Lesson 6: Probability Distribution of a Discrete Random Variable

Student Outcomes

- Given the probability distribution of a discrete random variable in table or graphical form, students describe the long-run behavior of the random variable.
- Given a discrete probability distribution in table form, students construct a graph of the probability distribution.

Lesson Notes

This lesson builds on the prior lesson about discrete probability distributions by asking students to describe the long-run behavior of a random variable. Students should note that there is variability in the long-run behavior. For example, $P(\text{success}) = 0.5$ does not mean that in 100 trials, there will be exactly 50 successes. Understanding the behavior of a random variable provides a sense of what is likely or very unlikely in the long run. For example, if the probability of a seal pup being female is 60%, that does not mean that in every litter there will be 60% females; however, over a long period of time and many samples, the probability will approach 60%. The focus here is not on exactly what is likely to be observed, but on developing an understanding that some values describing long-run behavior of a random variable for a given probability distribution seem reasonable, while others do not.

You may want to discuss the context for the exercises to make sure students understand the situations.

Classwork

Exercises 1–3 (15 minutes): Credit Cards

Students can work through the first three exercises individually or in small groups. Consider providing 3 to 5 minutes for students to first work through the exercises individually. After this time, ask them to listen and share their answers with another student. Discuss the exercises as a class after each group has discussed the work.

MP.1 In the exercises, students create a histogram given a relative frequency table, and use the graph to make sense of given scenarios. Use the histogram to review several previous topics students studied. For example, the histogram indicates a skewed distribution of this data. As a result, the median number of credit cards is the most appropriate measure of center to describe this data. The measure of center for a distribution, often used to indicate a typical value of the data, was an important topic for students in Grades 6 and 9. The histogram students create is a relative frequency histogram that was also developed in Grades 6 and 9. Point out to students that the relative frequencies that summarize this sample have a sum of 1. This result indicates that all of the adults in this sample had between 0 and 10 credit cards. Relative frequencies are used to interpret probability distributions.

Exercises 1–3: Credit Cards

Credit bureau data from a random sample of adults indicating the number of credit cards is summarized in the table below.

Table 1: Number of credit cards carried by adults

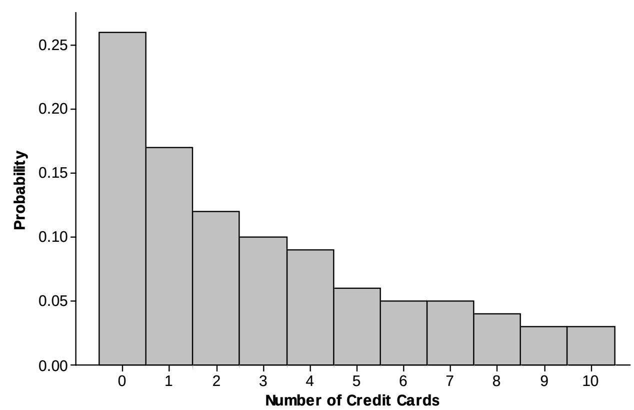
Number of Credit Cards	Relative Frequency
0	0.26
1	0.17
2	0.12
3	0.10
4	0.09
5	0.06
6	0.05
7	0.05
8	0.04
9	0.03
10	0.03

Scaffolding:

- Have students who may be below grade-level model the creation of the histogram or analyze a completed histogram that you provide.
- Have students who are above grade-level construct their own hypothetical histogram and explain the meaning of the probabilities in context.

1. **Consider the chance experiment of selecting an adult at random from the sample. The number of credit cards is a discrete random variable. The table above sets up the probability distribution of this variable as a relative frequency. Make a histogram of the probability distribution of the number of credit cards per person based on the relative frequencies.**

   Answer:



2. **Answer the following questions based on the probability distribution.**

   a. **Describe the distribution.**

   Responses will vary.

   The distribution is skewed right with a peak at 0. The mean number of cards (the balance point of the distribution) is about 3 or 4. The median number of cards is between 2 or 3 cards. Adults carry anywhere from 0 to 10 credit cards.

   b. **Is a randomly selected adult more likely to have 0 credit cards or 7 or more credit cards?**

   The probability that an adult has no credit cards is 0.26, while the probability of having 7 or more credit cards is about 0.15, so the probability of having no credit cards is larger.

c. **Find the area of the bar representing 0 credit cards.**

The area is $0.26 \cdot 1 = 0.26$.

d. **What is the area of all of the bars in the histogram? Explain your reasoning.**

The total area is 1 because the area of each bar represents the probability of one of the possible values of the random variable, and the sum of all of the possible values is 1.

3. **Suppose you asked each person in a random sample of 500 people how many credit cards he or she has. Would the following surprise you? Explain why or why not in each case.**

a. **Everyone in the sample owned at least one credit card.**

This would be surprising because 28% of adults do not own a credit card. It would be unlikely that in our sample of 500, no one had zero credit cards.

b. **65 people had 2 credit cards.**

This would not be surprising because 12% (0.12) of adults own two credit cards, so we would expect somewhere around 60 out of the 500 people to have two credit cards. 65 is close to 60.

c. **300 people had at least 3 credit cards.**

Based on the probability distribution, about 45% of adults have at least 3 credit cards, which would be about 225. 300 is greater than 225, but not enough to be surprising.

d. **150 people had more than 7 credit cards.**

This would be surprising because about 10% of adults own more than 7 credit cards. In this sample, $\frac{150}{500}$ or about 30% (three times as many as the proportion in the population) own more than 7 credit cards.

Before moving on to the next exercise set, check for understanding by asking students to answer the following:

- Explain to your neighbor how the histogram enabled us to answer questions about probabilities.

Exercises 4–7 (22 minutes): Male and Female Pups

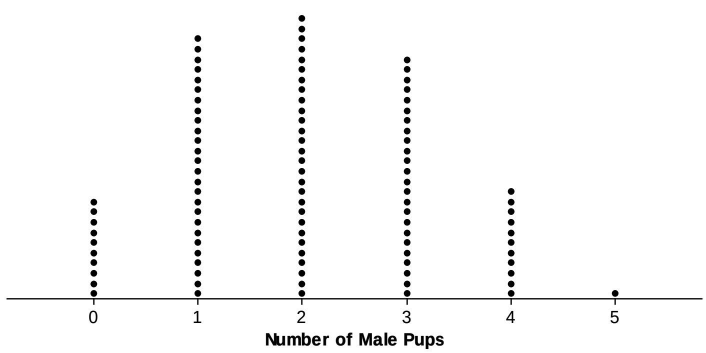
Depending on the class, these exercises might be done as a whole class discussion or with students working in small groups. Exercise 4 is intended to remind students of their earlier work with probability. This exercise describes a scenario involving animals, some species of seals, for example, that have biased sex ratios in their offspring. Biased sex ratios are ratios that are different from the expected probability due to conditions in the population studied. The reasons for this, and the reasons for studying the sex ratios of certain animals, are based on understanding the survival of the animal. Ask students to think of reasons why the probability of a female might be greater than the probability of a male for certain animals in order for the species to survive. Consider encouraging students to do independent research on the probability of a male or a female at birth for seals, elephants, or other endangered species of animals.

To help students think about some of the questions related to the scenario described in Exercise 4, you might have students look at data from a simulation. Conducting a simulation that records the number of males in a litter of the seals described in Exercise 4 will indicate variability, as well as how the long-run behavior follows the pattern of the probability distribution. For this lesson, generate 100 random samples of size 6 from the set $\{1, 0, 0\}$. Selecting a 1 from this sample set represents the birth of a male; selecting a 0 represents selecting a female. Have students work in groups to create random selections of 6 (with replacement) from this set. The 6 random selections represent a litter. Consider using the random features of graphing calculators or other probability software to simulate 100 litters of 6 pups. Organize students to record the number of males from the simulated litter. You might also consider having students record the simulated number of males in a litter on a whiteboard or class poster, or have students post their results to a class dot plot.

Simulated results of the number of male pups in a litter of 6 was conducted and summarized below. Use this data distribution if there is limited time to conduct a simulation.

1	2	2	4	1	2	2	2	2	1
2	1	1	1	2	3	1	3	3	3
4	0	3	0	1	1	4	2	2	1
3	2	4	1	2	3	3	0	3	4
2	3	1	2	3	2	0	2	1	3
4	1	2	3	2	1	3	4	3	4
2	4	5	2	3	3	2	0	1	3
3	1	3	4	1	2	0	3	2	4
1	3	2	1	0	0	1	2	1	0
0	1	3	2	1	1	3	1	2	2

A graph of this distribution is also provided below and will help students connect the results to the probability distribution. Provide this graph or as a class develop the dot plot for the simulated 100 litters of seals.



The graph will help students answer several of the questions in the exercises. For example, in thinking about Exercise 6, part (a), students could look at the first five litters in the simulated distribution, the second five, and so on to see about how many of the litters had fewer males than females. They could count the number of times one male showed up in every two litters for Exercise 6, part (b). See table below.

Exercises 4–7: Male and Female Pups

4. The probability that certain animals will give birth to a male or a female is generally estimated to be equal, or approximately 0.50. This estimate, however, is not always the case. Data are used to estimate the probability that the offspring of certain animals will be a male or a female. Scientists are particularly interested about the probability that an offspring will be a male or a female for animals that are at a high risk of survival. In a certain species of seals, two females are born for every male. The typical litter size for this species of seals is six pups.

   a. What are some statistical questions you might want to consider about these seals?

   Statistical questions to consider include the number of females and males in a typical litter, or the total number of males or females over time. (The question about the number of males in a typical litter will be explored in the exercises.)

   b. What is the probability that a pup will be a female? A male? Explain your answer.

   Out of every three animals that are born, one is male and two are female.

   $\frac{1}{3}$ *probability of a male,* $\frac{2}{3}$ *probability of a female*

   c. Assuming that births are independent, which of the following can be used to find the probability that the first two pups born in a litter will be male? Explain your reasoning.

   i. $\frac{1}{3}+\frac{1}{3}$

   ii. $\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)$

   iii. $\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)$

   iv. $2\left(\frac{1}{3}\right)$

   The two probabilities are multiplied if the events are independent; these are independent events, so the probability of the first two pups being male is $\left(\frac{1}{3}\right)\left(\frac{1}{3}\right)=\frac{1}{9}$.

5. The probability distribution for the number of males in a litter of six pups is given below.

Table 2: Probability distribution of number of male pups per litter*

Number of male pups	Probability
0	0.088
1	0.243
2	0.330
3	0.220
4	0.075
5	0.018
6	0.001

*The sum of the probabilities in the table is not equal to 1 due to rounding.

Use the probability distribution to answer the following questions.

   a. How many male pups will typically be in a litter?

   The most common will be two males and four females in a litter, but it would also be likely to have one to three male pups in a litter.

b. **Is a litter more likely to have six male pups or no male pups?**

It is more likely to have no male pups (probability of no males is 0.088) than the probability of all male pups (0.001).

6. **Based on the probability distribution of the number of male pups in a litter of six given above, indicate whether you would be surprised in each of the situations. Explain why or why not.**

a. **In every one of a female's five litters of pups, there were fewer males than females.**

Responses will vary.

This seems like it would not be surprising because in each litter, the chance of having more females than males is $0.33 + 0.243 + 0.088 = 0.621$. The probability that this would happen in five consecutive litters would be $(0.621)^5$, which is much smaller than 0.09, but which might be considered not too unlikely.

Scaffolding:

The word *litter* may be familiar to English language learners in terms of trash. Point out that in this context, *litter* refers to a group of young animals born at the same time to the same animal.

b. **A female had only one male in two litters of pups.**

Responses will vary.

The probability of no males in a litter is about 0.088 and of one male in a litter is 0.243, for a probability of 0.33 for either case (the two litters are independent of each other, so you can add the probabilities). While having only one male in two litters might be somewhat unusual, the probability of happening twice in a row would be 0.11, which is not too surprising if it happened twice in a row.

c. **A female had two litters of pups that were all males.**

Responses will vary.

This would be surprising because the chance of having all males is 0.001, and to have two litters with all males would be unusual (0.000001).

d. **In a certain region of the world, scientists found that in 100 litters born to different females, 25 of them had four male pups.**

Responses will vary.

This would be surprising because it shows a shift to about $\frac{1}{4}$ of the litters having four males, which is quite a bit larger than the probability of 0.075 given by the distribution for four males per litter where the probability of a male is $\frac{1}{3}$.

MP.2

7. **How would the probability distribution change if the focus was the number of females rather than the number of males?**

Responses will vary.

The probabilities would be in reverse order. The probability of 0 females (all males) would now be the probability of 6 females (no males), the probability of 1 female would be the same as the probability of 5 males, and so on.

Closing (3 minutes)

- Students should understand that probabilities are interpreted as long-run relative frequencies but that in a finite number of observations, the observed proportions will vary slightly from those predicted by a probability distribution.
- Ask students to summarize the key ideas of the lesson in writing or by talking to a neighbor. Use this as an opportunity to informally assess student understanding. The lesson summary provides some of the key ideas from the lesson.

> Lesson Summary
>
> **The probability distribution of a discrete random variable in table or graphical form describes the long-run behavior of a random variable.**

Exit Ticket (5 minutes)

Name ____ Date ____________________

Lesson 6: Probability Distribution of a Discrete Random Variable

Exit Ticket

The following statements refer to a discrete probability distribution for the number of songs a randomly selected high school student downloads in a week, according to an online music library.

Probability distribution of number of songs downloaded by high school students in a week:

Number of songs	0	1	2	3	4	5	6	7	8	9
Probability	0.06	0.14	0.22	0.25	0.15	0.09	0.05	0.024	0.011	0.005

Which of the following statements seem reasonable to you based on a random sample of 200 students? Explain your reasoning, particularly for those that are unreasonable.

a. 25 students downloaded 3 songs a week.

b. More students downloaded 4 or more songs than downloaded 3 songs.

c. 30 students in the sample downloaded 9 or more songs per week.

Exit Ticket Sample Solutions

The following statements refer to a discrete probability distribution for the number of songs a randomly selected high school student downloads in a week, according to an online music library.

Probability distribution of number of songs downloaded by high school students in a week:

Number of songs	0	1	2	3	4	5	6	7	8	9
Probability	0.06	0.14	0.22	0.25	0.15	0.09	0.05	0.024	0.011	0.005

Which of the following statements seem reasonable to you based on a random sample of 200 students? Explain your reasoning, particularly for those that are unreasonable.

a. 25 students downloaded 3 songs a week.

Responses will vary.

If the probability is 0.25 that a student will download 3 songs, then in 200 students, it would seem reasonable to have around 50 students downloading 3 songs. To have only 25 does not seem reasonable.

b. More students downloaded 4 or more songs than downloaded 3 songs.

Responses will vary.

The probability of 4 or more songs is $0.15 + 0.09 + 0.05 + 0.024 + 0.011 + 0.005 = 0.33$, which is larger than 0.25, so it would seem reasonable to have more students downloading 4 or more songs.

c. 30 students in the sample downloaded 9 or more songs per week.

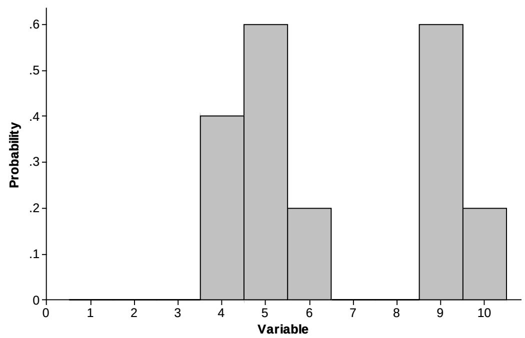
Responses will vary.

The probability that a student would download 9 or more songs per week is 0.005 or 0.5%; 30 out of 200 is about 15%. This does not seem reasonable.

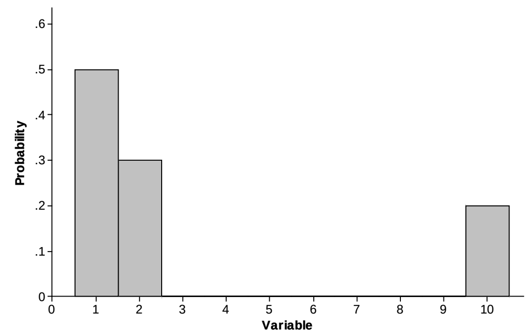
Problem Set Sample Solutions

1. **Which of the following could be graphs of a probability distribution? Explain your reasoning in each case.**

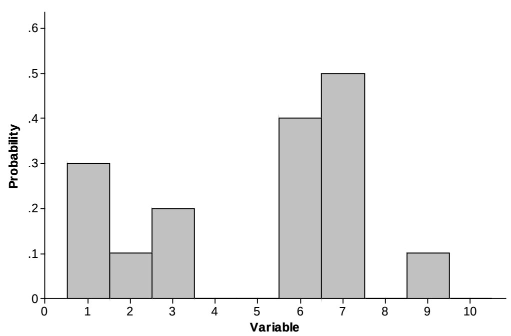
a.



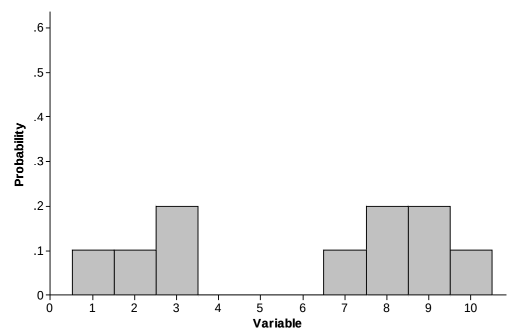
b.



c.



d.



Graphs (b) and (d) are probability distributions of discrete random variables because the outcomes are discrete numbers from 1 to 10, the probability of every outcome is less than 1, and the sum of the probabilities of all the outcomes is 1. The data distribution for graphs (a) and (c) are not probability distributions of discrete random variables because the sum of the probabilities is greater than 1.

2. Consider randomly selecting a student from New York City schools and recording the value of the random variable number of languages in which the student can carry on a conversation. A random sample of $1,000$ students produced the following data.

Table 3: Number of languages spoken by random sample of students in New York City

Number of languages	1	2	3	4	5	6	7
Number of students	542	280	71	40	34	28	5

a. Create a probability distribution of the relative frequencies of the number of languages students can use to carry on a conversation.

Answer:

Table 4: Number of languages spoken by random sample of students in New York City

Number of languages	1	2	3	4	5	6	7
Probability	0.542	0.280	0.071	0.040	0.034	0.028	0.005

b. If you took a random sample of 650 students, would it be likely that 350 of them only spoke one language? Why or why not?

About 54% of all the students speak only one language; $\frac{350}{650}$ is about 54%, so it seems likely to have 350 students in the sample who could carry on a conversation in only one language.

c. If you took a random sample of 650 students, would you be surprised if 100 of them spoke exactly 3 languages? Why or why not?

$\frac{100}{650}$ is about 15%. The data suggest that only about 7% of all students speak exactly 3 languages. This could happen but does not seem too likely.

d. Would you be surprised if 448 students spoke more than two languages? Why or why not?

$\frac{448}{650}$ is about 69%, which seems to be a lot more than the 18% of all students who speak more than two languages, so I would be surprised.

3. Suppose someone created a special six-sided die. The probability distribution for the number of spots on the top face when the die is rolled is given in the table.

Table 5: Probability distribution of the top face when rolling a die

Face	1	2	3	4	5	6
Probability	$\frac{1-x}{6}$	$\frac{1-x}{6}$	$\frac{1-x}{6}$	$\frac{1+x}{6}$	$\frac{1+x}{6}$	$\frac{1+x}{6}$

a. If x is an integer, what does x have to be in order for this to be a valid probability distribution?

The probabilities have to add to 1, so we need $6-3x+3x=6$. However, this is true for any value of x. Since the probabilities also have to be greater than or equal to 0, x can only be 1 or 0.

b. Find the probability of getting a 4.

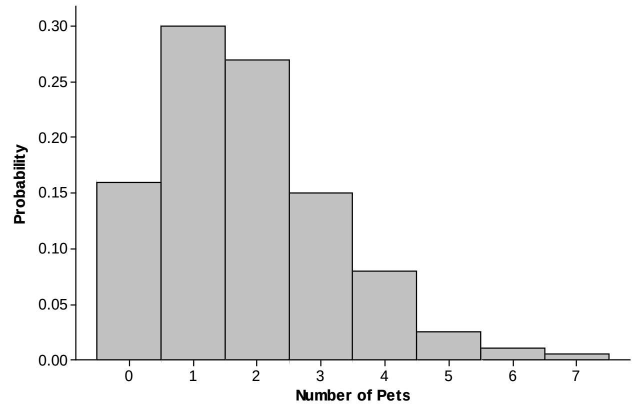
If $x = 0$, $P(4)=\frac{1}{6}$; if $x = 1$, then $P(4) = \frac{1}{3}$.

c. **What is the probability of rolling an even number?**

If $x = 0$, *then* $P(\text{even}) = \frac{3}{6} = \frac{1}{2}$.

If $x = 1$, *then* $P(\text{even}) = \frac{4}{6} = \frac{2}{3}$.

4. **The graph shows the relative frequencies of the number of pets for households in a particular community.**



a. **If a household in the community is selected at random, what is the probability that a household would have at least 1 pet?**

0.84

b. **Do you think it would be likely to have 25 households with 4 pets in a random sample of 225 households? Why or why not?**

$\frac{25}{225}$ *is about* 11%. *The probability distribution indicates about* 8% *of the households would have* 4 *pets. These are fairly close, so it would seem reasonable to have* 25 *households with* 4 *pets.*

c. **Suppose the results of a survey of 350 households in a section of a city found 175 of them did not have any pets. What comments might you make?**

Responses will vary.

Based on this probability distribution, there should be about 16% *or* 56 *of the* 350 *households without pets. To have* 175 *houses without pets suggests that there is something different; perhaps the survey was not random, and it included areas where pets were not allowed.*