CHAPTER 4

Probability Distributions for Discrete Variables

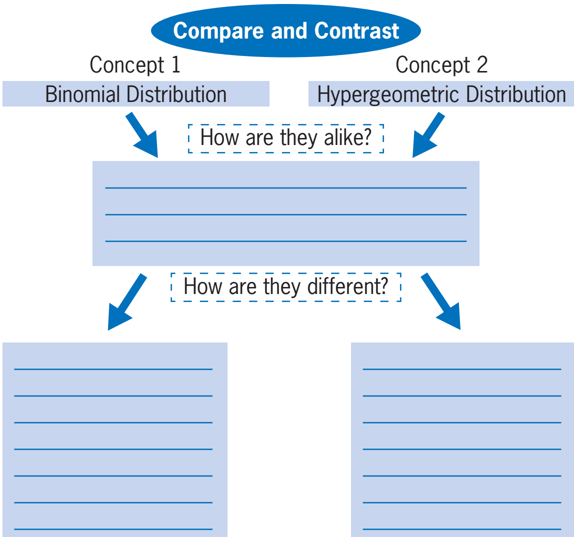
In the study of probability, the mathematical models used to analyse data are called distributions. They describe the probabilities that arise under particular situations. Insurance companies use probabilities to calculate the expected cost of health insurance for different age groups. Game show producers consider both probabilities and payouts when designing the rules of the game.

Key Terms

- probability distribution
- random variable
- discrete random variable
- continuous random variable
- probability histogram
- weighted mean
- expectation
- uniform distribution
- binomial probability distribution
- hypergeometric probability distribution

Literacy Strategy

A compare and contrast graphic organizer helps organize and connect ideas and compare concepts and strategies. As you complete the chapter, compare binomial and hypergeometric distributions.



Career Link

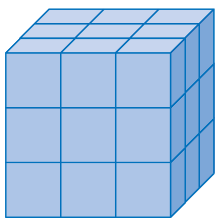
Meteorologist

A meteorologist uses probability and statistics to analyse weather patterns and make predictions. For example, under a given set of conditions, there may be an 80% chance of rain in about 50% of the area. The meteorologist multiplies the numbers to estimate a 40% chance of rain in the given area ($80\% \times 50\% = 40\%$). Meteorologists also often predict an expected amount of precipitation, such as 8 mm of rain over the next 24 hours. What other factors does a meteorologist consider, and how are they related to probability?

Chapter Problem

Painted Cube

A large cube is painted on all six faces. It is then cut into 27 smaller, congruent cubes as shown. These smaller cubes have a variety of paint patterns. You will calculate the probabilities associated with each number of painted faces. You will also design a simulation to help compare the theoretical to the statistical probability. How many patterns of painted faces do you think there are?



Prerequisite Skills

Simple Probability

1. When selecting a card from a standard deck, what is the probability that it is

a) a king?

b) a red card?

c) a spade?

2. An experiment involves rolling three dice and recording the sum.

a) Calculate the probability of each sum.

b) Verify that the sum of the probabilities is 1.

3. In an experiment, a lab rat searches for food behind one of eight doors. Three doors are red, and the remaining doors are green.

a) What is the probability that the food was placed behind a red door?

b) What is the probability that the rat will select the correct door?

Mutually Exclusive Events

4. A bag contains six red, one green, four blue, and three yellow marbles. A marble is selected at random. What is the probability that the marble is

a) red or green?

b) neither red nor blue?

5. If a deck containing only the face cards is shuffled and one card is selected, what is the probability that the card is

a) a queen or a king?

b) a red card or the queen of spades?

c) a red card and a spade?

Independent and Dependent Events

6. Classify each pair of events as independent or dependent.

a) rolling two dice

b) selecting two cards at the same time from a standard deck

c) flipping a head on one coin and a tail on another

d) selecting two males from a list of four males and five females

7. a) Use a tree diagram to illustrate the probabilities associated with the number of heads when three coins are flipped.

b) Are the events (head, tail, tail) independent or dependent? Explain.

8. A card game uses only the hearts. Players select two cards without replacement. What is the probability that a player will

a) select a queen followed by a king?

b) select a queen and a king?

c) not select a face card on either draw?

Combinations

9. Twelve men and 10 women apply to attend a special event. Six names are selected.

a) In how many ways could three men and three women be selected?

b) In how many ways could more men than women be selected?

10. A cookie jar contains three chocolate chip, four peanut butter, and six butterscotch cookies. Hansa reaches in and grabs a handful of five cookies. In how many ways could she select

a) two chocolate chip, two peanut butter, and one butterscotch cookie?

b) no chocolate chip cookies?

c) at least one of each type of cookie?

11. In how many ways could 15 different books be divided equally

a) among five different people?

b) among three different people?

Evaluating Expressions

12. Evaluate.

a) ${}_{12}C_5 \times {}_8C_3$ **b)** $\dfrac{{}_8C_5}{{}_{12}C_7}$

c) $\dfrac{{}_7C_3 \times {}_9C_5}{{}_{16}C_8}$ **d)** ${}_5C_2(0.6)^2(0.4)^3$

e) ${}_6C_4\left(\dfrac{1}{3}\right)^4\left(\dfrac{2}{3}\right)^2$

Binomial Theorem

13. Use the binomial theorem to expand.

Example:

$$(a + b)^3 = {}_3C_0a^3 + {}_3C_1a^2b + {}_3C_2ab^2 + {}_3C_3b^3$$
$$= a^3 + 3a^2b + 3ab^2 + b^3$$

a) $(x + y)^4$ **b)** $(4x + 3y)^5$

c) $(0.3 + 0.7)^6$ **d)** $\left(\dfrac{1}{4} + \dfrac{3}{4}\right)^5$

Graphing Calculator Keystrokes

14. Use a graphing calculator to make a frequency histogram of the data:

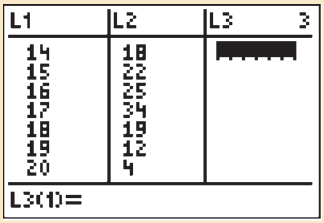
Homework Hours	0	1	2	3	4	5	6	7	8
Frequency	24	35	30	18	12	1	0	2	2

Example:

Student Age	14	15	16	17	18	19	20
Frequency	18	22	25	34	19	12	4

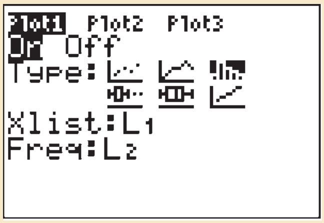
a) Setting up lists:

- To clear all lists, press **2ND** **MEM**. Then select **4:ClrAllLists** and press **ENTER**.
- To enter the data in the lists, press **STAT**, then select **1:Edit**.
- Enter the ages individually from 14 to 20 in list **L1**.
- Enter the frequency of each age in list **L2**.

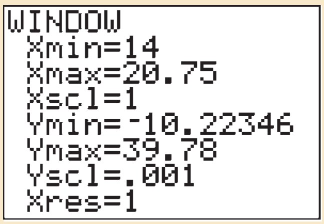


b) Graph a frequency histogram of the ages.

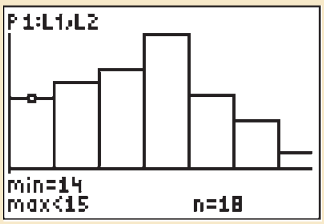
- To access the statistical plot, press **2ND** **STAT PLOT** and select **Plot 1**.
- Set the plot parameters as shown.



- To set the appropriate parameters for the graph, press **ZOOM**, then select **9:ZoomStat**.
- To adjust the scale so that the x-axis counts by 1, press **WINDOW** and change Xscl to 1.



- Press **TRACE** to see the frequency histogram and to read the values associated with each bar.



4.1

Probability Distributions

Learning Goals

I am learning to

- recognize and identify a discrete random variable
- generate a probability distribution by calculating the probabilities for all values of a random variable
- represent a probability distribution using a table and a probability histogram
- make connections between the frequency histogram and the probability histogram
- calculate and interpret the expected value for a probability distribution
- make connections between the expected value and the weighted mean of the values of the discrete random variable

Minds On...

When a seemingly rare event occurs, we often wonder, "What are the chances?"

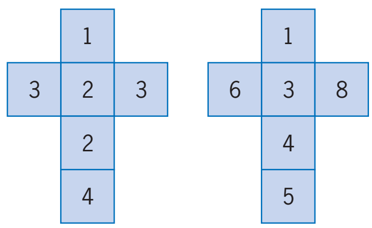
- In a family with five children, what is the probability of having three boys?
- What about all the other possibilities, such as five boys? Are they equally rare?
- When rolling two dice, double ones and double sixes are rare, but are they any rarer than other doubles or other combinations of the dice?

Action!

Investigate Probability Distributions

A relative frequency histogram shows the probability of each outcome as the height of each bar in the histogram.

Sicherman dice have faces as shown. They were invented by George Sicherman, an American military strategist. Consider all the possible sums when rolling the two dice.



1. Copy and complete the table showing the sums of the two dice.

		Die 1					
		1	**2**	**2**	**3**	**3**	**4**
Die 2	**1**						
	3						
	4						
	5						
	6						
	8						

2. Make a table of values showing the probability of each sum.

3. Construct a histogram to illustrate the probability distribution.

probability distribution

- the probabilities for all possible outcomes of an experiment or sample space
- often shown as a graph of probability versus the value of a **random variable**

random variable

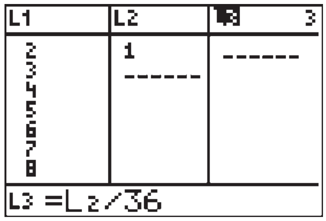
- a quantity that can have a range of values
- designated by a capital letter X, with individual values designated by a lower-case x

Method 1: Use a Graphing Calculator

Refer to the Prerequisite Skills on page 143 to review how to use a graphing calculator.

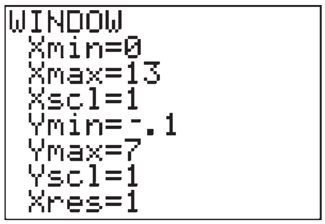
a) Enter the data into the lists.

- Enter the sums in list **L1**. Enter the frequencies in list **L2**.
- To calculate the probability of each sum, enter L2÷36 in the **L3** column heading.



b) Graph a frequency histogram of the sums.

- Press 2ND STAT PLOT and select **Plot 1**.
- To adjust the scale, select **9:ZoomStat**, then WINDOW. Set the scale as shown.
- Press TRACE to see the frequency histogram.
- Use the left and right arrows to see how the frequency (n = ?) changes for each sum.

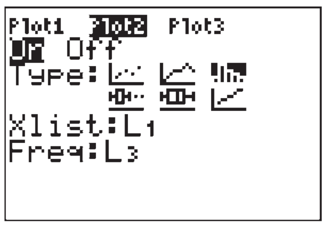


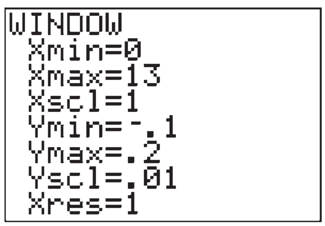
The parameters provide a range of x between 0 and 13 with a scale of 1 unit. The range of y is 0 to 7, but –0.1 is the lower limit so you can easily view the x-axis.

c) Describe the graph. What does it tell you?

d) Graph a relative frequency histogram.

- Press 2ND STAT PLOT, then select **Plot 1** and change it to **Off**.
- Select **Plot 2** and set the plot parameters as shown.
- To adjust the scale, select **9:ZoomStat**, then WINDOW. Set the scale as shown.
- Press TRACE to see the relative frequency histogram.
- Use the left and right arrows to see how the frequency (n = ?) changes for each sum.





Change the scale and range of y-values to reflect that relative frequencies are between 0 and 1.

Method 2: Use Fathom™

a) Set up the lists.

- Drag a new **Collection** to the workspace and name it Dice.
- With the Collection highlighted, drag a new **Table** down.
- Double click on **<new>** and rename it **SUM**.
- Enter all 36 sums in the list, making sure you enter each sum the appropriate number of times (e.g., 4 occurs three times).

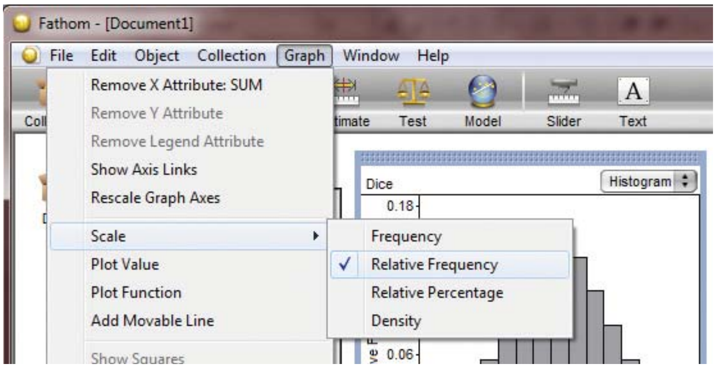
b) Graph a frequency histogram of the sums.

- With the **Collection** highlighted, drag a new **Graph** down.
- Drag the **SUM** attribute from the table to the x-axis of the graph.
- At the top right, select **Histogram**.

c) Describe the graph. What does it tell you?

d) Graph a relative frequency histogram.

- With the graph highlighted, in the **Graph** menu select **Scale**, then **Relative Frequency**.



4. Reflect How are the relative frequency histogram from step 3d) and the frequency histogram from step 3b) the same? different?

5. Extend Your Understanding How is this distribution similar to the probability distribution of the sum of two standard dice? How is it different?

discrete random variable

- a variable that can have only certain values within a given range, such as the sum of two dice

continuous random variable

- a variable that can have an infinite number of possible values in a given range, often measurements, such as volume or time

Discrete Sample Space

A discrete sample space is the set of all values of a discrete random variable. In the Investigation, you created a discrete probability distribution to show the probabilities of the discrete sample space. A discrete probability distribution maps each value x, of a discrete random variable X, to a corresponding probability.

Continuous Sample Space

A continuous sample space is the set of all values of a continuous random variable. A continuous probability distribution shows the probabilities of a continuous sample space. You will investigate continuous probability distributions in chapter 7.

Example 1

Constructing a Probability Histogram

The table gives the probability distribution of the number of digits in street addresses of a large city.

Number of Digits, x	Probability, $P(x)$
1	14%
2	31%
3	42%
4	11%
5	2%

a) Identify the random variable.

b) Construct a probability histogram.

c) Explain the meaning of the individual bars in the histogram.

d) Describe the distribution.

e) Calculate the sum of the probabilities. Comment on the result.

probability histogram

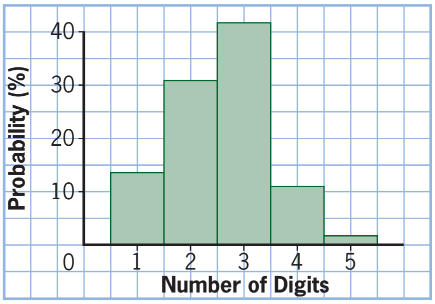
- a graph of a probability distribution in which equal intervals are marked on the horizontal axis and the probabilities associated with these intervals are indicated by the areas of the bars

Solution

a) The random variable, X, is the number of digits used in a street address.

b) Label the horizontal axis with the digits from 1 to 5, equally spaced. Centre each bar on the discrete variable it represents.

c) The area of each bar represents its probability. The width of each bar is 1, so the probability is shown on the vertical axis.



Why would the distribution look the way it does? Consider how street numbers are assigned.

d) Three-digit numbers occur most frequently, and the probability decreases as the digit value increases or decreases from three digits.

e) The sum of all the probabilities in any distribution is 1 because the distribution covers 100% of all cases.

Project Prep

Your probability project may involve a probability distribution. How do you know the difference between discrete and continuous random variables?

Your Turn

The table gives the percent breakdown of the number of rooms in apartments in a particular complex.

Number of Rooms, x	Percent, $P(x)$
2	15
3	30
4	42
5	10
6	3

a) Identify the random variable.

b) Construct a probability histogram.

c) Explain the meaning of the individual bars in the histogram.

d) Describe the distribution.

e) Calculate the sum of the probabilities. Does this confirm the results in the example?

Example 2

Expectation of a Probability Distribution

a) Make a tree diagram and show the probability distribution for the number of girls in a family of three children.

b) Make a probability histogram for this distribution.

c) Calculate the weighted mean number of girls in a "typical" family of three children.

d) Calculate the expectation for the number of girls in a family of three children. Compare it to the weighted mean.

e) Interpret the results in parts c) and d).

weighted mean

- the mean of a set of numbers that are given weightings based on their frequency
- multiply each number by its weight (or frequency) and divide by the sum of the weights

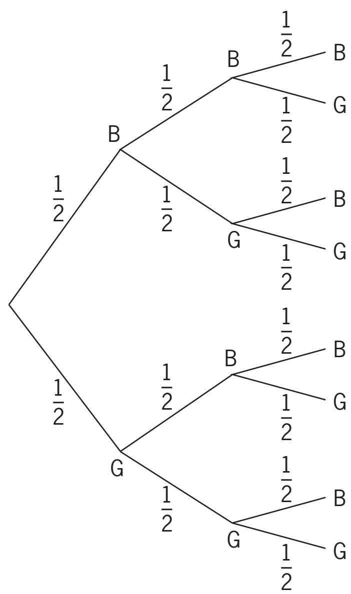
expectation (expected value)

- written $E(X)$
- $E(X)$ of a probability distribution is the predicted average of all possible outcomes
- $E(X)$ is equal to the sum of the products of each outcome, x, with its probability, $P(x)$
- $E(X) = \sum_{i=1}^{n} x_i \cdot P(x_i)$

Solution

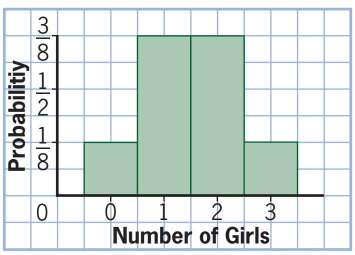
a) Boys and girls are equally likely, so each branch in the tree diagram has a probability of $\frac{1}{2}$.

There are eight strings of branches and each has a probability of $\left(\frac{1}{2}\right)^3 = \frac{1}{8}$.



Number of Girls, x	Distribution of Girls	Frequency	Probability, $P(x)$
0	BBB	1	$\frac{1}{8}$
1	GBB BGB BBG	3	$\frac{3}{8}$
2	GGB GBG BGG	3	$\frac{3}{8}$
3	GGG	1	$\frac{1}{8}$

b)



c) To determine the weighted mean, use the values in the table in part a). For each row, multiply the number of girls, x, by the frequency, since this is how many times the outcome can occur. Then, find the sum of the products and divide by the total frequency.

$$\begin{aligned}&\text{Weighted mean}\\ &= \frac{(x_1 \cdot \text{frequency}) + (x_2 \cdot \text{frequency}) + (x_3 \cdot \text{frequency}) + (x_4 \cdot \text{frequency})}{\text{total frequency}}\\ &= \frac{(0 \times 1) + (1 \times 3) + (2 \times 3) + (3 \times 1)}{8}\\ &= 1.5\end{aligned}$$

d) Add a fourth column to the chart and multiply each value of the random variable with its probability.

Number of Girls, x	Distribution of Girls	Frequency	Probability, $P(x)$	$x \cdot P(x)$
0	BBB	1	$\frac{1}{8}$	0
1	GBB BGB BBG	3	$\frac{3}{8}$	$\frac{3}{8}$
2	GGB GBG BGG	3	$\frac{3}{8}$	$\frac{6}{8}$
3	GGG	1	$\frac{1}{8}$	$\frac{3}{8}$
Sum		8		$\frac{12}{8}$

$$\begin{aligned}\sum_{x=0}^{3} x \cdot P(x) &= \frac{12}{8}\\ &= 1.5\end{aligned}$$

The expected number of girls in a family of three children is 1.5. The expectation equals the weighted mean of the outcomes.

e) On average, a family of three would have 1.5 girls.

Although it is impossible to have 1.5 children, do not round because this is a predicted average value.

Literacy Link

Capital sigma $\sum_{i=1}^{n} f(x_i)$ is a symbol for the sum of the values of the function $f(x)$. The limits below and above the sigma show that the sum is from the first term ($i = 1$) to the nth term.

For example, $\sum_{i=1}^{5} i^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2$.

Your Turn

A spinner has two equal sectors, coloured red and blue.

a) Make a tree diagram to show the probability distribution for the number of times the spinner lands on blue when it is spun four times.

b) Make a probability histogram for this distribution.

c) Calculate the expected number of times the spinner lands on blue.

d) Interpret the results in part c).

Example 3

Expected Value

A hospital is having a fundraising lottery to raise money for cancer research. A ticket costs \$10, and 2 000 000 tickets are available. There are four levels of prizes: one \$5 000 000 grand prize, three \$100 000 second prizes, ten \$1000 prizes, and 2000 free tickets for next year's lottery.

a) What is the expected value of each ticket?

b) Explain its meaning.

Solution

a) **Method 1: Use the Equation for the Expectation of a Probability Distribution**

$$E(X) = x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + \cdots + x_n \cdot P(x_n) - \text{price per ticket}$$

$$\begin{aligned} E(X) &= 5\,000\,000 \times \frac{1}{2\,000\,000} + 100\,000 \times \frac{3}{2\,000\,000} + 1000 \times \frac{10}{2\,000\,000} \\ &\quad + 10 \times \frac{2000}{2\,000\,000} - 10 \\ &= 5\,330\,000 \times \frac{1}{2\,000\,000} - 10 \\ &= 2.665 - 10 \\ &= -7.335 \end{aligned}$$

Method 2: Subtract the Cost of the Ticket From the Expected Payout

$$E(X) = \frac{\text{total value of all the prizes}}{\text{the number of tickets sold}} - \text{price per ticket}$$

What is the value of a free ticket?

$$\begin{aligned} E(X) &= \frac{(5\,000\,000 + 3 \times 100\,000 + 10 \times 1000 + 2000 \times 10)}{2\,000\,000} - 10 \\ &= 2.665 - 10 \\ &= -7.335 \end{aligned}$$

The expected value per ticket is –\$7.335.

b) On average, a ticket is worth a loss of \$7.335.

Just as with averages, expected values should not necessarily be rounded.

Processes

Connecting
Why might someone choose to buy a ticket for this lottery knowing they probably will not win?

Your Turn

A lottery has a \$10 000 000 grand prize, a \$500 000 second prize, and ten \$50 000 third prizes. A ticket costs \$5, and 4 000 000 tickets were sold.

a) What is the expected value of each ticket?

b) Using the results of this question and of Example 3, are lottery tickets a good investment?

c) How could the lottery be adjusted to make buying a ticket more attractive?

Consolidate and Debrief

Key Concepts

- A probability distribution shows the probabilities of all possible outcomes in an experiment.
- The sum of all probabilities in any distribution is 1.
- A probability histogram graphs the relative frequency of the random variable. The area of each bar represents the probability of the variable.
- Expectation, or expected value, is the weighted average value of the random variable.

$$E(X) = x_1 \cdot P(x_1) + x_2 \cdot P(x_2) + \cdots + x_n \cdot P(x_n)$$
$$= \sum_{i=1}^{n} x_i \cdot P(x_i)$$

The expectation can be a non-integer value.

Reflect

R1. The expected number of children in a Canadian family is 1.8. Should this be rounded to 2 or left as is? Explain.

R2. Give two examples of a discrete probability distribution. Explain what makes them discrete.

R3. Describe the steps in setting up a probability distribution for the sum of two 12-sided dice.

Practise

Choose the best answer for #2 and #3.

1. Classify each of the random variables as discrete or continuous:

a) the number of points scored in a basketball game

b) the length of time players played in a basketball game

c) the mass of the weights in a weight room

d) the number of windows in the classrooms in a school

e) the area of the windows in the classrooms in a school

2. Which of the following is a false statement about expectation?

A The sum in the expected value calculations is equal to 1.

B $E(X) = \sum_{i=1}^{n} x_i \cdot P(x_i)$

C It is the predicted average of all possible outcomes.

D It is equal to the mean of the outcomes weighted according to their respective frequencies.

3. In Example 2 on page 148, what is the discrete random variable?

   A x

   B $P(x)$

   C the number of girls in a family of three children

   D the expected number of girls in a family of three children

4. Draw a probability histogram for each of the distributions.

   a)

x	$P(x)$
1	0.35
2	0.42
3	0.11
4	0.12

   b)

x	$P(x)$
5	$\frac{1}{8}$
10	$\frac{1}{4}$
15	$\frac{5}{12}$
20	$\frac{1}{12}$
25	$\frac{1}{8}$

5. Calculate the expectation for each of the distributions.

   a)

x	$P(x)$
1	0.3
2	0.2
3	0.1
4	0.4

   b)

x	$P(x)$
0	$\frac{1}{5}$
2	$\frac{3}{10}$
4	$\frac{1}{5}$
6	$\frac{1}{10}$
8	$\frac{1}{10}$
10	$\frac{1}{10}$

Apply

6. **Communication** The distribution of marble sizes in a bag is shown in the table.

Diameter (mm)	Frequency
12.0	5
13.0	11
14.0	24
20.0	15
25.0	5

   a) Identify the random variable.

   b) Is the random variable discrete? Explain.

   c) Draw a probability histogram for this distribution.

   d) Describe what each bar in the histogram represents.

   e) Calculate the weighted mean of the diameters. How does this relate to the expectation?

7. **Application** Two 8-sided dice are rolled.

   a) Show the probability distribution for the sums of the two dice.

   b) Draw a probability histogram by hand or using technology.

   c) Calculate the expectation. Explain its meaning in this context.

8. A rectangle is to be drawn on a grid with perimeter of 24 cm. The dimensions are integers, and are randomly selected. Show the probability distribution for either the dimensions or the area. Include a probability histogram.

9. **Thinking** A school is holding a fundraising raffle. The first prize is \$500, the three second prizes are \$100 each, and the five third prizes are \$50 each. A total of 2000 tickets were sold at \$5 each.

   a) What is the probability of winning a prize?

   b) What is the expected payout per ticket?

   c) What is the expected profit per ticket?

   d) What price should have been charged to have a 90% profit per ticket?

Achievement Check

10. A card is chosen from a standard deck, replaced, then another is chosen. This process is repeated three times.
    a) Show the probability distribution for the number of face cards in three trials.
    b) Sketch a graph of this distribution.
    c) Is the number of face cards a discrete random variable? Justify your response.
    d) Calculate the expected value. Explain its meaning.

11. In many games, rolling doubles has beneficial results. Three people are playing a board game in which two dice are rolled.
    a) Use a tree diagram to illustrate the probability distribution of the number of doubles in three rolls of two dice.
    b) Calculate the probability of each outcome in the sample space.
    c) What is the expected number of doubles in the three rolls?

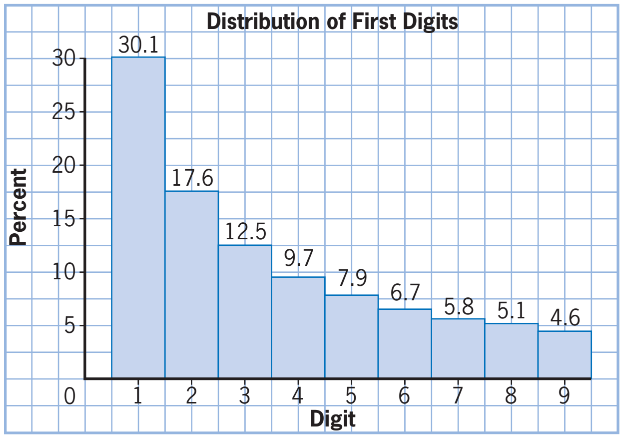
12. Build a probability distribution for the sums of three dice. Include all pertinent components of a distribution, and appropriate explanations.

13. **Open Question** A random device is one that generates a random result. Spinners and dice are typical random devices. Design a random device that has at least four outcomes with non-equal probabilities. Develop the probability distribution for your device and illustrate it using a probability histogram.

Project Prep

You may need to make a random device in your probability project. Think of an appropriate device that you can use.

14. **Communication** The graph shows the percent of numbers that start with each digit when applied to many different data sets, such as hydro bills, addresses, stock prices, population sizes, death rates, and lengths of rivers.



    a) Why would the distribution look this way?
    b) Calculate the expectation. Explain what it means.

Extend

15. When continuously cutting a card from a deck with replacement, what is the probability that the first ace will be cut
    a) on the first try?
    b) on the second try?
    c) on the third try?
    d) on the nth try?

16. Use technology to show the probability histogram for a spinner with five unequal sectors, labelled 1 to 5, respectively. The sectors are proportionally equal in arc length to their labelled numbers.

17. What is the expected sum of two weighted dice on which the number 5 occurs twice as often as the other numbers?

4.2

Uniform Distributions

Learning Goal

I am learning to

- solve problems involving uniform probability distributions

Minds On...

On a TV game show, Allie has the option of taking home \$750 or guessing which one of 26 briefcases contains \$1 000 000.

- What are her chances of winning?
- Is each briefcase equally likely to hold the money? What would you do?
- How would all of this change if she were allowed a second chance after checking the contents of five briefcases?

Action!

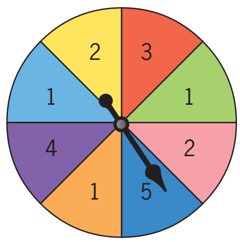
Investigate Uniform Distributions

1. Create the probability distribution for each experiment.

a) the upper face on a single roll of a die

b) the result of a single spin of the spinner

c) the position a person could be assigned when the eight runners in a race are randomly assigned a starting lane from lanes 1 to 8



2. Sketch a probability histogram for each distribution in step 1.

3. Compare the three distributions. How are they different? the same?

4. **Reflect** Which distribution(s) would be considered a **uniform distribution**? Why?

5. **Extend Your Understanding** Rewrite the descriptions of the non-uniform examples so they are uniform.

uniform distribution

- occurs when, in a single trial, all outcomes are equally likely
- for all outcomes, x, $P(x) = \frac{1}{n}$, where n is the number of possible outcomes in the experiment

Example 1

Uniform Distribution

A calculator has been programmed to generate a random number between 1 and 5.

a) Classify this distribution.

b) Calculate the probability distribution.

c) Sketch a graph of the distribution. Comment on the shape of the graph.

d) Calculate the expectation. Interpret its meaning.

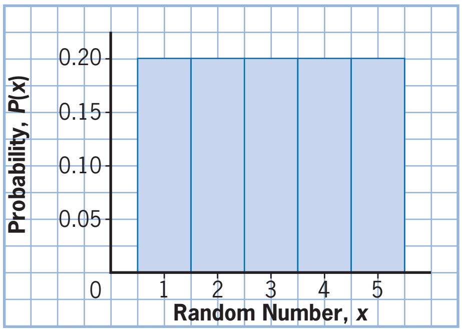
Solution

a) Each random number is equally likely and there is a single trial. So, this is a uniform distribution.

b)

Random Number, x	$P(x)$	$x \cdot P(x)$
1	$\frac{1}{5}$	$\frac{1}{5}$
2	$\frac{1}{5}$	$\frac{2}{5}$
3	$\frac{1}{5}$	$\frac{3}{5}$
4	$\frac{1}{5}$	$\frac{4}{5}$
5	$\frac{1}{5}$	$\frac{5}{5}$

c) Since this is a uniform distribution and all the probabilities are equal, the bars all have the same dimensions.



d) Method 1: Use the Sum of the Values of $x \cdot P(x)$

$$E(X) = \frac{1}{5} + \frac{2}{5} + \frac{3}{5} + \frac{4}{5} + \frac{5}{5}$$
$$= \frac{15}{5}$$
$$= 3$$

Method 2: Use the Common Fraction $\frac{1}{5}$

$$E(X) = \frac{1}{5}(1 + 2 + 3 + 4 + 5)$$
$$= \frac{1}{5}(15)$$
$$= 3$$

You can use $\frac{n(n + 1)}{2}$ to calculate the sum of the numbers from 1 to n. Show how this works.

The expectation is 3. The predicted average value of the random number will be 3.

Your Turn

A screen saver has been programmed to draw a circle with a randomly chosen radius of integer length between 1 and 8 cm.

a) Is the probability distribution of areas uniform? Explain.

b) Calculate the probability distribution.

c) Sketch a graph of the distribution. Comment on the shape of the graph.

d) Calculate the expectation. Interpret its meaning.

Example 2

Fair Game

A game involves rolling a die. A player who rolls an even number receives points equal to two times the face value of the die. If the player rolls an odd number, the player loses three times the face value of the die. Is this a fair game?

Solution

Calculate the expectation. Points received indicate a positive value for the random variable, x. Points lost indicate a negative value for x.

$$E(X) = \sum x \cdot P(x)$$
$$= -\frac{3}{6}$$
$$= -0.5$$

Roll	Point Value, *x*	*P*(*x*)	*x* · *P*(*x*)
1	–3	$\frac{1}{6}$	$-\frac{3}{6}$
2	4	$\frac{1}{6}$	$\frac{4}{6}$
3	–9	$\frac{1}{6}$	$-\frac{9}{6}$
4	8	$\frac{1}{6}$	$\frac{8}{6}$
5	–15	$\frac{1}{6}$	$-\frac{15}{6}$
6	12	$\frac{1}{6}$	$\frac{12}{6}$
		Sum	$-\frac{3}{6}$

A fair game will have an expectation equal to 0. This is not a fair game because the player will lose 0.5 points on each turn, on average.

Project Prep

Your culminating probability project may require you to calculate the expectation and to determine whether it is a fair game. How would you make those calculations for any particular game?

Your Turn

A spinner has eight equally spaced sectors labelled from 1 to 8. In a particular game, a player wins points equal to double the sector's face value if a power of two is spun. For all other spins, the player loses the face value of the spin. Is this a fair game?

Consolidate and Debrief

Key Concepts

- A uniform distribution occurs when, in a single trial, all outcomes are equally likely.
- For a uniform distribution, $P(x) = \frac{1}{n}$, where n is the number of possible outcomes in the experiment.
- When calculating expectation for a uniform distribution, you can factor $\frac{1}{n}$ to make the calculations easier: $E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$
- When calculating expectation, you can calculate the sum of the numbers from 1 to n using the expression $\frac{n(n+1)}{2}$.
- The expected outcome of a fair game is equal to 0.

Reflect

R1. A school board randomly selects students by their student number to take part in a survey. Is this process a uniform distribution? Explain.

R2. A spinner has 10 equally spaced sectors. Draw an example of a spinner with

a) a uniform distribution

b) a non-uniform distribution

R3. A school raffle has payouts totalling \$1500. The school expects 2000 tickets to be sold. Will a price of \$2 per ticket give an advantage to the school, to the customers, or will it be a fair game? Justify mathematically.

Practise

Choose the best answer for #2 and #3.

1. Explain whether each of the following is a uniform distribution:

a) recording the sum of two dice

b) cutting a card from a well-shuffled deck

c) an MP3 player randomly selecting a song from a playlist

d) the number of boys in a family of five children

e) randomly selecting five students to be members of a committee

2. Which of these is not a uniform distribution?

A political parties using robo-callers to telephone all constituents in a riding

B three people being selected at random from a group of four girls and five boys

C dealing one card, face down, to each of five players

D a school randomly selecting a student to attend a conference

3. A random number generator provides a number between 1 and 10. What is the expected outcome?

A 5 **B** 50

C 55 **D** 5.5

4. A jar contains red and green balls. A person reaches in and randomly selects a ball to indicate the number of points earned or lost. There are four red balls, each labelled +3 points. How many green balls, each labelled –2 points, would be required for this to be a fair game?

5. Given the probability distributions, determine the expected values.

a)

x	$P(x)$
5	$\frac{1}{5}$
10	$\frac{1}{5}$
15	$\frac{1}{5}$
20	$\frac{1}{5}$
25	$\frac{1}{5}$

b)

x	$P(x)$
0	12.5%
1	12.5%
2	12.5%
3	12.5%
4	12.5%
5	12.5%
6	12.5%
7	12.5%

Apply

6. A random number between 1 and 12 is generated to decide on the hour during which a special contest will be played on a radio station.

a) Develop the probability distribution for the contest hour, and calculate the expected outcome.

b) Does this mean that the time represented by the expectation is the most likely to be selected? Explain.

7. **Communication** A card is randomly selected from a deck.

a) What is the probability that it is any specific card?

b) Is this an example of a uniform distribution? Explain.

c) The card is not placed back into the deck and a second card is selected. What is the probability it is any specific card?

d) Are the two card choices an example of a uniform distribution? Explain.

8. A multiple choice test has five possible answers, labelled A, B, C, D, E. If the position of the correct answer is to be chosen at random, draw a probability histogram for this distribution.

9. The Prisoner's Dilemma involves two prisoners, P and Q, who are being held for a crime. If both P and Q confess to the crime, each of them goes to prison for two years. If P confesses but Q denies the crime, P will be set free but Q will serve three years in prison (and vice versa). If P and Q both deny the crime, both will serve only one year in prison.

a) If each prisoner's decision is randomly chosen, show the probability distribution for the number of years in prison for prisoner P.

b) If you were prisoner P, what would your decision be? Base your decision on mathematical reasoning.

10. **Application** There are only five platonic solids: tetrahedron (4 faces), cube (6 faces), octahedron (8 faces), dodecahedron (12 faces), and icosahedron (20 faces).

Literacy Link

A *platonic solid* is a regular, convex polyhedron with congruent faces.

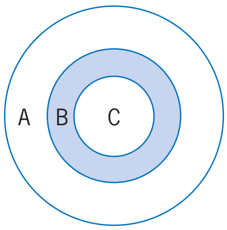
a) Predict the expected outcome for each die.

b) Check your prediction with appropriate calculations for the four smallest platonic solids.

c) Use your findings to confirm or refine your prediction for the icosahedron die.

11. On a TV game show, a giant wheel has 10 equally spaced sectors as shown. To play the game, contestants must risk a certain amount of their previous winnings. What amount of risk money would make this a fair game?

12. **Thinking** A target contains circles with radii of 8 cm, 12 cm, and 20 cm.



a) If a dart randomly lands on one of the three regions, show mathematically that this is not a uniform distribution.

b) Assign points to each area to make this a fair game.

c) Create a similar target with a uniform distribution.

13. In its Flip Your Lid contest, a coffee chain offers prizes of 50 000 free coffees, each worth \$1.50; two new TVs, each worth \$1200; a snowmobile worth \$15 000; and a sports car worth \$35 000. A total of 1 000 000 promotional coffee cups have been printed for this contest. Coffee sells for \$1.50 per cup. What is the expected value of a cup of coffee to the consumer?

14. A charity raffle offers a first prize of \$1 000 000, a second prize of \$100 000, and a third prize of \$10 000. A total of 500 000 tickets will be sold. What price should be charged for a ticket in order for the charity to make a 60% profit on this raffle?

✓ Achievement Check

15. Describe or draw an example of a random number generator to be used in a uniform distribution that

a) has $P(x) = \frac{1}{9}$ for each value of x.

b) has an expected outcome of 8.

c) provides an outcome for an unfair game.

16. The game show Deal or No Deal involves trying to guess which of 26 briefcases contains \$1 000 000. Each briefcase contains a different amount of money. Your teacher will direct you to a website where you can read the rules of Deal or No Deal and play the game yourself online.

a) What is the expectation of this game? How does it compare to the offer given to "quit now"?

b) Calculate the expectation after each of the next two rounds. How does it compare to the offers given to quit?

c) Is this a fair game? Explain.

Extend

17. A uniform distribution has possible outcomes from 1 to n. Develop a formula to calculate the expected outcome.

18. A contest involves a contestant choosing a number between 1 and 10. One of two cards, each containing a formula, is selected at random. The first card indicates that the contestant will win \$40 plus double the contestant's chosen number. The second card indicates that the contestant will win \$100 minus the square of the contestant's chosen number. Describe an appropriate strategy to win the most money.

4.3

Binomial Distributions

Learning Goals

I am learning to

- recognize conditions that give rise to a binomial probability distribution
- make connections among the table, histogram, and algebraic representation of a binomial probability distribution
- solve problems involving binomial probability distributions

Minds On...

Many events in games and industry rely on success or failure, and these often can be quantified with probabilities. For example, in the game of Monopoly, success in getting out of jail means rolling doubles, and failure means any other roll. When measuring the fit of car doors, success could be being within a given gap tolerance. Think of other examples where success or failure could be quantified by probability.

Action!

Investigate Binomial Distributions

Materials

- 4 red tiles and 3 green tiles
- computer with Fathom™ software (optional)

binomial probability distribution

- a distribution with independent trials whose outcomes are either success or failure
- the random variable is the number of successes in a given number of trials

In this activity, you will develop a **binomial probability distribution** for the number of red tiles selected in four independent trials. Randomly select one tile from four red and three green tiles. Repeat four times with replacement.

1. What is the probability of drawing a red tile on a single draw?

2. Make a tree diagram that illustrates the probability distribution for the number of red tiles selected in four trials. Label each branch with the outcome and its independent probability.

3. a) How many paths represent two red tiles?

b) How is this related to ${}_nC_r$ and/or Pascal's triangle?

4. Copy and complete the table for 0 to 4 red tiles.

Number of Red Tiles	Number of Paths	Number of Paths in ${}_nC_r$ Form	Probability	Probability in the form *axbxc*
0				
1				

5. How could you use your understanding of independent probabilities with ${}_nC_r$ to determine the probability of exactly two red tiles?

6. Describe the relationship between the probability column and the other columns.

7. Make a probability histogram for this distribution.

8. **Reflect**

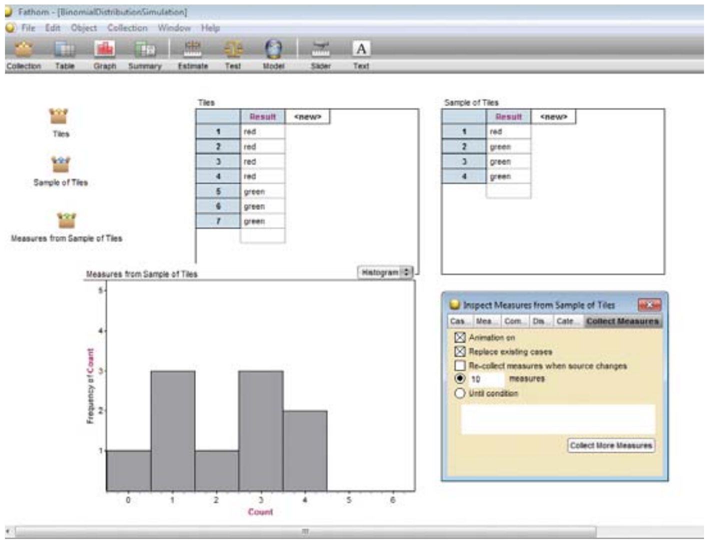
   a) Write a formula for calculating the probability of x red tiles in four independent trials.

   b) Write a formula for calculating the probability of x red tiles in n independent trials.

   c) Use your formula to calculate the probability of selecting two red tiles in five independent trials.

9. **Extend Your Understanding** Your teacher will provide you with a file called **BinomialProbabilityDistribution.ftm**. When you open the file, you will see four open collections.

   - The “Tiles” collection gives the tiles available for selection.
   - “Sample of Tiles” simulates the results when four independent trials are completed.
   - “Inspect Measures from Sample of Tiles” allows you to simulate any number of experiments of four trials each, and is set to 10.
   - The graph is a histogram of the results from these 10 experiments.



   a) Press **Collect More Measures**. You can repeat this any number of times. Describe what happens to the graph.

   b) De-select **Replace existing cases** and change 10 measures to 100. You may wish to turn off animation to speed up the process. Press **Collect More Measures**. How closely does the graph resemble your histogram in step 6? Explain.

Example 1

Counting Successes

Two dice are rolled five times. What is the probability that doubles occur twice?

Solution

The probability of success (rolling doubles) on any individual roll of two dice is $\frac{1}{6}$. The probability of failure is $\frac{5}{6}$.

There will be two successes and three failures in the five rolls. So, there is some combination of $\left(\frac{1}{6}\right)\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right)\left(\frac{5}{6}\right) = \left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3$

The two doubles can occur on any two of the five rolls, in ${}_5C_2$ ways. The three non-doubles can occur in the remaining ${}_3C_3$ ways.

The probability of success on two dice and failure on the other three is ${}_5C_2\left(\frac{1}{6}\right)^2 \times {}_3C_3\left(\frac{5}{6}\right)^3$.

$$
\begin{aligned}
P(x = \text{doubles}) &= {}_5C_2\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 \\
&\approx 0.1608
\end{aligned}
$$

Remember, ${}_3C_3 = 1$.

The probability that doubles occur twice in five rolls is about 0.1608.

Your Turn

A card is repeatedly cut from a deck and replaced each time. What is the probability that, in 10 tries, an ace is cut

a) once? **b)** three times?

The method used in Example 1 can be applied to the general case.

Probability in a Binomial Distribution

The probability of x successes in n identical independent trials is $P(x) = {}_nC_x p^x q^{n-x}$, where p is the probability of success in an individual trial, and $q = 1 - p$ is the probability of failure.

Each term in the expansion of $(p + q)^n$ represents the probability of one possible outcome in the probability distribution.

Expectation for a Binomial Distribution

When determining the expectation for a binomial distribution, you can multiply the number of trials by the probability of success in an individual trial instead of using the standard process.

$E(X) = np$

Example 2

Binomial Distribution

A random number generator provides a number between 1 and 100 over a total of five trials with repetition permitted. Calculate a probability distribution for the number of times a prime number is output.

a) Identify the discrete random variable.

b) Calculate the probability distribution.

c) Verify that the sum of the probabilities is 1.

d) Graph the probability distribution.

e) Describe the shape of the probability histogram.

f) What does $P(5)$ tell you?

g) Calculate the expectation. Interpret its meaning.

Solution

a) X = the number of occurrences of a prime number

b) The 25 prime numbers between 1 and 100 are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, and 97.

Method 1: Use a Scientific Calculator

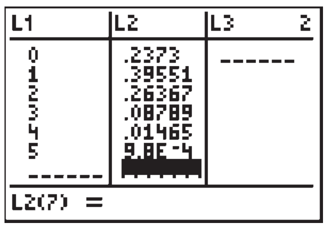
Each trial is independent, and on each trial the probability of a prime number is $\frac{25}{100} = 0.25$.

Number of Primes, x	Probability, $P(x)$	$x \cdot P(x)$
0	${}_5C_0(0.25)^0(0.75)^5 = 0.2373$	0
1	${}_5C_1(0.25)^1(0.75)^4 = 0.3955$	0.3955
2	${}_5C_2(0.25)^2(0.75)^3 = 0.2638$	0.5273
3	${}_5C_3(0.25)^3(0.75)^2 = 0.0879$	0.2637
4	${}_5C_4(0.25)^4(0.75)^1 = 0.0146$	0.0586
5	${}_5C_5(0.25)^5(0.75)^0 = 0.0010$	0.005

Method 2: Use a Graphing Calculator

Refer to the Prerequisite Skills on page 143. After setting up the lists, complete the following steps:

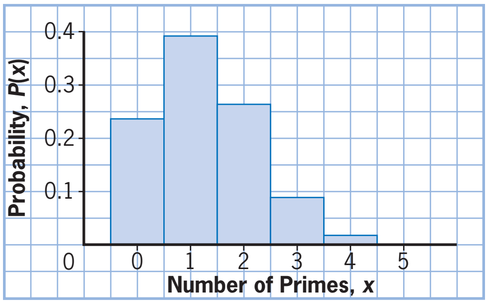
- In the **L1** column, enter the x-values, 0 to 5. To program ${}_5C_x(0.25)^x(0.75)^{5-x}$ in the **L2** column heading, enter 5 MATH **PRB 3:nCr** ENTER 2ND L1 × 0.25 ^ 2ND L1 × 0.75 ^ (5 − 2ND L1) ENTER



c) $0.2373 + 0.3955 + 0.2638 + 0.0879 + 0.0146 + 0.0010 = 1$

The sum of the probabilities is 1.

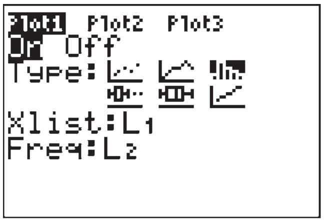
d) Method 1: Use Paper and Pencil



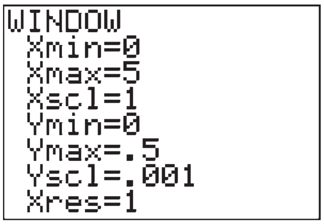
Method 2: Use a Graphing Calculator

Refer to the Prerequisite Skills on page 143.

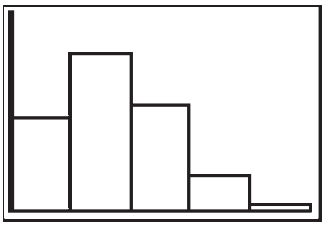
- Set up your Stat Plot screen as shown. Make sure all other Stat Plots have been turned off.



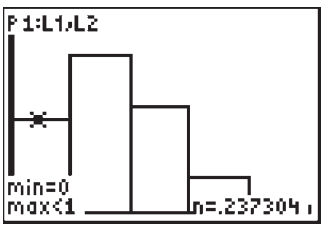
- Press **WINDOW** and change the parameters as shown.



- Press **GRAPH**.



- Press **TRACE** to read the coordinates for each bar.



e) The probability is somewhat bell-shaped, with the mode at $x = 1$ prime. It is skewed to the right.

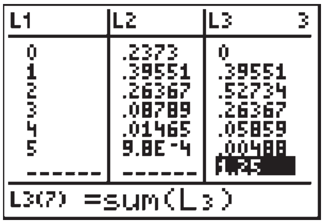
f) $P(5) = 0.001$, which means that selecting five prime numbers is extremely rare.

g) **Method 1: Use the Sum of the $x \cdot P(x)$ Values**

$$\begin{aligned}\sum_{x=0}^{5} x \cdot P(x) &= 0(0.2373) + 1(0.3955) + 2(0.2638) + 3(0.0879) + \\ &\quad\; 4(0.0146) + 5(0.001) \\ &= 1.2501\end{aligned}$$

Method 2: Use a Graphing Calculator

- In the column heading of list **L3**, enter
  2ND L1 × 2ND L2 ENTER.
- In the bottom cell of list **L3**, enter
  2ND LIST MATH **Sum** (2ND L3) ENTER.



Method 3: Use the Formula

$$\begin{aligned}E(X) &= np \\ &= (5)(0.25) \\ &= 1.25\end{aligned}$$

On average, you can expect 1.25 prime numbers out of 5 randomly chosen numbers.

Your Turn

A family has six children. Consider a probability distribution for the number of girls in the family.

a) Identify the discrete random variable.

b) Calculate the probability distribution.

c) Verify that the sum of the probabilities is 1.

d) Graph the probability distribution. Compare the shape of the probability histogram to the one in Example 2.

e) Calculate the expectation. Interpret its meaning.

Literacy Link

A graph is *skewed* when the data are spread out more on one side of the median than the other.

Example 3

Apply the Binomial Distribution

The failure rate is 5% in the initial production run of a new computer chip. A quality control inspector selects 30 chips for testing.

a) What is the probability that more than two of them are defective?

b) What is the expected number of defective chips?

Solution

a) In this case, the probability of success means the *failure* of the chip, so

$$p = 5\% = 0.05 \qquad q = 1 - 0.05 = 0.95$$

You could add the probabilities of having exactly 3, 4, 5, …, 30 failures. It is easier to use the indirect method, $P(x > 2) = 1 - P(x \le 2)$.

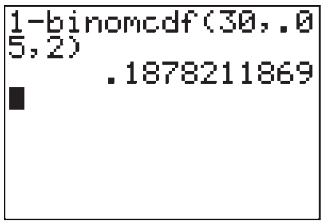
Method 1: Use a Scientific Calculator

$$\begin{aligned} P(x > 2) &= 1 - P(0) - P(1) - P(2) \\ &= 1 - {}_{30}C_0(0.05)^0(0.95)^{30} - {}_{30}C_1(0.05)^1(0.95)^{29} - {}_{30}C_2(0.05)^2(0.95)^{28} \\ &\approx 0.1878 \\ &= 18.78\% \end{aligned}$$

Method 2: Use a Graphing Calculator

Use 2ND DISTR **A:binomcdf(n, p, x)**. This returns the cumulative probability of x successes on n trials with probability of success on each trial of p.

$$\begin{aligned} P(x > 2) &= 1 - P(0) - P(1) - P(2) \\ &= 1 - \text{binomcdf}(30, 0.05, 2) \\ &= 0.1878 \end{aligned}$$



There is about an 18.78% chance that more than two chips are defective.

b) $$\begin{aligned} E(X) &= np \\ &= (30)(0.05) \\ &= 1.5 \end{aligned}$$

On average, there will be 1.5 defective chips in a selection of 30 chips.

Your Turn

With a certain set of atmospheric conditions, the probability of rain is 40%. During a one-month period, eight days had those conditions.

a) What is the probability that it rained on fewer than six of those days?

b) What is the expected number of rainy days?

Consolidate and Debrief

Key Concepts

- A binomial distribution has a specific number of identical independent trials in which the result is success or failure.
- You can represent a binomial distribution using a table, a histogram, and a formula.
- The probability of x successes in n independent trials is $P(x) = {}_nC_x p^x q^{n-x}$, where p is the probability of success in an individual trial, and $q = 1 - p$ is the probability of failure.
- The expectation for the binomial distribution is $E(X) = np$.

Reflect

R1. The formula for the binomial distribution is $P(x) = {}_nC_x p^x q^{n-x}$. What is the purpose of the ${}_nC_x$ coefficient?

R2. In a binomial distribution, how are p and q related? Include an example using a deck of cards.

R3. About 11% of Canadians are left-handed. A newspaper columnist interpreted this to mean that there is an 11% chance that any one of the newspaper's 25 reporters will be left-handed. Discuss the accuracy of this statement.

Practise

Choose the best answer for #1 to #3.

1. Which expression best represents the probability of three successes in seven independent trials in a binomial distribution?

A ${}_7C_3p^3q^4$ **B** ${}_7C_3p^4q^3$

C ${}_7C_4p^3q^4$ **D** ${}_7C_4p^4q^3$

2. What is the expectation for a binomial distribution with $p = 0.5$ and $n = 8$?

A 0.4 **B** 4

C 16 **D** 0.0625

3. Which of the following is an example of a binomial distribution?

A probabilities of the number of queens in a five-card hand

B probability of each sum when two dice are rolled

C probability of each lane for a 100 m race

D probabilities of the number of times a 5 occurs when spinning a spinner six times

4. A tetrahedral die has four triangular faces. Three faces are labelled **1** and the fourth is labelled **2**. The die is rolled four times.

a) Draw a tree diagram to illustrate the possible outcomes.

b) Use the tree diagram to assign probabilities for this distribution.

c) Verify these probabilities using the binomial distribution formula.

d) Substitute values for p and q and expand $(p + q)^4$, but do not simplify. How does the expansion relate to the above results?

5. Prepare a probability table and a graph for a binomial distribution with
   a) $n = 6$ and $p = 0.3$
   b) $n = 8$ and $p = \frac{1}{9}$

6. What is the expected number of times a 6 appears when rolling a die 2000 times?

7. In a family of five children, what is the probability that there are exactly
   a) two girls?
   b) three boys?

Apply

8. Six people are asked to choose a number between 1 and 20. What is the probability that
   a) two people choose the number 9?
   b) at least two people choose the number 9?

9. Two dice are rolled repeatedly and their sum is recorded.
   a) Show the probability distribution for the number of sums of 7 in five rolls.
   b) Graph the distribution with a probability histogram.
   c) Verify the formula $E(X) = np$.

10. In archery competitions, Paul hits the bull's-eye 45% of the time.
    a) Show the probability distribution for the number of bull's-eyes in eight attempts.
    b) What is the expected number of bull's-eyes in eight attempts?
    c) What does $P(8)$ tell you?

11. a) You roll five dice at the same time. What is the probability that you roll two 3s?
    b) Expand the binomial $\left(\frac{1}{6} + \frac{5}{6}\right)^5$.
    c) Which term in the expansion matches the answer in part a)?
    d) How does the binomial probability distribution relate to the binomial theorem?

12. **Application** A machine makes light bulbs, and 6% do not meet the specifications. An inspector randomly chooses 10 light bulbs for testing.

    a) What is the probability that three bulbs do not meet specifications?
    b) What is the probability that seven bulbs do not meet specifications?
    c) What is the probability that between three and seven bulbs do not meet specifications?
    d) What method did you use in part c)? Describe an alternate method.
    e) Should the inspector be concerned if two bulbs do not meet specifications? Explain your reasoning.

13. **Thinking** On a game show, five contestants are each given a box containing 10 car keys, one of which fits their assigned new car. Each contestant is allowed to choose one key and try to start their car. If no car starts, or only one car starts, nobody wins their car. If two or more cars start, then those contestants win their car. Do the results of the game favour the contestants or the game show? Justify mathematically.

14. Jamaal is successful on basketball free throws 80% of the time.
    a) How likely is he to be successful on eight of 10 free-throw attempts?
    b) How likely is he to be successful on at least eight of 10 free-throw attempts?

✓ **Achievement Check**

15. Jean forgot to study for an eight-question multiple choice quiz. Each question contains four possible answers. Jean will guess the answer to each question.

a) What is the probability that she will get only two questions correct?

b) What is the probability that she will pass?

c) What is the expected number of correct answers on the quiz?

d) Predict the shape of the probability histogram for this distribution. Explain your reasoning.

e) Describe how the graph will change if Jean feels that she has a 40% chance of guessing correctly on each question.

f) Use technology to check your predictions to parts d) and e).

16. **Communication** A jar contains 12 red balls and eight green balls. Six balls are removed without replacement. What is the probability that four of the balls are red?

a) Explain why the binomial distribution is not a suitable model for this problem.

b) Write a new question using the same set of balls so it can be modelled using a binomial distribution.

c) Solve the new problem.

17. Opinion polls based on small samples often yield misleading results. In a particular city, 65% of residents are opposed to a new light rail transit system.

a) If a poll were taken, calculate the probabilities of a majority of people approving the transit system with a sample of

- 7 people
- 100 people
- 1000 people

b) Explain any differences in the results.

18. **Thinking** A store offers a scratch and win discount for each customer who spends over $100. Each card has six spots that give a discount of $10, three spots that give a discount of $25, and one spot that gives a discount of $50. What is the expected cost to the store if it has 200 customers one particular day?

19. Your teacher will provide you with the file **BinomialProbabilityDistribution.ftm** that was used in the Investigation at the beginning of this section.

a) Edit the file to simulate the success and failure options for your choice of questions 8, 10, 11, 15, 16, or 18.

b) Try the simulation for 5 experiments, 10 experiments, 100 experiments, and so on, until the graph becomes close to matching the theoretical probabilities.

c) How many experiments did it take for the simulation to come close to the theoretical probabilities?

20. **Open Question** Use one of the following rates to develop your own problem involving the binomial distribution. Then, trade problems with a classmate.

- 19% of the Canadian population live in rural areas
- 39% of the Canadian population live in Ontario

Extend

21. A standard die is painted so that opposite faces are green, red, and yellow, respectively. In 10 rolls of this die, how many could be red or green or yellow? This leads to what could be a trinomial distribution.

a) Develop a formula to calculate the probability distribution in which there are three outcomes with individual probabilities of p, q, and r.

b) Use your formula to determine the probability of rolling three reds, two greens, and five yellows in 10 rolls of the die described above.

22. Derive the formula for expectation of a binomial distribution $E(X) = np$ algebraically.

4.4

Hypergeometric Distributions

Learning Goals

I am learning to

- recognize conditions that give rise to a hypergeometric distribution
- calculate the probability associated with each random variable of a hypergeometric distribution
- represent the hypergeometric distribution using a table and a probability histogram
- solve problems involving hypergeometric probability distributions

Minds On...

The binomial distribution involves independent trials. This section develops a distribution involving dependent trials. Cutting cards with a standard deck provides independent trials, whereas dealing the cards involves dependent trials. Similarly, selecting a jury pool and catching and tagging animals for scientific research involve dependent trials. Brainstorm other examples of dependent trials.

Action!

Investigate Hypergeometric Distributions

Materials

- 4 red tiles and 3 green tiles
- computer with Fathom™ software (optional)

hypergeometric probability distribution

- a distribution with dependent trials whose outcomes are either success or failure
- the random variable is the number of successes in a given number of trials

In this activity, you will develop a hypergeometric probability distribution for the number of red tiles selected in four dependent trials. Randomly select one tile from four red and three green tiles. Repeat four times without replacement.

1. What is the total number of ways of selecting four tiles from seven tiles, without replacement?

2. Make a tree diagram that illustrates the probability distribution for the number of red tiles selected in four trials. Label each branch with the outcome and its dependent probability.

3. a) How many paths represent two red tiles?

   b) How is this related to ${}_nC_r$ and/or Pascal's triangle?

4. Copy and complete the table for 0 to 4 red tiles.

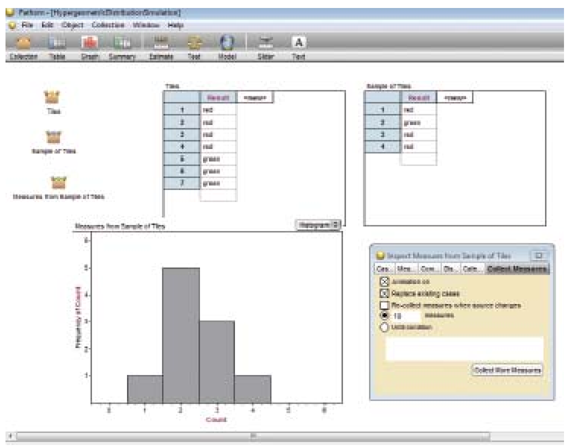
Number of Red Tiles	Number of Paths	Number of Paths in ${}_nC_r$ Form	Probability	Probability as $\frac{axb}{c}$, where a, x, b, c are written as ${}_nC_r$
0				
1				

5. Describe the relationship between the probability column and the other columns.

6. Make a probability histogram for this distribution.

7. **Reflect**
   a) Write a formula for the probability of choosing three red tiles in four dependent trials.
   b) Write a formula for calculating the probability of x red tiles in four dependent trials.
   c) Write a formula for calculating the probability of x red tiles in r dependent trials.
   d) Use your formula to calculate the probability of selecting two red tiles in five dependent trials.
   e) Compare your answer to that in section 4.3 Investigate Binomial Distributions, step 7c) on page 161.

8. **Extend Your Understanding** Your teacher will provide you with a file called **Hypergeometric ProbabilityDistribution.ftm**. When you open the file, you will see four open collections.



   - The "Tiles" collection gives the tiles available for selection.
   - "Sample of Tiles" simulates the results when four dependent trials are completed.
   - "Inspect Measures from Sample of Tiles" allows you to simulate any number of experiments of four trials each, and is set to 10.
   - The graph is a histogram of the results from these 10 experiments.

   a) Press **Collect More Measures**. You can repeat this any number of times. Describe what happens to the graph.
   b) De-select **Replace existing cases** and change 10 measures to 100, then to 1000. You may wish to turn off animation to speed up the process. Press **Collect More Measures**. How closely does the graph resemble your histogram in step 6? Explain.

Example 1

Hypergeometric Probability

A committee of six people is to be formed from a pool of six grade 11 students and seven grade 12 students. Determine the probability that the committee will have two grade 11 students.

Solution

The population size is 13 students. The size of the sample space is six students.

$n(S) = {}_{13}C_6$

For the successful outcome, select two of the six grade 11 students and four of the seven grade 12 students.

$n(2 \text{ grade } 11\text{s}) = {}_6C_2 \times {}_7C_4$

$$\begin{aligned} P(2 \text{ grade } 11\text{s}) &= \frac{{}_6C_2 \times {}_7C_4}{{}_{13}C_6} \\ &= \frac{525}{1716} \\ &\approx 0.3059 \end{aligned}$$

The probability of having two grade 11 students on the committee is about 30.6%.

Your Turn

On a team of 15 astronauts, six are women and nine are men. If four astronauts are selected at random for a flight simulation, what is the probability that two men and two women are selected?

Probability in a Hypergeometric Distribution

The probability of x successful outcomes in r dependent trials is

$$P(x) = \frac{{}_aC_x \cdot {}_{n-a}C_{r-x}}{{}_nC_r}$$

where a is the number of successful outcomes available in a population of size n.

Expectation for a Hypergeometric Distribution

The ratio of the expectation to the number of trials is proportional to the ratio of the number of available successes to the size of the population.

$$\frac{E(x)}{r} = \frac{a}{n}$$

So, $E(X) = \frac{ra}{n}$.

Example 2

Hypergeometric Distribution

A five-card hand is dealt from a standard deck of cards.

a) Show the probability distribution for the number of hearts in the hand.

b) Illustrate the distribution with a probability histogram.

c) Describe the shape of the graph.

d) What does $P(5)$ tell you?

e) Calculate the expectation and explain its meaning.

Solution

a) $n(S) = {}_{52}C_5$

$\quad = 2\ 598\ 960$

There can be 0 to 5 hearts in the hand. The number of ways the hearts can be chosen is ${}_{13}C_x$. The number of ways of choosing the remaining cards is ${}_{39}C_{5-x}$.

$P(x) = \dfrac{{}_{13}C_x \times {}_{39}C_{5-x}}{{}_{52}C_5}$, where x is the number of hearts in the hand.

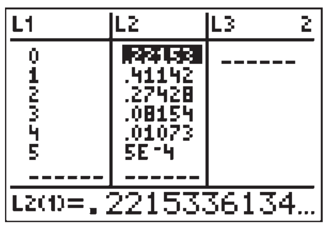
Method 1: Use Paper and Pencil

Number of Hearts, *x*	Probability, *P*(*x*)	*x* · *P*(*x*)
0	$\dfrac{{}_{13}C_0 \times {}_{39}C_5}{{}_{52}C_5} \approx 0.2215$	0
1	$\dfrac{{}_{13}C_1 \times {}_{39}C_4}{{}_{52}C_5} \approx 0.4114$	0.4114
2	$\dfrac{{}_{13}C_2 \times {}_{39}C_3}{{}_{52}C_5} \approx 0.2743$	0.5486
3	$\dfrac{{}_{13}C_3 \times {}_{39}C_2}{{}_{52}C_5} \approx 0.0815$	0.2445
4	$\dfrac{{}_{13}C_3 \times {}_{39}C_1}{{}_{52}C_5} \approx 0.0107$	0.0428
5	$\dfrac{{}_{13}C_5 \times {}_{39}C_0}{{}_{52}C_5} \approx 0.0005$	0.0025

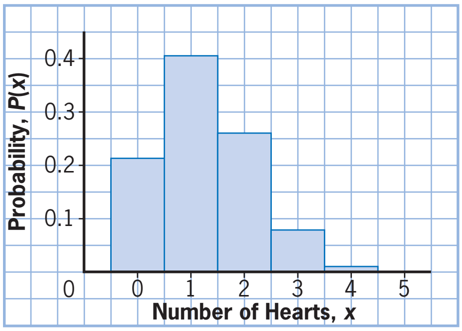
Method 2: Use a Graphing Calculator

Refer to the Prerequisite Skills on page 143.

- In list **L1**, enter the x-values, 0 to 5.
- To program $\dfrac{{}_{13}C_x \times {}_{39}C_{5-x}}{{}_{52}C_5}$ in the **L2** column heading, enter 13 MATH **PRB 3:nCr** ENTER 2ND L1 × 39 MATH **PRB 3:nCr** (5 − 2ND L1) ÷ 52 MATH **PRB 3:nCr** 5 ENTER



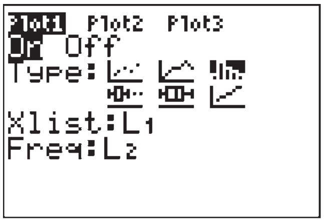
b) Method 1: Use Pencil and Paper



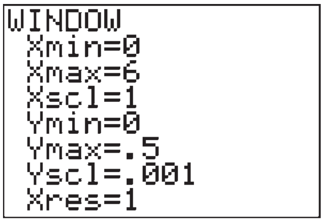
Method 2: Use a Graphing Calculator

Refer to the Prerequisite Skills on page 143.

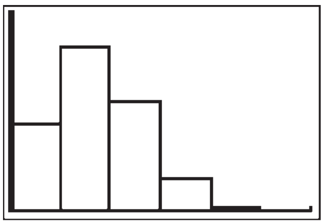
- Set up your **Plot1** screen as shown. Make sure all other Stat Plots have been turned off.



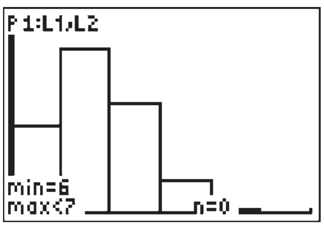
- Press **WINDOW** and change the parameters as shown.



- Press **GRAPH**.



- Press **TRACE** to read the values associate with each bar.



c) The graph has the shape of a bell curve with the mode at $x = 1$. It is skewed to the right.

d) $P(5) = 0.0005$
$\qquad = 0.05\%$

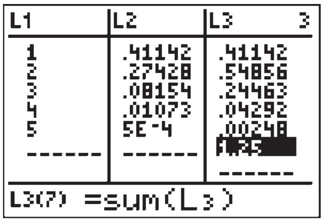
Getting five hearts is extremely rare.

e) **Method 1: Use the Sum of the $x \cdot P(x)$ Values**

$$\sum_{x=0}^{5} x \cdot P(x) = 0(0.221\,53) + 1(0.411\,42) + 2(0.274\,28) + 3(0.081\,54) + 4(0.010\,73) + 5(0.0005)$$
$$= 1.2498$$

Method 2: Use a Graphing Calculator

- In the column heading of list **L3**, enter 2ND L1 × 2ND L2 ENTER
- In the bottom cell of list **L3**, enter 2ND LIST MATH Sum (2ND L3) ENTER



Method 3: Use the Formula

$$E(X) = \frac{ra}{n}$$
$$= \frac{(5)(13)}{52}$$
$$= 1.25$$

Do not round this value, since it represents an average.

On average, there will be 1.25 hearts in a five-card hand.

Your Turn

A bag contains 10 jellybeans. Four are blue and six are green. Four jellybeans are selected at random.

a) Show the probability distribution for the number of green jellybeans selected.

b) Illustrate the distribution with a probability histogram.

c) Compare the shape of the graph to the one in Example 2. Explain any differences.

d) What does $P(0)$ tell you?

e) Calculate the expectation and explain its meaning.

Project Prep

You might use cards when designing your probability project. Will they involve a hypergeometric distribution? How can you use cards appropriately?

Example 3

Apply the Hypergeometric Distribution to Selections

In a class of 30 students, 18 have a driver's licence. Ten students are selected at random.

a) What is the probability that at least four have their driver's licence?

b) What is the expected number of students with their driver's licence?

Solution

a) $n = 30$, $r = 10$, $a = 18$

You could add the probabilities of 4, 5, … 10 students with their driver's licence. It is easier to use the indirect method by subtracting the probabilities of 0, 1, 2, and 3 students with their driver's licence from 1.

$P(x \geq 4) = 1 - P(x \leq 3)$.

Method 1: Use a Scientific Calculator

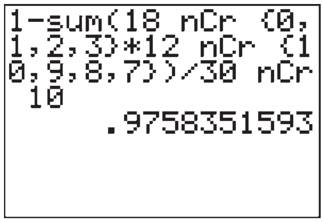
$$P(x \geq 4) = 1 - P(0) - P(1) - P(2) - P(3)$$
$$= 1 - \frac{{}_{18}C_0 \times {}_{12}C_{10}}{{}_{30}C_{10}} - \frac{{}_{18}C_1 \times {}_{12}C_9}{{}_{30}C_{10}} - \frac{{}_{18}C_2 \times {}_{12}C_8}{{}_{30}C_{10}} - \frac{{}_{18}C_3 \times {}_{12}C_7}{{}_{30}C_{10}}$$
$$\approx 0.9758$$

The probability that at least four students will have a driver's licence is about 0.9758.

Method 2: Use a Graphing Calculator

A graphing calculator can combine operations on sets of numbers by placing them in lists using curly brackets {} for both the x values of 0, 1, 2, and 3, and the $10 - x$ values of 10, 9, 8, and 7. Use the sum command to add the results. To enter $1 - {}_{18}C_x - {}_{12}C_{10-x} - {}_{30}C_{10}$, type

1 [–] [2ND] [LIST] **MATH 5:sum(** [ENTER] 18 [MATH] **PRB 3:nCr** [ENTER] [2ND] [{] 0,1,2,3 [2ND] [}] [×] 12 [MATH] **PRB 3:nCr** [ENTER] [2ND] [{] 10,9,8,7 [2ND] [}] [)] [÷] 30 [MATH] **PRB 3:nCr** [ENTER] 10 [ENTER]



The probability that at least four students will have their driver's licence is about 0.9758.

b)
$$E(X) = \frac{ra}{n}$$
$$= \frac{(10)(18)}{30}$$
$$= 6$$

On average, six students will have their driver's licence in a selection of 10 students.

Your Turn

Twenty-four students have signed up to attend a workshop. Fourteen are female and ten are male. Seven are randomly chosen to attend.

a) What is the probability that at least three are male?

b) What is the expected number of male and female students chosen?

Example 4

Apply the Expectation Formula

Wildlife officials tagged 350 seals from a particular colony. Forty seals were caught later, and 17 of them had been tagged. What is the approximate size of the seal population in this colony?

Solution

The 40 seals caught were all independent from each other, so the trials were dependent. If they were tagged, the trial was deemed a success; if not, the trial was deemed a failure. This is represented by a hypergeometric distribution.

$n =$ size of seal population

$a =$ number originally tagged (number available)
$\quad = 350$

$r =$ number later caught (sample size)
$\quad = 40$

$E(X) =$ number of seals that had been tagged (expectation from the sample)
$\quad = 17$

Method 1: Use the Expectation Formula

$$E(X) = \frac{ra}{n}$$
$$17 = \frac{(40)(350)}{n}$$
$$n = \frac{(40)(350)}{17}$$
$$n = 823.53$$

The seal population is about 824 seals.

Method 2: Use a Proportion

$$\frac{E(X)}{r} = \frac{a}{n}$$
$$\frac{17}{40} = \frac{350}{n}$$
$$n = \frac{(40)(350)}{17}$$
$$n = 823.53$$

When a whole number answer is needed, rounding the expectation is appropriate.

Your Turn

During one summer, 500 foxes were caught and vaccinated against rabies. At that time, they were also tagged. Eighty foxes were later caught to estimate the size of the fox population, and 34 of them had been tagged. Estimate the size of the fox population.

Consolidate and Debrief

Key Concepts

- A hypergeometric probability distribution occurs when there are two outcomes, success and failure, and all trials are dependent. The random variable is the number of successes in a given number of trials.
- You can represent a hypergeometric distribution using a table, a probability histogram, or a formula.
- The probability of x successes in r dependent trials is $P(x) = \dfrac{{}_aC_x \cdot {}_{n-a}C_{r-x}}{{}_nC_r}$, where a is the number of successful outcomes available in a population of size n.
- Expectation $E(X) = \frac{ra}{n}$.

Reflect

R1. For each example of a hypergeometric distribution, identify the random variable, the size of the sample space, the size of the population, and the range of the random variable.

a) A bag contains six red and four green marbles. Five marbles are randomly selected from the bag. The number of red marbles is recorded.

b) A seven-card hand is dealt from a standard deck. The number of hearts is recorded.

R2. A standard die is rolled five times and the number of 3s is noted. Explain why this would or would not be a valid hypergeometric probability situation.

Practise

Choose the best answer for #1 to #2.

1. A bag contains five red and six blue blocks. What is the probability of getting three red blocks if four blocks are randomly selected?

A $\dfrac{{}_{11}C_3 \times {}_6C_4}{{}_{18}C_4}$ **B** $\dfrac{{}_5C_3}{{}_{11}C_4}$

C $\dfrac{{}_5C_3}{{}_6C_4}$ **D** $\dfrac{{}_5C_3 \times {}_6C_1}{{}_{11}C_4}$

2. Which is an example of a hypergeometric distribution?

A probability of the number of aces in a seven-card hand

B probability of each sum when two dice are rolled

C probability of a number being randomly chosen from the numbers 1 to 10

D probability of the number of times a 3 occurs when rolling a die six times

3. Each expression represents the probability of a hypergeometric probability. State the values of the unknowns.

a) $\dfrac{{}_6C_3 \times {}_9C_2}{{}_nC_r}$ **b)** $\dfrac{{}_aC_5 \times {}_7C_b}{{}_{10}C_6}$

c) $\dfrac{{}_6C_3 \times {}_cC_2}{{}_{25}C_d}$

4. Show the hypergeometric probability distribution for an experiment with

a) $n = 15$, $r = 4$, $a = 7$.

b) $n = 8$, $r = 4$, $a = 4$.

Apply

5. A five-card hand is dealt from the honour cards in a standard deck (10, J, Q, K, A).
   a) Show the probability distribution for the number of hearts in the hand.
   b) Calculate the expectation in two ways.

6. In a box of 20 light bulbs, five are defective. Three light bulbs are selected at random.
   a) What is the probability that at least one is defective?
   b) What is the expected number of defective light bulbs?
   c) What is the meaning of $P(3)$ in this context?

7. In the card game of bridge, 13 cards are dealt to each player. Find the probability of each of the following hands:
   a) 4 aces
   b) at least 1 king
   c) 5 clubs, 8 diamonds

8. **Communication** A laboratory has 30 mice, eight of which have a specific genetic mutation. A lab assistant randomly selects 10 of the mice. Which has a greater probability of the mice having the genetic mutation, fewer than three, or more than seven? Explain.

9. In a provincial park, 200 foxes are tagged. In 100 sightings, 14 were tagged. Estimate the size of the fox population.

10. **Application** Wildlife officials tagged 80 deer in an area that had approximately 120 deer.
    a) If they later took a sample of 25 deer, how many would they expect to have been tagged?
    b) Should the officials be surprised if the sample has fewer than 13 tagged deer? Explain your thinking.

11. **Thinking**
    a) Which will have a greater probability, a seven-card hand with no spades, or a five-card hand with no spades.
    b) Verify mathematically and explain any discrepancies with your prediction.

✓ Achievement Check

12. The members of an antique car club own one car from the 1920s, two cars from the 1930s, four cars from the 1940s, six cars from the 1950s, and seven cars from the 1960s. They have been invited to send four cars to a car show and will choose them at random.
    a) Compare the probabilities of sending all four cars from any given decade.
    b) Show the probability distribution for the number of cars selected from the 1950s or 1960s, in table form and graphically.
    c) What is the expected number of cars sent from the 1940s?
    d) How would the graph of the distribution in part b) change if there were an additional three cars from the 1920s?

13. A jury of 12 people is to be chosen from a pool of 9 men and 11 women.
    a) Graph the probability distribution for the number of men on the jury.
    b) Simulate the probability distribution by using 9 red and 11 black playing cards. Perform 10 trials, then combine your results with other classmates. Alternatively, use the file **HypergeometricProbability Distribution.ftm** that was used in the Investigation on page 170. Make appropriate changes to accommodate the different choices available. Run the simulation for 10 experiments, then 100, and so on.
    c) After how many experiments did the simulation reasonably match the theoretical probabilities?

Extend

14. Calculate the probability that a bridge hand of 13 cards contains four cards of one suit and three of each other suit.

15. A bag contains three nickels, five dimes, and four quarters. Three coins are removed at random. What is the probability that the value of the coins will total 75 cents?

4.5

Comparing and Selecting Discrete Probability Distributions

Learning Goals

I am learning to

- compare the probability distributions of discrete random variables
- solve problems involving uniform, binomial, and hypergeometric distributions

Minds On...

In earlier chapters, you looked at independent and dependent events. When determining the appropriate probability distribution, this is an important criterion to consider. How would you differentiate between independent and dependent events? Come up with two or three examples of each. Consider board games, card games, or other types of games.

Action!

Investigate Comparing Binomial and Hypergeometric Distributions

Consider the binomial and hypergeometric distributions.

1. If you have not already done so, complete the compare and contrast graphic organizer on page 140. Consider the following criteria: population, discrete or continuous, independence of trials, counting outcomes, number of trials, random variable, what needs to be known?, expectation, parameters (n, r).

2. **Reflect** Use your graphic organizer to help classify each of the probability distributions as binomial, hypergeometric, or neither. Justify your classification.
   a) the probability of successfully shooting 13 free throws in 15 tries given the probability of success on a free throw
   b) the probability of each possible outcome when a card is drawn from a standard deck
   c) selecting 25 grizzly bears at random and determining how many of them were tagged with radio chips over the last year
   d) the probability that three or more batteries are defective in a batch of 35 batteries when batteries have a rate of defect of 0.05%

3. **Extend Your Understanding** Compare and contrast the uniform distribution with the binomial and hypergeometric distributions.

Example

Compare Two Similar Distributions

a) Compare and contrast the following probability distributions. Include the values of the parameters.

- cutting five cards from a standard deck, with replacement, and counting the number of face cards
- dealing five cards at the same time from a standard deck and counting the number of face cards

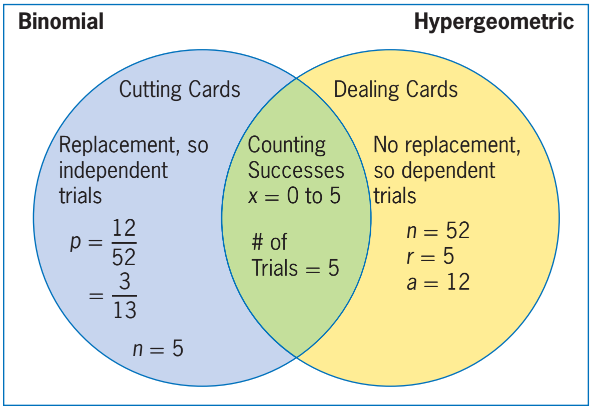
b) Graph the two probability histograms.

c) How are the graphs alike? How are they different?

> **Literacy Link**
>
> A *parameter* is a constant that can have different values in an expression, but that does not change the form of the expression. For example, in $y = mx + b$, m and b are parameters, while x and y are variables.

Solution

a) Use a Venn diagram to compare and contrast.

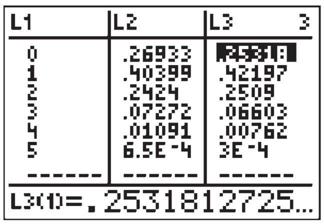


b) Method 1: Use a Graphing Calculator

Refer to the Prerequisite Skills on page 143.

- In list **L1**, enter the x-values, 0 to 5.
- For the binomial distribution, to program ${}_5C_x\left(\frac{3}{13}\right)^x\left(\frac{10}{13}\right)^{5-x}$ in the **L2** column heading, enter:

  5 MATH **PRB 3:nCr** ENTER 2ND L1 × (3 ÷ 13) ^ 2ND L1 × (10 ÷ 13) ^ (5 − 2ND L1) ENTER

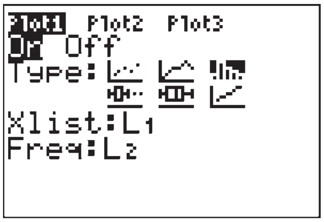


- For the hypergeometric distribution, to program $\frac{{}_5C_x \times {}_{40}C_{5-x}}{{}_{52}C_5}$ in the **L3** column heading, enter:

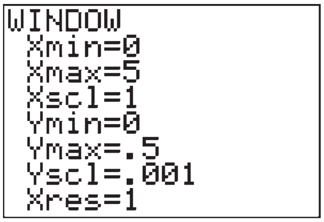
  5 MATH **PRB 3:nCr** ENTER 2ND L1 × 40 MATH **PRB 3:nCr** ENTER (5 − 2ND L1) ÷ 52 MATH **PRB 3:nCr** ENTER 5 ENTER

  To see the graphs, press: 2ND STAT PLOT **1:Plot 1** ENTER

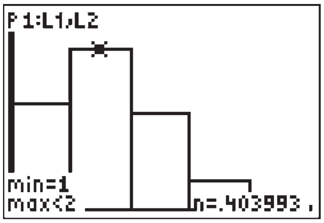
- Set up your **Plot1** screen as shown. Make sure all other Stat Plots have been turned off.



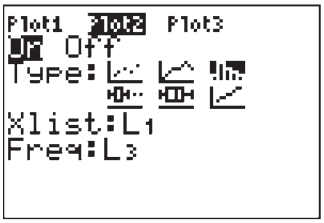
- Press WINDOW and change the parameters as shown.



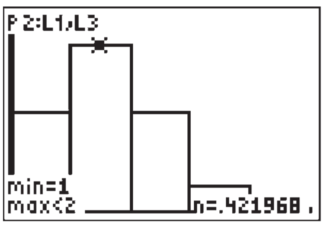
- Press TRACE. Use the left and right arrows to read the coordinates of each bar.



For the hypergeometric distribution, turn off **Plot1** and, when setting up the Stat Plot, use **Plot2** and **Freq:L3**.



- Press TRACE. Use the left and right arrows to read the coordinates of each bar.



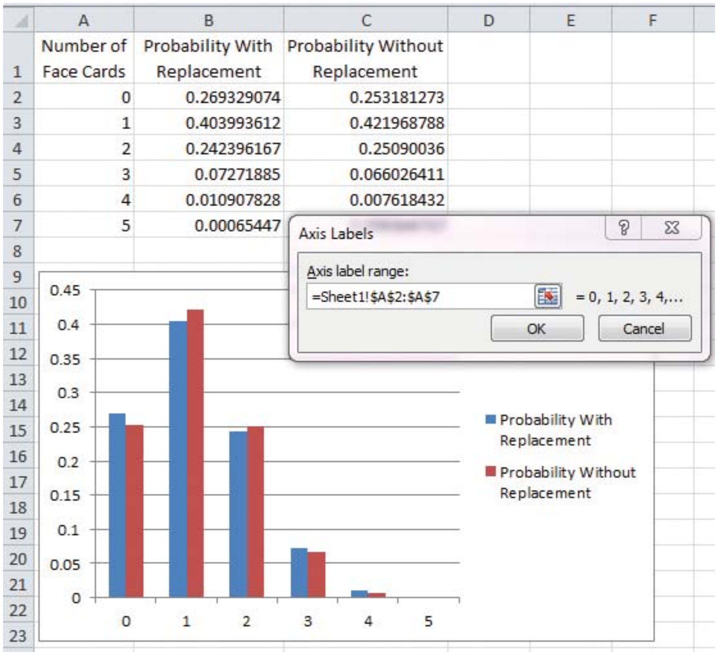
Method 2: Use a Spreadsheet

Open a spreadsheet.

- Create three columns called **Number of Face Cards**, **Probability With Replacement**, and **Probability Without Replacement**.
- In the **Number of Face Cards** column, enter the values from 0 to 5.
- In the **Probability With Replacement** column, enter the binomial formula in B2: =COMBIN(5,A2)*(3/13)^A2*(10/13)^(5–A2). Copy it from B2 to B7.
- In the **Probability Without Replacement** column, enter the hypergeometric in C2: =COMBIN(12,A2)*COMBIN(40,5–A2)/COMBIN(52,5). Copy it from C2 to C7.

	A	B	C
1	Number of Face Cards	Probability With Replacement	Probability Without Replacement
2	0	0.269329074	0.253181273
3	1	0.403993612	0.421968788
4	2	0.242396167	0.25090036
5	3	0.07271885	0.066026411
6	4	0.010907828	0.007618432
7	5	0.00065447	0.000304737

- Highlight columns B and C, select **Insert** from the ribbon, and then select **Clustered Column**.
- Right click on the chart and click on **Select Data...** .
- Under Horizontal (Category) Axis Labels, click **Edit**. Highlight cells A2 to A7 for the axis label range. Click **OK**. Click **OK** again.



c) The graphs have the same bell-like shape, with the $x = 1$ face card being the most likely outcome. The hypergeometric graph is slightly taller than the binomial graph at $x = 1$ (0.422 vs 0.404) and $x = 2$ (0.251 vs. 0.242), and shorter at the other values of x. This occurs due to the dependent nature of the hypergeometric distribution, causing probabilities to increase when fewer choices are available.

Your Turn

a) Use a Venn diagram to compare and contrast the probability distributions if a hat contains five male and six female names.
- Selecting four names with replacement, and counting the number of female names.
- Selecting four names without replacement, and counting the number of female names.

b) Graph the two probability histograms.

c) How are the graphs alike? How are they different?

Consolidate and Debrief

Key Concepts

- The chart summarizes the general conditions of the distributions.

	Uniform	Binomial	Hypergeometric
Parameters and What They Represent	n = number of items	n = number of trials p = probability of success on an individual trial q = probability of failure on an individual trial	n = size of the population r = number of trials a = number of successful items available
Definition of Random Variable, x	Value of the outcome	Number of successful outcomes	Number of successful outcomes
Range of Values for x	Depends on the situation	$x = 0, 1, 2, \ldots, n$	$x = 0, 1, 2, \ldots, r$
Probability Formula	$P(x) = \frac{1}{n}$	$P(x) = {}_nC_x p^x q^{n-x}$	$P(x) = \frac{{}_aC_x \cdot {}_{n-a}C_{r-x}}{{}_nC_r}$
Expectation Formula	$E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$	$E(X) = np$	$E(X) = \frac{ra}{n}$
Identifying Characteristics	All items are equally likely A single trial	Trials are independent Successes are counted	Trials are dependent Successes are counted

Reflect

R1. Refer to the graphic organizer in the Investigation on page 180, and to the general conditions chart in the Key Concepts above. Make a Venn diagram or a graphic organizer to compare and contrast the general conditions for the binomial and hypergeometric probability distributions.

R2. Sam wrote that the difference between binomial and hypergeometric distributions is that with the binomial distribution each trial has the same probability, but with hypergeometric the individual probabilities change with the sampling. Is this an accurate statement? Explain.

Practise

Choose the best answer for #2 and #3.

1. State which type of probability distribution (uniform, binomial, hypergeometric, none of these) would model each situation.

a) A health inspector is in charge of inspecting 75 restaurants, 15 of which have had health code violations in the past. The inspector randomly selects 10 of the 75 restaurants for inspection. What is the probability that four of these will have had health code violations?

b) It is estimated that 12% of all restaurants in a city have had health code violations. Ten restaurants are selected at random for inspection. What is the probability that four of these will have had health code violations?

c) For a charity lottery, you picked 1, 2, 3, 5, and 8 from the numbers 1 to 20. Five different winning numbers are selected at random. What is the probability of three of your numbers matching the five winning numbers?

d) For a school fundraising draw, 1000 tickets are sold, each with a number from 0001 to 1000. The winning ticket is drawn from a bin. What is the probability of winning the draw?

2. If, in a probability distribution, the number of successes is counted, then the distribution

A must be binomial

B must be hypergeometric

C may be either binomial or hypergeometric

D may be neither binomial nor hypergeometric

3. On a TV game show there are nine squares, five of which have a winning sum of money. The contestant selects four different squares. The probability distribution for the number of squares chosen that contain money is

A uniform

B binomial

C hypergeometric

D none of the above

4. Identify possible random variables for the following experiments and the values the variables may take:

a) dealing five cards from a deck

b) naming four members of a committee selected from five grade 11 and seven grade 12 students

c) cutting a card from a deck

d) rolling a die 10 times

e) testing 20 bottles of ginger ale for quality control

f) selecting a winning square on a TV game show

Apply

5. **Communication** Make a flowchart to help you decide when to use each type of distribution (uniform, binomial, hypergeometric, neither).

6. A game consists of randomly selecting a number from 1 to 15. Your favourite number is 13, and you are hoping your number will come up.

a) Is this a uniform or binomial distribution? Explain.

b) Rewrite the situation to convert it to uniform or binomial, as appropriate.

7. At Bill's Burger Barn, there is a one in eight chance of winning a free hamburger. Nicolas bought a hamburger every day for five days, hoping to win as many free hamburgers as possible.

a) Is this a binomial or hypergeometric distribution? Explain.

b) Rewrite the situation to convert it to binomial or hypergeometric, as appropriate.

8. For a random draw, 20 slips of paper containing people's names are placed into a bin. Barb noted that four of the names were her friends. Five names will be selected to win a prize, and Barb is hoping at least one of the prizes goes to a friend.

a) Is this a binomial or hypergeometric distribution? Explain.

b) Rewrite the situation to convert it to binomial or hypergeometric, as appropriate.

c) Calculate the probability of success for Barb in each distribution.

d) Which distribution would make Barb happier? Why?

9. a) With or without technology, simulate the binomial and hypergeometric distributions in #8. See the Investigates in section 4.3, page 160, and section 4.4, page 170, for instructions on using Fathom™ to simulate probability distributions.

b) Compare the theoretical probabilities to the simulation.

c) What would make the simulation closer to the theoretical?

d) What would make the binomial and hypergeometric close to being the same?

10. **Application** Compare the expectations for cutting a card from a deck four times and for dealing four cards. Then, explain the results

a) for the number of aces

b) for the number of red cards

c) for the number of hearts

✓ Achievement Check

11. A basket contains 20 slips of paper, each with a different student's name on it. Eight of the names are boys and 12 are girls. Six different names are selected at random, and those students will win fantastic prizes!

a) Explain why this scenario can be modelled using a hypergeometric distribution.

b) Show a full probability distribution for the number of girls who win prizes.

c) Determine the expected number of girls who win prizes using two methods and confirm that they are equal. If not, explain any differences.

d) Rewrite the situation described above to change it to a binomial distribution.

12. **Thinking** At a fall fair, players in a ring-toss game are successful 8% of the time.

a) Design a problem that would involve a binomial distribution.

b) Design a problem that would involve a hypergeometric distribution.

c) Design a problem that would involve the uniform distribution.

13. Thinking An activity involves selecting six people from a population in which six are males and the rest are females. The results will be different if the population is 10 people than if it is 200 people. Use technology to develop solutions to the questions:

a) Make graphs for the two population sizes using the hypergeometric distribution.

b) Make graphs for the two population sizes using the binomial distribution.

c) Compare the two distributions for $n = 10$ and for $n = 200$.

d) Comment on the accuracy of using the binomial distribution to approximate the hypergeometric distribution when r is small in relation to n.

Extend

14. When the population in a binomial distribution is very large and p is very small, it can be modelled using the Poisson distribution, named for the French mathematician Siméon-Denis Poisson (1781–1840). It uses the formula $P(x) = \frac{e^{-np}(np)^x}{x!}$, where e is the irrational number 2.718 28…. An estimated 1.5% of the world's population has green eyes. If 2000 people were selected at random, use the Poisson distribution to calculate the probability that fewer than 10 have green eyes. Compare the results using a graphing calculator. How close is the approximation?

15. a) Determine the expected values of the following:

i) a single trial in which the random variable can take the values $1, 2, 3 \ldots n$

ii) multiple independent trials in which there are a successful items in a population of n items, and the random variable can take the values $1, 2, 3 \ldots n$

iii) multiple dependent trials in which there are a successful items in a population of n items, and the random variable can take the values $1, 2, 3 \ldots n$

b) Compare the results and explain any similarities or differences.

Processes

Representing
How can you represent these expected values algebraically?

16. The geometric probability distribution involves the probability that a given waiting time will occur before success. A certain traffic light is programmed to be red 40% of the time.

Literacy Link

Waiting time refers to the number of failures before success.

a) What is the probability that your first red light will be on

i) your first trip through the intersection?

ii) your second trip through the intersection?

iii) your third trip through the intersection?

iv) your nth trip through the intersection?

b) Describe how to calculate the probability of an event occurring for the first time after n initial failures, involving independent trials.

c) Derive a formula that calculates this probability.

d) You repeatedly roll a pair of dice until you roll doubles. Build a probability distribution for up to eight rolls of the dice.

e) Compare and contrast the geometric distribution with the binomial and hypergeometric distributions.

Chapter 4 Review

Learning Goals

Section	After this section, I can
4.1	• recognize and identify a discrete random variable • generate a probability distribution by calculating the probabilities for all values of a random variable • represent a probability distribution using a table and a probability histogram • make connections between the frequency histogram and the probability histogram • calculate and interpret the expected value for a probability distribution • make connections between the expected value and the weighted mean of the values of the discrete random variable
4.2	• solve problems involving uniform probability distributions
4.3	• recognize conditions that give rise to a binomial probability distribution • make connections among the table, histogram, and algebraic representation of a binomial probability distribution • solve problems involving binomial probability distributions
4.4	• recognize conditions that give rise to a hypergeometric distribution • calculate the probability associated with each random variable of a hypergeometric distribution • represent the hypergeometric distribution using a table and a probability histogram • solve problems involving hypergeometric probability distributions
4.5	• compare the probability distributions of discrete random variables • solve problems involving uniform, binomial, and hypergeometric distributions

4.1 Probability Distributions, pages 144–153

1. Classify each random variable as discrete or continuous.

a) length of time you play in a hockey game

b) number of times you successfully shoot a basket in a basketball game

c) number of candies in a bag

d) mass of candies in a bag

2. Graph each distribution using a probability histogram.

a)

x	$P(x)$
0	$\frac{1}{12}$
1	$\frac{5}{12}$
2	$\frac{1}{3}$
3	$\frac{1}{6}$

b)

x	$P(x)$
2	0.05
4	0.13
6	0.24
8	0.38
10	0.12
12	0.05
14	0.03

3. Calculate the expectation for each distribution in #2.

4.2 Uniform Distributions, pages 154–159

4. Describe the criteria for a distribution to be uniform.

5. A spinner has six equal sectors, numbered from 1 to 6.

a) Show the probability distribution for a single spin, using a table and a graph.

b) Calculate the expected outcome. Interpret its meaning.

6. An urn contains 25 balls, 40% of which are green. A contestant reaches in the urn to choose three balls; the contestant will win $200 if he or she selects a green ball, but will lose $120 for any other colour. Is each version of the game a fair game? Justify your response.

a) The ball is replaced after each draw.

b) The ball is not replaced after each draw.

4.3 Binomial Distributions, pages 160–169

7. Prepare a distribution table and probability histogram for the number of 5s when a die is rolled six times.

8. The chart shows the percent of Canadians with each blood type.

Blood Type	Percent
O	46
A	42
B	9
AB	3

a) If 120 people are donating blood, what is the expected number of people with type O blood? Why would this be considered a binomial distribution?

b) Calculate the expected number of people with each of the other types of blood.

9. A restaurant gives customers a card with each purchase; customers scratch a box to see if they have won a prize. Twelve percent of the cards are winners.

a) What is the probability of winning a prize only once in 10 tries?

b) What is the probability of winning a prize at least three times in 10 tries?

c) What is the expected number of winning cards in 10 tries?

4.4 Hypergeometric Distributions, pages 170–179

10. a) Prepare a table and a graph for a hypergeometric distribution with $n = 25$, $a = 10$, and $r = 7$.

b) Calculate the expected outcome using two methods.

11. In a collection of 56 coins, 18 are rare. If you select 10 of the coins, what is the probability that

a) all of them are rare?

b) none of them is rare?

c) at least two of them are rare?

12. The fisheries department caught and tagged 420 seals. Recently, 100 seals were caught and 42 had been tagged. Estimate the size of the seal population.

4.5 Comparing and Selecting Discrete Probability Distributions, pages 180–187

13. Classify each situation as uniform, binomial, hypergeometric, or none of these.

a) Forty-five percent of women aged 18 to 25 are currently enrolled in post-secondary education. The random variable is the number of women between the ages of 18 and 25, out of 25 polled, who attend post-secondary education.

b) Twenty out of 30 people at a party are non-smokers. The random variable is the number of smokers in a selection of 8 partiers.

c) The flaws in pieces of timber average 0.2 per metre. The random variable is the number of flaws in the next 50 m of timber.

d) A spinner has 20 equally likely spaces, numbered from 1 to 20. The random variable is the number on which the spinner lands.

14. a) Seven cards are dealt from a standard deck. What is the probability that five are face cards?

b) Seven cards are chosen from a standard deck, with replacement. What is the probability that five are face cards?

c) Compare the two answers in parts a) and b). Explain any differences.

Chapter 4 Test Yourself

✓ Achievement Chart

Category	Knowledge/ Understanding	Thinking	Communication	Application
Questions	1, 2, 3, 4, 5	8, 12, 14, 15, 16	8, 14, 15	6, 7, 9, 10, 11, 13

Multiple Choice

Choose the best answer for #1 to #5.

1. An 8-sided die has its faces numbered 2, 4, 6 … 16. What is the expected outcome on a typical roll?

A 7 **B** 8

C 9 **D** 16

2. The binomial and hypergeometric distributions are similar in that

A they both use independent trials

B they both use dependent trials

C they use the same formula for calculating the expectation

D they both involve counting successes

3. The expectation for a uniform distribution is calculated using

A $\frac{1}{n}\sum_{i=1}^{n} x_i$ **B** $\frac{ra}{n}$

C np **D** $\frac{x}{n}$

4. Counting the number of tails when a coin is flipped 20 times is an example of a

A binomial distribution

B hypergeometric distribution

C uniform distribution

D none of the above

5. The probability that exactly two students will be selected when five people are selected from four students and three teachers is

A $\frac{2}{5}$ **B** $\frac{{}_4C_2 \times {}_3C_3}{{}_7C_5}$

C ${}_5C_2\left(\frac{4}{7}\right)^2\left(\frac{3}{7}\right)^3$ **D** $\frac{5 \times 4}{7}$

Short Response

6. A particular traffic light is programmed to be red 40% of the time. On his daily Monday to Friday commute to and from work, what is the expected number of times Jack can expect to have a red light?

7. Three cards are selected, without replacement, from the honour cards (10, J, Q, K, A) in a standard deck. What is the probability that two of them are face cards?

8. a) Is the situation in #7 modelled by a binomial or a hypergeometric distribution? Explain.

b) Describe how to change the situation to the binomial or hypergeometric distribution, as appropriate.

9. The beaver population in a particular provincial park is known to be 452. Two hundred beavers were caught and tagged. If 65 beavers were later caught and checked for tags, how many would you expect to be tagged?

Extended Response

10. Two dice are rolled a total of eight times, and the sum is recorded each time.

a) Show the probability distribution for a sum of 7.

b) Make a probability histogram of the distribution.

c) Determine and interpret the expected outcome.

11. A certain cell phone provider's help line is busy 95% of the time.

a) In 15 calls to the help line, what is the probability that it will be busy every time? at least 12 times?

b) What is the expected number of times a caller should expect the line to be busy in 15 attempts?

12. Ten males and five females applied for four job promotions. The union's affirmative action committee is concerned that no females were hired, saying that at least one should have been female. Use appropriate calculations to support or refute their claim.

13. The incidence of a disease in the population is 12%. Six people are in an elevator.

a) What is the probability that at least two of them will have the disease?

b) What is the expected number of these people with the disease?

14. Eighteen of thirty players selected in the NHL first–round draft were Canadian. If seven drafted players are randomly selected, what is the probability that

a) only one is Canadian?

b) all are Canadian?

c) most of them are Canadian?

15. If $n \div r > 200$, the binomial distribution can be used to approximate the hypergeometric distribution. Why would this be?

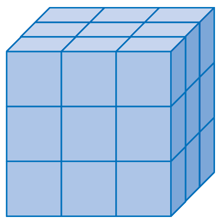
16. Four numbers are chosen from six positive and eight negative numbers. What is the probability that the product of these four numbers will be positive, given that

a) there is no repetition of numbers?

b) repetition of numbers is permitted?

Chapter Problem

Painted Cube



A large cube is painted on all six faces. It is then divided into 27 smaller, congruent cubes.

a) Use a table and a histogram to show the probability distribution for the number of painted faces on a randomly selected cube.

b) If you select 10 of the cubes at the same time, what is the probability that at least half of them will have two painted faces?

c) If you select a cube 10 times, with replacement, what is the probability that at least half of them will have two painted faces?

d) Justify your choice of distributions in parts a), b), and c).

e) For each distribution, calculate the expectation and interpret its meaning.

f) A person is blindfolded and then randomly selects one cube and rolls it. What is the probability that it lands paint side up?

g) Design a simulation that models the distribution in either part b) or c). Conduct 10 trials and compare the results to the theoretical probability that at least half will have two painted faces. Comment on any differences.