Lesson 12: Estimating Probability Distributions Empirically

Student Outcomes

- Students use empirical data to estimate probabilities associated with a discrete random variable.
- Students interpret probabilities in context.

Lesson Notes

This lesson engages students in collecting empirical data associated with a discrete random variable by simulation. You may want to have them each do a few trials of the simulation and pool the data as a class before discussing the outcomes. Or, you may choose to have a small group of students collectively do as many as 50 trials of the simulation and then compare the results across groups. The use of technology for generating random samples will make the work easier for students, and some of the simulations can be completely done with simple steps to generate, collect, and display results.

In the first example, each student will need 2 dice if you plan to have them carry out a physical simulation. Or, you may choose to use technology to simulate tossing the dice. Students should use technology for the calculations so they can focus on the statistical concepts. Students make a dot plot of the results of the simulation and use that to estimate the probabilities in a probability distribution. In the second example, they go directly from the simulation to constructing a graph of the estimated probability distribution. Be sure they understand that the simulations produce frequencies, and these must be converted into relative frequencies to create the estimated probability distribution.

This lesson may take more than one class period. If you are short on time, consider choosing to do either Example 1 and Exercises 1–2 or Example 2 and Exercises 3–5.

Classwork

Exploratory Challenge 1/Exercises 1–2 (18 minutes): Moving Along

Students investigate tossing 2 dice and moving along a number line according to the absolute value of the difference of numbers showing on the faces. For example, if they toss a 6 and a 3 on the first toss, they would move from 0 to 3 on the number line. They may actually play the game, or they can simulate tossing 2 dice. They record the distance moved for each toss and the sum of the distances moved, targeting a total distance of at least 20. Using these data, students find the expected value and interpret it in context.

MP.5 Let students work with a partner to toss the dice and play a few games. Students should keep track of the distance moved for each toss and the number of tosses it takes to move past 20 on the number line. After playing the game a few times, have students work on Exercises 1 and 2 with their partner to carry out the physical (or technological simulation) and estimate probabilities. (If using technology, students may choose to use either a graphing calculator or computer software to perform the simulation they describe in their answer to Exercise 1.)

Exploratory Challenge 1/Exercises 1–2: Moving Along

In a certain game, you toss 2 dice and find the difference of the numbers showing on the faces. You move along a number line according to the absolute value of the difference. For example, if you toss a 6 and a 3 on the first toss, then you move 3 spaces from your current position on the number line. You begin on the number 0, and the game ends when you move past 20 on the number line.

1. **How many rolls would you expect it to take for you to get to 20? Explain how you would use simulation to answer this question.**

   Answers will vary. It appears that many rolls result in a difference of 1, 2, or 3. Considering that I need to move 20 spaces, I would say that I expect that it would take around 8, 9, or 10 rolls to get to 20.

   The random variable of interest is the distance moved on the number line for a toss of the dice. The possible values are 0, 1, 2, 3, 4, and 5. I would play the game with my partner at least 30 times and record the distance moved for each toss. Then, I would use the results to create an estimated probability distribution to determine expected value for the distance moved on 1 toss of the dice.

2. **Perform the simulation with your partner.**

MP.2

   a. **What is the expected value for the distance moved on 1 toss of 2 dice? Interpret your answer in terms of playing the game.**

   Responses will vary. One possible response based on 60 tosses is shown here:

   Distance moved on 1 toss of 2 dice

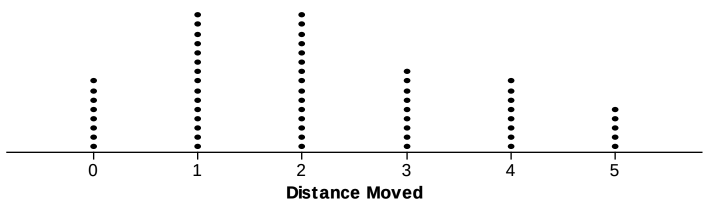


Table: Estimated probability distribution for the distance moved on 1 toss of 2 dice

Number of Moves	0	1	2	3	4	5
Estimated Probability	$\frac{8}{60} = 0.133$	$\frac{15}{60} = 0.25$	$\frac{15}{60} = 0.25$	$\frac{9}{60} = 0.15$	$\frac{8}{60} = 0.133$	$\frac{5}{60} = 0.083$

   The example results are

   $0(0.133) + 1(0.25) + 2(0.25) + 3(0.15) + 4(0.133) + 5(0.083) = 2.147.$

   In the long run, after tossing 2 dice many times, we expect the average distance moved to be close to 2.15.

   b. **Use your expected value from part (a) to find the expected number of tosses that would put you past 20 on the number line.**

   Responses will vary.

   For the sample data, in 10 tosses, you would expect to be at about 21 or 22.

Exploratory Challenge 2/Exercises 3–4 (15 minutes): Lemon Flavor

MP.5

Students can work alone or in small groups to collect their data. One strategy for carrying out the simulation is to generate about 200 random numbers from the set $\{0,1\}$ in a list with 1 representing lemon. Count down the list marking the number of cough drops before you have two 1s in a row. Note that the question is not whether there are 2 lemon-flavored cough drops in a row but rather how far into the package you have to go to get 2 lemon-flavored ones in a row.

Have students work alone or in a small group to complete the exercises.

Exploratory Challenge 2/ Exercises 3–4: Lemon Flavor

Cough drops come in a roll with 2 different flavors, lemon and cherry. The same number of lemon and cherry cough drops are produced. Assume the cough drops are randomly packed with 30 per roll and that the flavor of a cough drop in the roll is independent of the flavor of the others.

3. **Suppose you really liked the lemon flavor. How many cough drops would you expect to go through before finding 2 lemon cough drops in a row? Explain how you would use simulation to answer this question.**

   Responses will vary. Students might answer anywhere from 3 to 20.

   To simulate, I would generate random numbers from the set $\{0, 1\}$, with 1 representing lemon. The random variable is the number of cough drops before you have 2 lemon cough drops in a row. The possible values are $2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,$ on up to 30. In a case where 2 lemon cough drops do not occur in the pack of 30, I would use 0 to represent the outcome.

   I would simulate the experiment at least 50 times and record the number of cough drops I would go through before getting 2 lemon-flavored ones in a row. I would use the results to create a probability distribution that could be used to estimate the number of cough drops I expect to go through before finding 2 lemon-flavored ones in a row.

MP.2

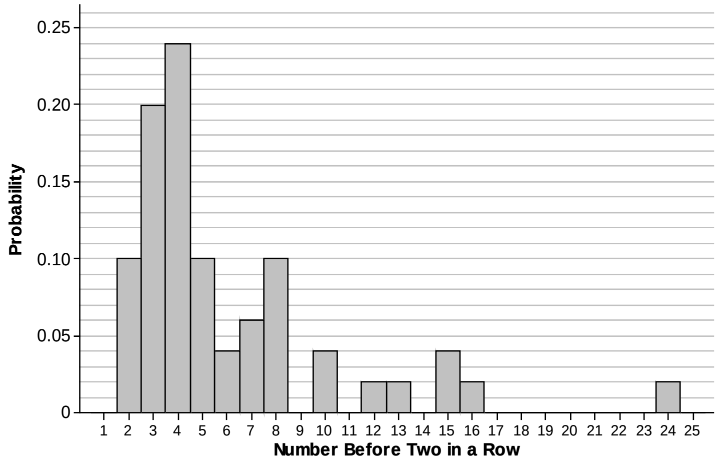
4. **Carry out the simulation and use your data to estimate the average number of cough drops you would expect to go through before you found 2 lemon-flavored ones in a row. Explain what your answer indicates about 2 lemon-flavored cough drops in a row.**

   Responses will vary. Sample results are below. Students might be surprised that in one case, it took 24 cough drops before they had 2 lemon-flavored ones in a row.

   Number before 2 in a row/frequency:

2	11111
3	1111111111
4	111111111111
5	11111
6	11
7	111
8	11111
9	

10	11
11	
12	1
13	1
14	
15	11
16	1
⋮	⋮
24	1



Using sample data above, the expected value would be 6.02. If you opened many packages of cough drops, on average you would go through about 6 cough drops before you got 2 lemon-flavored ones in a row.

Closing (2 minutes)

- How was simulation helpful in this lesson? What tools were used to help answer the questions posed in this lesson?
  - *Responses will vary. Sample response:*

    Simulation was used to help build estimated probability distributions. I used my graphing calculator to simulate the tossing of the 2 dice for my plan in Exploratory Challenge 1/Exercise 1.

- How did the estimated probability distributions for the distance moved in the game in Exploratory Challenge 1/Exercise 1 vary across the class?
  - *Responses will vary. Sample response:*

    They were fairly close but not exactly the same because of the variability in the outcomes of tossing 2 dice.
- Comment on the following statement: If the estimated expected value is 12 for the number of cough drops in a roll before 2 lemon-flavored ones, then you would expect to have 10 that were not lemon at the beginning.
  - *Responses will vary. Sample response:*

    You could have a mix of 10 lemon and cherry flavors but with no 2 lemons in a row. The eleventh and twelfth ones would be lemon.
- Ask students to summarize the main ideas of the lesson in writing or with a neighbor. Use this as an opportunity to informally assess comprehension of the lesson. The Lesson Summary below offers some important ideas that should be included.

Lesson Summary

In this lesson, you learned that

- **You can estimate probability distributions for discrete random variables using data from simulating experiments.**
- **Probabilities from an estimated probability distribution for a discrete random variable can be interpreted in terms of long-run behavior of the random variable.**
- **An expected value can be calculated from an estimated probability distribution and interpreted as a long-run average.**

Exit Ticket (10 minutes)

Name ____ Date ____________________

Lesson 12: Estimating Probability Distributions Empirically

Exit Ticket

A bus company has 9 seats on a shuttle between two cities, but about 10% of the time people do not show up for the bus even though they reserve a seat. The company compensates by reserving 11 seats instead of 9. Assume that whether or not a person with a reservation shows up is independent of what happens with the other reservation holders.

a. Consider the random variable number of people who are denied a seat because more than 9 people showed up for the shuttle. What are the possible values of this random variable?

b. The table displays the number of people who reserved tickets but did show up based on simulating 50 trips between the two cities. Use the information to estimate a probability distribution of the number of people denied a seat on the shuttle.

Table: Number of people who showed up for their reservation

10	10	11	9	10	11	9	9	11	11
7	10	10	8	9	10	11	9	10	10
8	11	11	10	11	10	10	11	9	11
11	10	8	11	9	11	9	10	11	9
10	9	9	10	10	9	9	10	10	9

c. In the long run, how many people should the company expect to be denied a seat per shuttle trip? Explain how you determined the answer.

Exit Ticket Sample Solutions

1. A bus company has 9 seats on a shuttle between two cities, but about 10% of the time people do not show up for the bus even though they reserve a seat. The company compensates by reserving 11 seats instead of 9. Assume that whether or not a person with a reservation shows up is independent of what happens with the other reservation holders.

   a. Consider the random variable number of people who are denied a seat because more than 9 people showed up for the shuttle. What are the possible values of this random variable?

   0, 1, and 2

   b. The table displays the number of people who reserved tickets but did show up based simulating 50 trips between the two cities. Use the information to estimate a probability distribution of the number of people denied a seat on the shuttle.

   Table: Number of people who showed up for their reservation

10	10	11	9	10	11	9	9	11	11
7	10	10	8	9	10	11	9	10	10
8	11	11	10	11	10	10	11	9	11
11	10	8	11	9	11	9	10	11	9
10	9	9	10	10	9	9	10	10	9

   Table: Number of people denied a seat on shuttle

Number denied a seat	2	1	0
Frequency	14	18	18
Relative frequency	0.28	0.36	0.36

   c. In the long run, how many people should the company expect to be denied a seat per shuttle trip? Explain how you determined the answer.

   The company should expect about 0.92 passengers will be denied a seat over the long run, or just less than 1 per trip. I used the sample data to determine the expected value: $0 \cdot 0.36 + 1 \cdot 0.36 + 2 \cdot 0.28 = 0.92$.

Problem Set Sample Solutions

1. **Suppose the rules of the game in Exploratory Challenge 1 changed.**

   If you had an absolute difference of

   - **3 or more, you move forward a distance of 1;**
   - **1 or 2, you move forward a distance of 2;**
   - **0, you do not move forward.**

   a. **Use your results from Exploratory Challenge 1/Exercise 2 to estimate the probabilities for the distance moved on 1 toss of 2 dice in the new game.**

   Responses will vary. Sample answer:

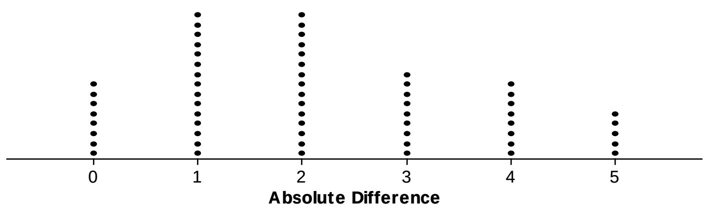


Table: Distance moved on 1 toss of the dice

Distance Moved	0 *(difference* = 0)	2 *(difference* = $1, 2$)	1 *(difference* = $3, 4, 5$)
Probability	$\frac{8}{60} = 0.133$	$\frac{15}{60} + \frac{15}{60} = 0.5$	$\frac{9}{60} + \frac{8}{60} + \frac{5}{60} = 0.366$

   b. **Which distance moved is most likely?**

   Responses will vary. Sample response:

   For the sample data, a distance of 2 is the most likely with a probability of 0.5.

   c. **Find the expected value for distance moved if you tossed 2 dice 10 times.**

   Responses will vary. Sample response:

   For the sample data, expected distance moved in 1 toss is $0(0.133) + 2(0.5) + 1(0.366) = 1.366$, so the expected distance moved in 10 tosses is 13.66.

   d. **If you tossed the dice 20 times, where would you expect to be on the number line, on average?**

   Responses will vary. Sample response:

   For the sample data, over the long run, you would expect to be at about 27 on the number line.

2. Suppose you were playing the game of Monopoly, and you got the Go to Jail card. You cannot get out of jail until you toss a double (the same number on both dice when 2 dice are tossed) or pay a fine.

   a. If the random variable is the number of tosses you make before you get a double, what are possible values for the random variable?

   $1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, \ldots$

   b. Create an estimated probability distribution for how many times you would have to toss a pair of dice to get out of jail by tossing a double. (You may toss actual dice or use technology to simulate tossing dice.)

   Responses will vary.

   For the sample data below, an estimated probability distribution would be as follows:

   Table: Number of times tossing a pair of dice before a double occurs

Number rolls	1	2	3	4	5	6	7	8
Probability	0.08	0.2	0.12	0.08	0.08	0.04	0.04	0.04
Number rolls	9	10	11	12	13	14	15	
Probability	0.08			0.08		0.04	0.04	

The table below displays the simulated tosses. The table is read down each column, from left to right. The first number represents the difference in the faces of simulated tosses of 2 dice, and the number 0 represents a double was tossed. All doubles have been highlighted. The number of times the dice were tossed between doubles appears in parentheses.

There were 8 tosses between doubles.

$-1.$	$3.$	$0.(4)$	$3.$	$-4.$	$0.(2)$	$-2.$	$2.$
$0.(2)$	$1.$	$4.$	$-4.$	$3.$	$2.$	$-2.$	$-1.$
$2.$	$0.(3)$	$4.$	$-3.$	$-3.$	$-1.$	$5.$	$-1.$
$-3.$	$3.$	$-2.$	$1.$	$0.(14)$	$2.$	$0.(9)$	$-1.$
$-1.$	$-3.$	$-4.$	$3.$	$1.$	$0.(4)$	$1.$	$-2.$
$1.$	$0.(3)$	$-3.$	$5.$	$-2.$	$3.$	$1.$	$0.(12)$
$-3.$	$0.(1)$	$-1.$	$-4.$	$2.$	$-4.$	$4.$	$1.$
$-4.$	$-4.$	$-4.$	$5.$	$5.$	$-1.$	$4.$	$0.(2)$
$1.$	$-5.$	$1.$	$3.$	$0.(5)$	$-3.$	$-1.$	$-1.$
$0.(8)$	$-4.$	$0.(9)$	$2.$	$2.$	$-1.$	$-1.$	$1.$

$-3.$	$-1.$	$4.$	$0.(12)$	$-3.$	$-1.$	$-1.$	$-3.$
$3.$	$3.$	$1.$	$0.(1)$	$0.(3)$	$1.$	$-4.$	$0\ (2)$
$1.$	$4.$	$-3.$	$-1.$	$3.$	$2.$	$-1.$	
$0.(6)$	$-4.$	$4.$	$4.$	$0.(2)$	$-2.$	$2.$	
$-2.$	$1.$	$2.$	$-1.$	$-2.$	$0.(6)$	$-5.$	
$-2.$	$-2.$	$-2.$	$4.$	$-3.$	$-2.$	$-2.$	
$2.$	$2.$	$-1.$	$2.$	$1.$	$3.$	$-1.$	
$-2.$	$-2.$	$3.$	$-4.$	$2.$	$2.$	$-2.$	
$-1.$	$0.(15)$	$-5.$	$0.(7)$	$0.(5)$	$1.$	$1.$	
$4.$	$-2.$	$-3.$	$-1.$	$-1.$	$4.$	$0.(15)$	

Summary of number of tosses of dice from simulation

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
11	11111	111	11	11	11	1	1	11			11		1	11

c. What is the expected number of tosses of the dice before you would get out of jail with a double?

Responses will vary.

For the sample data, the expected number is 4.16. Over the long run, you would expect to toss the dice about 4 times before you got a double.

3. The shuttle company described in the Exit Ticket found that when they make 11 reservations, the average number of people denied a seat per shuttle is about 1 passenger per trip, which leads to unhappy customers. The manager suggests they take reservations for only 10 seats. But his boss says that might leave too many empty seats.

a. Simulate 50 trips with 10 reservations, given that in the long run, 10% of those who make reservations do not show up. (You might let the number 1 represent a no-show and a 0 represent someone who does show up. Generate 10 random numbers from the set that contains one 1 and nine 0s to represent the 10 reservations and then count the number of $1'$s.)

Responses will vary. Sample response:

Table: Simulated results of the number of no shows given that 10% are no shows

Number of no shows	0	1	2	3	4
Frequency	11	22	15	2	0

b. If 3 people do not show up for their reservations, how many seats are empty? Explain your reasoning.

2 seats are empty because there are only 9 seats on the shuttle.

c. Use the number of empty seats as your random variable and create an estimated probability distribution for the number of empty seats.

Responses will vary.

Note that there are no empty seats when 9 or 10 people with reservations show up because there are only 9 seats on the shuttle.

Table: Number of empty seats on shuttle

Number of empty seats	0	1	2	3
Frequency	33	15	2	0
Relative frequency	0.66	0.30	0.04	0

d. What is the expected value for the estimated probability distribution? Interpret your answer from the perspective of the shuttle company.

Responses will vary.

For the sample data, the expected value is 0.38, which means that over the long run, 0.38 seats are empty per shuttle.

e. How many reservations do you think the shuttle company should accept and why?

Responses will vary.

In the long run, about 1 person per trip would be denied a seat if they make 11 reservations, and they will have an empty seat over $\frac{1}{3}$ of the time if they make only 10 reservations. What they choose to do would depend on how much they have to compensate those who are denied seats.

Table of Random Numbers

00000110010001100011

10111001111010001100

01101001011011010101

11011011010100110010

01110001000100110011

00010000001100111011

11101001010010000110

10110001110001000100

11101100101101100110

11100010010011100011