

Geometric Distributions

ID: 12101

 Time required
 45 minutes

Activity Overview

*In this activity, students will simulate a geometric distribution of rolling a die until a 5 appears. They will determine experimental probabilities and draw conclusions about the number of rolls it takes for the first 5 to appear. Students will then use the **geometpdf** command to calculate the theoretical probabilities and use a scatter plot to draw conclusions. They will also calculate expected value.*

Topic: Random Distributions

- *Geometric Distributions*
- *Cumulative Distribution*
- *Probability*
- *Simulations*

Teacher Preparation and Notes

- *This lesson is intended to be teacher-led to help students derive the formulas for geometric distributions and the expected value. Two extension problems are included at the end of the activity. The student worksheet is used to help students take notes and draw conclusions.*
- *Students will need to be familiar with binomial distributions to answer the first three questions. They also need to know combinations and basic probability rules to derive the formulas.*
- ***To download the student worksheet, go to education.ti.com/exchange and enter "12101" in the quick search box.***

Associated Materials

- *StatWeek13_GeoDist_worksheet_TI84.doc*

Suggested Related Activities

To download any activity listed, go to education.ti.com/exchange and enter the number in the quick search box.

- *Geometric Dartboards (TI-Nspire technology) — 8269*
- *Binomial and Geometric Distributions (TI-Navigator, TI-84 Plus family) — 8395*
- *Geometric Distribution (TI-Navigator) — 1956*
- *Law of Large Numbers: Adding it up(TI-84 Plus family) — 4238*

Problem 1 – Introduction

Introduce the definition for a geometric distribution, given on the student worksheet. Contrast this definition with the definition for a binomial distribution.

Discuss with students the 6 situations on their worksheet and why each variable is or is not geometrically distributed. Answers:

1. Yes. This situation meets all the requirements of a geometric distribution.
2. No. The variable of interest is the total number of boys expected, not the number of trials required to obtain the first success.
3. Yes. This situation meets all the requirements of a geometric distribution.
4. Yes. Even though the probability of success is small, this situation meets all the requirements of a geometric distribution.
5. No. Drawing a card from a deck without replacement is not independent.
6. No. The variable of interest is the total number of doubles expected, not the number of trials required to obtain the first success.

Problem 2 – Simulation

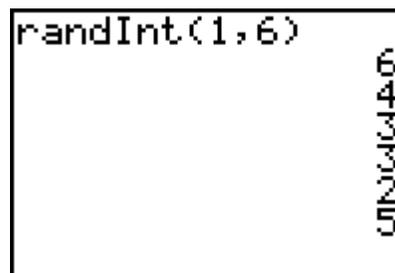
Students will use a simulation to find the number of dice rolls it takes to obtain the first 5. They will begin with a simpler question: What is the probability that it takes four rolls to roll the first 5?

Step 1: On the Home screen, students are to enter **randInt(1,6)** and then press **ENTER** until a 5 appears. They will need to count each entry until a 5 appears.

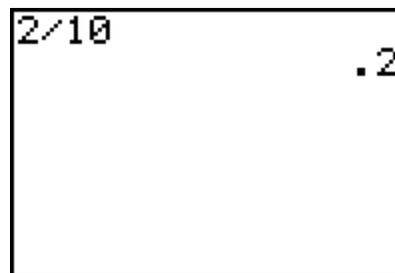
Step 2: Students are to press **STAT** **ENTER** and use list **L1** to record the number of simulated rolls it took to obtain the first 5.

Students need to repeat steps 1 and 2 until they have completed 10 simulations. They can press **CLEAR** to clear the screen after each simulation if needed.

Students are to determine their experimental probability of the first 5 appearing on the fourth roll by counting the number of 4s in L1 and dividing that number by 10, the number of trials.

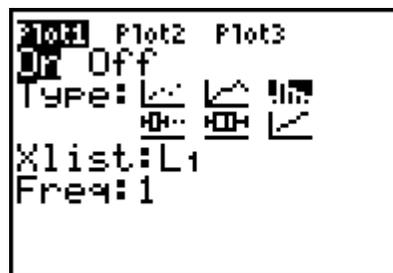


L1	L2	L3	Z
6	██████	-----	
5			
4			
4			
6			
L2(1)=			



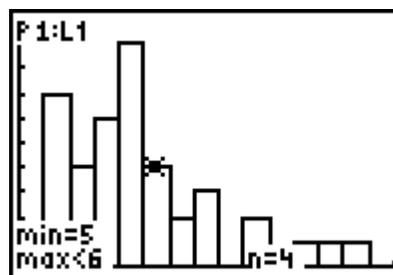
Step 3: In groups of four, students are to compile all of the results by adding others data to their data in L1. Students can use the data to create a histogram.

To create the histogram, students should set the **Plot1** options as shown and change the settings for the window as described in the worksheet.



Students may use the trace feature and the left and right arrows to navigate the graph and determine values. They are to determine the experimental probability of the first 5 appearing on the fourth roll based on the data simulated by the group.

Discuss the results of the different groups.



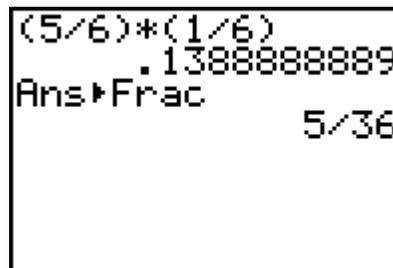
Ask students what conclusions they can draw from the graph. What can they determine about the number of rolls it takes for the first 5 to appear? They should be able to see that the graph is skewed left, meaning that it is more likely that the first 5 appears on the sixth roll or earlier and less likely that it appears on seventh roll or later. The graph drops off steeply to the right, meaning that there is a slim to zero chance of rolling the first 5 on the thirteenth or fourteenth roll. Allow this to lead into a discussion of the theoretical probability.

Problem 3 – Investigation

The goal of this problem is to derive the formula for $P(X=n)$, where n is the number of trials it takes for the first success.

On the Home screen, students are to calculate the probabilities of the first 6 appearing on the 1st, 2nd, 3rd, 4th, and 10th roll. Students should see a pattern leading them to the formula for calculating the probability of the first 6 appearing on the n th roll.

Students can convert their decimal answers to fraction form by pressing **MATH** **ENTER** **ENTER**.

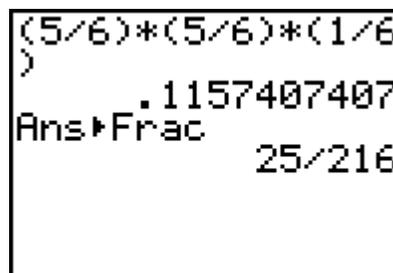


$$P(x = 1) = \frac{1}{6}$$

$$P(x = 2) = \frac{5}{6} \cdot \frac{1}{6} \text{ (Failure, then success)}$$

$$P(x = 3) = \left(\frac{5}{6}\right)^2 \frac{1}{6} \text{ (2 failures, then success)}$$

$$P(x = n) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} \text{ (n - 1 failures, then success)}$$



Lists can be used to determine the probability of the first 5 appearing on rolls 1 through 100.

Students will need to clear any data in L1, L2, and L3.

Step 4: L1 is the number of trials, n . Students are to enter the command **seq(x,x,1,100)** in L1 to generate a list of the integers from 1 to 100.

Step 5: L2 is the probability that the first success will occur in n trials. Students are to enter the command **geometpdf(1/6, L1)** in L2 to generate the probabilities.

Step 6: The first cell in L3 is the sum of the probabilities in L2. Students are to enter the command **sum(L2)** cell L3(1).

Discuss with students why the sum of the probabilities in list L2 is equal to 1.

Discussion Questions:

- How can you verify that this is a probability distribution?
- Theoretically, is the distribution finite or infinite?
- Practically, how long does it take before the $P(X=n) \approx 0$?
- How do your calculations above compare with the values in the list?

Step 7: The students may create a scatter plot of the data stored in L1 and L2 by changing the settings of **Plot1** as shown on the worksheet. They can then press **ZOOM** and choose **ZoomStat** to view the graph.

Discuss with students that this is a graphical representation of the geometric distribution. This graph should reaffirm to students that the first 5 will most likely appear on one of the first several rolls.

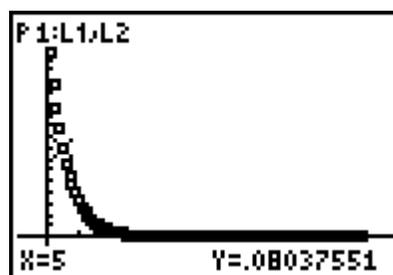
Discussion Questions:

- What is the shape of the graph?
- What type of function is represented?
- What conclusions can be drawn about the first 5 appearing as the number of rolls increases?

L1	L2	L3	1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	-----	-----	
L1 =seq(X, X, 1, 100)			

L1	L2	L3	2
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	.16667 .13889 .11574 .09645 .08038 .06698 .05582	-----	
L2 =geometpdf(1/6			

L1	L2	L3	3
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	.16667 .13889 .11574 .09645 .08038 .06698 .05582	1	
L3(1)=sum(L2)			



Step 8: Students will now find an exponential regression for the data in the scatter plot.

(**Note:** The syntax for the exponential regression formula is **ExpReg L1,L2**).

If the student elects to, he or she may graph the regression line on top of the scatter plot. A fast way to do this is to press **STO►** and type **Y1** after calculating the regression.

```
ExpReg
y=a*b^x
a=.2
b=.8333333333
r^2=1
r=-1
```

How does the regression equation relate to theoretical probability?

Through some algebra steps, students should see that the regression equation is the same as the formula for the first 6 appearing on the *n*th roll.

$$\left(\frac{5}{6}\right)^{n-1} \cdot \frac{1}{6} = \left(\frac{5}{6}\right)^{-1} \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^n = \frac{6}{5} \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^n = \frac{1}{5} \cdot \left(\frac{5}{6}\right)^n = 0.2(8.33333333^n)$$

Discussion Questions:

- Why is this distribution called a geometric distribution? (Show the ties to a geometric sequence, which can be used to **prove** that the sum of a geometric distribution is 1.)
- How can this formula be made into a general formula for geometric distributions?

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

Problem 4 – Expected Value

Students will use the lists they created in Problem 3 to calculate the expected value (also known as the mean) of this probability distribution.

Discuss with students the definition and meaning of the expected value of a random distribution.

If one repeated the experiment of counting the number of rolls it takes to roll a 5 many times, the expected value is the number of rolls one could expect to occur most often.

Step 1: Students should multiply payoff × probability with the formula **L1 * L2**.

L1	L2	L3
1	.16667	.16667
2	.13889	.27778
3	.11574	.34722
4	.09645	.3858
5	.08038	.40188
6	.06698	.40188
7	.05582	.39071

L3 = L1 * L2

Step 2: In first cell of L4, they will find the sum of the products by entering **sum(L3)**.

Students should understand that the expected value of 6 represents the average number of rolls until the first 5 appears. Six is approximately the mean of the distribution.

L2	L3	L4
.16667	.16667	6
.13889	.27778	-----
.11574	.34722	
.09645	.3858	
.08038	.40188	
.06698	.40188	
.05582	.39071	

L4(1)=SUM(L3)

Discussion Questions:

- Is this answer reasonable? Why?
- How can the general formula for expected value be found? (This is can be done using a geometric series).

In general, the expected value of a geometric distribution is $\frac{1}{p}$.

Extension – More or less

Students are asked two questions that involve adding geometric probabilities.

What is the probability that it will take less than 4 rolls to obtain a 5?

Solution: $P(X < 4) = P(X = 1) + P(X = 2) + P(X = 3)$

This can also be found by using the **geometcdf** command. Students can choose **geometcdf(** and enter the probability of success (in this case, 1/6) and the upper limit of rolls (in this case, 3).

```
geometcdf(1/6,3)
.4212962963
```

What is the probability that it will take more than 4 rolls to obtain a six?

Solution:

$P(X > 4) = 1 - P(X \leq 4) = 1 - \text{geometcdf}(1/6,4)$

```
1-geometcdf(1/6,4)
.4822530864
```