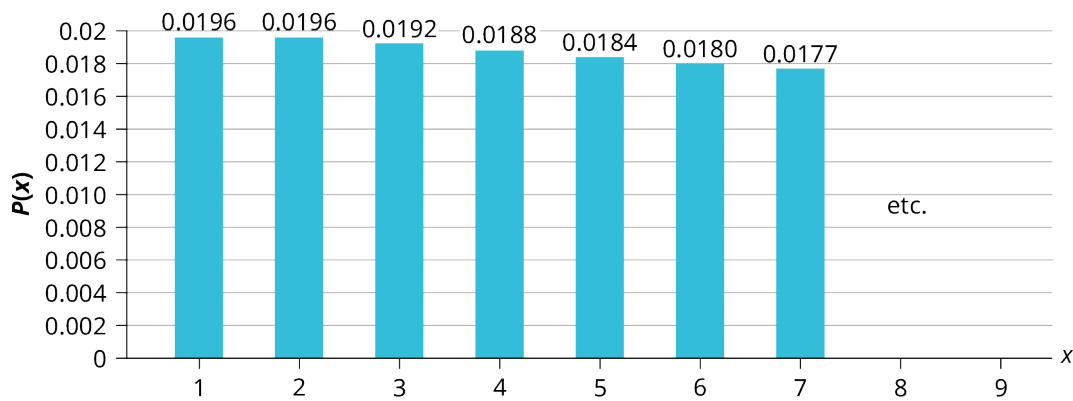


Figure 4.4

GEOMETRIC:

 $P = .02$ , Common ratio  $= r = .98$ 

The y-axis in Figure 4.4 contains the probability of  $x$ , and the x-axis is the random number components tested. For example, at  $x = 1$  the probability it will be found to be defective is 0.0196. With two components tested, the probability the second component is defective is graphed at a probability of 0.0196 at  $x = 2$  on the x axis. For the probability that the third component is defective we find  $P(X = 3) = 0.019208$ . (The first two are the same because of rounding in the computations.)

Notice on Figure 4.4 that the probabilities decline by the same step down with each change in the value of  $x$ . This increment is called the common ratio. This exists for the geometric probability distribution uniquely. The common ratio, called  $r$ , can be calculated by dividing any value by the previous value, e.g.,  $\frac{P(x=5)}{P(x=4)} = \frac{0.018447}{0.018823} = 0.98$ .

For this set of data, therefore, the common ratio is 0.98. The common ratio then multiplied by any other probability value will provide the next probability value in the sequence. For example, the probability that the sixth component tested is a failure is 0.018078. Check this using the formula from Case I. Now we have the  $P(x = 6)$ . Knowing this and the common ratio we can calculate the  $P(x = 7)$  by simple multiplication:  $P(x = 7) = (0.018078) * 0.98 = 0.017716$ , the same value we found earlier by using the geometric probability distribution. This common ratio increment is the same ratio between every number and is called a geometric progression and thus the name for this probability density function. Once the common ratio is calculated, any  $P(x = x_a)$  one desires to know can be easily found.

The number of components that you would expect to test until you find the first defective component is the mean,  $\mu = 50$  for this case of defective components. The formula for the mean of the geometric distribution for the random variable defined as number of failures before first success is  $\mu = \frac{1}{p} = \frac{1}{0.02} = 50$

See Example 4.20 for an example where the geometric random variable is defined as number of trials until first success. The expected value of this formula for the geometric distribution will be different from this version of the distribution. Case II also has a variance, but this is changed from the Case I formula. This formula for the variance is:

$$P(X = x) = (1 - p)^{x-1} p$$

for  $x = 1, 2, 3, \dots$

$$\text{The formula for the variance is } \sigma^2 = \left(\frac{1}{p}\right) \left(\frac{1}{p} - 1\right) = \left(\frac{1}{0.02}\right) \left(\frac{1}{0.02} - 1\right) = 2,450$$

$$\text{The standard deviation is } \sigma = \sqrt{\left(\frac{1}{p}\right) \left(\frac{1}{p} - 1\right)} = \sqrt{\left(\frac{1}{0.02}\right) \left(\frac{1}{0.02} - 1\right)} = 49.5$$

**EXAMPLE 4.20**

Assume that the probability of a defective computer component is 0.02. Components are randomly selected. Find the probability that the first defect is caused by the seventh component tested. How many components do you expect to test until one is found to be defective?

Let  $X$  = the number of computer components tested until the first defect is found.

$X$  takes on the values 1, 2, 3, ... where  $p = 0.02$ .

Find  $P(x = 7)$ .  $P(x = 7) = 0.0177$ .



#### USING THE TI-83, 83+, 84, 84+ CALCULATOR

To find the probability that  $x = 7$ ,

- Enter 2<sup>nd</sup>, DISTR
- Scroll down and select `geometpdf(`
- Press ENTER
- Enter 0.02, 7); press ENTER to see the result:  $P(x = 7) = 0.0177$

To find the probability that  $x \leq 7$ , follow the same instructions EXCEPT select `E:geometcdf(` as the distribution function.

The probability that the seventh component is the first defect is 0.0177.

The formula for the variance is  $\sigma^2 = \left(\frac{1}{p}\right) \left(\frac{1}{p} - 1\right) = \left(\frac{1}{0.02}\right) \left(\frac{1}{0.02} - 1\right) = 2,450$

The standard deviation is  $\sigma = \sqrt{\left(\frac{1}{p}\right) \left(\frac{1}{p} - 1\right)} = \sqrt{\left(\frac{1}{0.02}\right) \left(\frac{1}{0.02} - 1\right)} = 49.5$



#### TRY IT 4.20

The probability of a defective steel rod is 0.01. Steel rods are selected at random. Find the probability that the first defect occurs on the ninth steel rod. Use the TI-83+ or TI-84 calculator to find the answer.

#### EXAMPLE 4.21

##### Problem

The lifetime risk of developing cancer is about one in 67 (1.5%). Let  $X$  = the number of people you ask until one says they have cancer. Then  $X$  is a discrete random variable with a geometric distribution:  $X \sim G\left(\frac{1}{67}\right)$  or  $X \sim G(0.015)$

- What is the probability of that you ask ten people before one says they have cancer?
- What is the probability that you must ask 20 people?
- Find the (i) mean and (ii) standard deviation of  $X$ .

##### Solution

- $P(x = 10) = \text{geometricpdf}(0.015, 10) = 0.0131$
- $P(x = 20) = \text{geometricpdf}(0.015, 20) = 0.0113$
- Mean  $= \mu = \frac{1}{p} = \frac{1}{0.015} = 66.67$

$$\text{Standard Deviation} = \sigma = \sqrt{\frac{1-p}{p^2}} = \sqrt{\frac{1-0.015}{[0.015]^2}} = 66.16$$



#### TRY IT 4.21

The literacy rate for a nation measures the proportion of people age 15 and over who can read and write. The literacy rate for women in Afghanistan is 12%. Let  $X$  = the number of Afghani women you ask until one says that she is literate.

- What is the probability distribution of  $X$ ?
- What is the probability that you ask five women before one says she is literate?

- c. What is the probability that you must ask ten women?
- d. Find the (i) mean and (ii) standard deviation of  $X$ .

## 4.5 Hypergeometric Distribution

There are five characteristics of a hypergeometric experiment.

1. You take samples from **two** groups.
2. You are concerned with a group of interest, called the first group.
3. You sample **without replacement** from the combined groups. For example, you want to choose a softball team from a combined group of 11 men and 13 women. The team consists of ten players.
4. Each pick is **not** independent, since sampling is without replacement. In the softball example, the probability of picking a woman first is  $\frac{13}{24}$ . The probability of picking a man second is  $\frac{11}{23}$  if a woman was picked first. It is  $\frac{10}{23}$  if a man was picked first. The probability of the second pick depends on what happened in the first pick.
5. You are **not** dealing with Bernoulli Trials.

The outcomes of a hypergeometric experiment fit a **hypergeometric probability** distribution. The random variable  $X$  = the number of items from the group of interest.

### EXAMPLE 4.22

#### Problem

A candy dish contains 100 jelly beans and 80 gumdrops. Fifty candies are picked at random. What is the probability that 35 of the 50 are gumdrops? The two groups are jelly beans and gumdrops. Since the probability question asks for the probability of picking gumdrops, the group of interest (first group) is gumdrops. The size of the group of interest (first group) is 80. The size of the second group is 100. The size of the sample is 50 (jelly beans or gumdrops). Let  $X$  = the number of gumdrops in the sample of 50.  $X$  takes on the values  $x = 0, 1, 2, \dots, 50$ . What is the probability statement written mathematically?

#### Solution

$$P(x = 35)$$



### TRY IT 4.22

A bag contains letter tiles. Forty-four of the tiles are vowels, and 56 are consonants. Seven tiles are picked at random. You want to know the probability that four of the seven tiles are vowels. What is the group of interest, the size of the group of interest, and the size of the sample?

### EXAMPLE 4.23

#### Problem

Suppose a shipment of 100 laptops is known to have ten defective laptops. An inspector randomly chooses 12 for inspection. He is interested in determining the probability that, among the 12 laptops, at most two are defective. The two groups are the 90 non-defective laptops and the 10 defective laptops. The group of interest (first group) is the defective group because the probability question asks for the probability of at most two defective laptops. The size of the sample is 12 laptops. (They may be non-defective or defective.) Let  $X$  = the number of defective laptops in the sample of 12.  $X$  takes on the values 0, 1, 2, ..., 10.  $X$  may not take on the values 11 or 12. The sample size is 12, but there are only 10 defective laptops. Write the probability statement mathematically.

#### Solution

$$P(x \leq 2)$$

### > TRY IT 4.23

A gross of eggs contains 144 eggs. A particular gross is known to have 12 cracked eggs. An inspector randomly chooses 15 for inspection. She wants to know the probability that, among the 15, at most three are cracked. What is  $X$ , and what values does it take on?

### EXAMPLE 4.24

You are president of an on-campus special events organization. You need a committee of seven students to plan a special birthday party for the president of the college. Your organization consists of 18 women and 15 men. You are interested in the number of men on your committee. If the members of the committee are randomly selected, what is the probability that your committee has more than four men?

This is a hypergeometric problem because you are choosing your committee from two groups (men and women).

#### ? Problem

- Are you choosing with or without replacement?
- What is the group of interest?
- How many are in the group of interest?
- How many are in the other group?
- Let  $X =$  \_\_\_\_\_ on the committee. What values does  $X$  take on?
- The probability question is  $P(\text{_____})$ .

#### ✓ Solution

- without
- the men
- 15 men
- 18 women
- Let  $X =$  **the number of men** on the committee.  $x = 0, 1, 2, \dots, 7$ .
- $P(x > 4)$

### > TRY IT 4.24

A palette has 200 milk cartons. Of the 200 cartons, it is known that ten of them have leaked and cannot be sold. A stock clerk randomly chooses 18 for inspection. He wants to know the probability that among the 18, no more than two are leaking. Give five reasons why this is a hypergeometric problem.

## Notation for the Hypergeometric: $H =$ Hypergeometric Probability Distribution Function

$$X \sim H(r, b, n)$$

Read this as "X is a random variable with a hypergeometric distribution." The parameters are  $r$ ,  $b$ , and  $n$ ;  $r$  = the size of the group of interest (first group),  $b$  = the size of the second group,  $n$  = the size of the chosen sample.



**EXAMPLE 4.25**

A school site committee is to be chosen randomly from six men and five women. If the committee consists of four members chosen randomly, what is the probability that two of them are men? How many men do you expect to be on the committee?

Let  $X$  = the number of men on the committee of four. The men are the group of interest (first group).

$X$  takes on the values 0, 1, 2, 3, 4, where  $r = 6$ ,  $b = 5$ , and  $n = 4$ .  $X \sim H(6, 5, 4)$

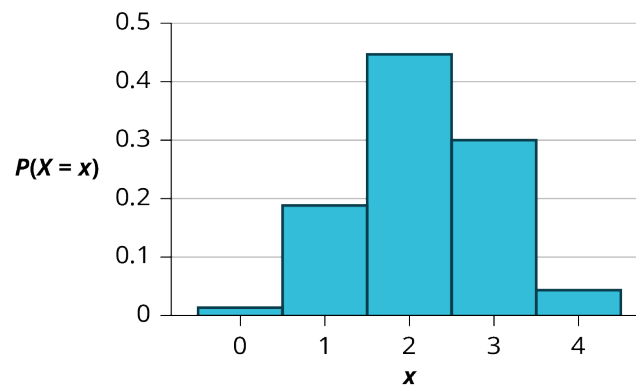
Find  $P(x = 2)$ .  $P(x = 2) = 0.4545$  (calculator or computer)

**NOTE**

Currently, the TI-83+ and TI-84 do not have hypergeometric probability functions. There are a number of computer packages, including Microsoft Excel, that do.

The probability that there are two men on the committee is about 0.45.

The graph of  $X \sim H(6, 5, 4)$  is:



**Figure 4.5**

The y-axis contains the probability of  $X$ , where  $X$  = the number of men on the committee.

You would expect  $m = 2.18$  (about two) men on the committee.

The formula for the mean is  $\mu = \frac{nr}{r+b} = \frac{(4)(6)}{6+5} = 2.18$

**TRY IT 4.25**

An intramural basketball team is to be chosen randomly from 15 boys and 12 girls. The team has ten slots. You want to know the probability that eight of the players will be boys. What is the group of interest and the sample?

## 4.6 Poisson Distribution

There are two main characteristics of a Poisson experiment.

1. The **Poisson probability distribution** gives the probability of a number of events occurring in a **fixed interval** of time or space if these events happen with a known average rate and independently of the time since the last event. For example, a book editor might be interested in the number of words spelled incorrectly in a particular book. It might be that, on the average, there are five words spelled incorrectly in 100 pages. The interval is the 100 pages.
2. The Poisson distribution may be used to approximate the binomial if the probability of success is "small" (such as 0.01) and the number of trials is "large" (such as 1,000). You will verify the relationship in the homework exercises.  $n$  is the number of trials, and  $p$  is the probability of a "success."

30.  $P(X = 4) = \underline{\hspace{2cm}}$
31.  $P(X < 4) = \underline{\hspace{2cm}}$
32. On average, how many years would you expect a child to study ballet with this teacher?
33. What does the column " $P(x)$ " sum to and why?
34. What does the column " $x \cdot P(x)$ " sum to and why?
35. You are playing a game by drawing a card from a standard deck and replacing it. If the card is a face card, you win \$30. If it is not a face card, you pay \$2. There are 12 face cards in a deck of 52 cards. What is the expected value of playing the game?
36. You are playing a game by drawing a card from a standard deck and replacing it. If the card is a face card, you win \$30. If it is not a face card, you pay \$2. There are 12 face cards in a deck of 52 cards. Should you play the game?

### 4.3 Binomial Distribution

Use the following information to answer the next eight exercises: The Higher Education Research Institute at UCLA collected data from 203,967 incoming first-time, full-time first-year students from 270 four-year colleges and universities in the U.S. 71.3% of those students replied that, yes, they believe that same-sex couples should have the right to legal marital status. Suppose that you randomly pick eight first-time, full-time first-year students from the survey. You are interested in the number that believes that same sex-couples should have the right to legal marital status.

37. In words, define the random variable  $X$ .
38.  $X \sim \underline{\hspace{1cm}}(\underline{\hspace{1cm}}, \underline{\hspace{1cm}})$
39. What values does the random variable  $X$  take on?
40. Construct the probability distribution function (PDF).

$x$	$P(x)$

Table 4.29

41. On average ( $\mu$ ), how many would you expect to answer yes?
42. What is the standard deviation ( $\sigma$ )?
43. What is the probability that at most five of the first-year students reply "yes"?
44. What is the probability that at least two of the first-year students reply "yes"?

### 4.4 Geometric Distribution

Use the following information to answer the next six exercises: The Higher Education Research Institute at UCLA collected data from 203,967 incoming first-time, full-time first-year students from 270 four-year colleges and universities in the U.S. 71.3% of those students replied that, yes, they believe that same-sex couples should have the right to legal marital status. Suppose that you randomly select first-year students from the study until you find one who replies "yes." You are interested in the number of first-year students you must ask.

45. In words, define the random variable  $X$ .
46.  $X \sim \underline{\hspace{1cm}}(\underline{\hspace{1cm}}, \underline{\hspace{1cm}})$

47. What values does the random variable  $X$  take on?
48. Construct the probability distribution function (PDF). Stop at  $x = 6$ .

$x$	$P(x)$
1	
2	
3	
4	
5	
6	

Table 4.30

49. On average ( $\mu$ ), how many first-year students would you expect to have to ask until you found one who replies "yes?"
50. What is the probability that you will need to ask fewer than three first-year students?

#### 4.5 Hypergeometric Distribution

Use the following information to answer the next five exercises: Suppose that a group of statistics students is divided into two groups: business majors and non-business majors. There are 16 business majors in the group and seven non-business majors in the group. A random sample of nine students is taken. We are interested in the number of business majors in the sample.

51. In words, define the random variable  $X$ .
52.  $X \sim \underline{\hspace{1cm}}(\underline{\hspace{1cm}}, \underline{\hspace{1cm}})$
53. What values does  $X$  take on?
54. Find the standard deviation.
55. On average ( $\mu$ ), how many would you expect to be business majors?

#### 4.6 Poisson Distribution

Use the following information to answer the next six exercises: On average, a clothing store gets 120 customers per day.

56. Assume the event occurs independently in any given day. Define the random variable  $X$ .
57. What values does  $X$  take on?
58. What is the probability of getting 150 customers in one day?
59. What is the probability of getting 35 customers in the first four hours? Assume the store is open 12 hours each day.
60. What is the probability that the store will have more than 12 customers in the first hour?
61. What is the probability that the store will have fewer than 12 customers in the first two hours?
62. Which type of distribution can the Poisson model be used to approximate? When would you do this?

Use the following information to answer the next six exercises: On average, eight teens in the U.S. die from motor vehicle injuries per day. As a result, states across the country are debating raising the driving age.

63. Assume the event occurs independently in any given day. In words, define the random variable  $X$ .