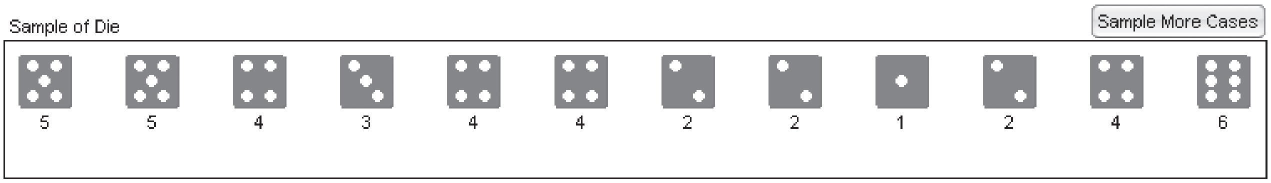
Geometric Distributions—Waiting-Time Problems

You will need
- one die
- **RollingDice.ftm**

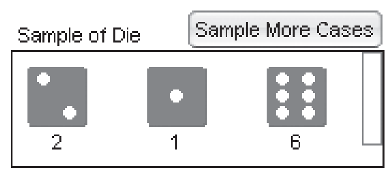
In the general population, about 40% of people have type A blood. Suppose that a worker in a blood bank needs type A blood today and wants to know about how many blood donations he might have to process before he finds the first donation that is type A. Or suppose you want to know how many times, on the average, you have to roll a die to get a six. Is the sample below unusual?



Both of these situations are called *waiting-time problems* because the variable in question is the number of trials you have to wait before the event of interest happens. The number of trials isn't fixed—you simply count the number of trials until you get the first success. In this activity you'll design simulations for these waiting-time situations and use them to construct approximate probability distributions.

COLLECT DATA

1. Roll a die. Is it a six? In other words, was your waiting time to success (a six) just one trial?
2. If you were successful on your first roll, then stop this run of the simulation. If you did not roll a six on the first trial, then continue rolling your die until you get your first six. Record the value of *x*, the number of the trial (roll) on which the first six occurred. For example, in this set of rolls, it took 3 trials (rolls) to get a six, so you would record a 3 and then start all over again.



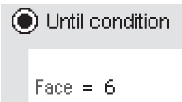
3. Repeat steps 1 and 2 at least ten times, recording the value of *x* each time.
4. Combine your values of *x* with others from the class and construct a plot to represent the simulated distribution of *X*. Describe the shape of this distribution and find its mean.

INVESTIGATE

Rolling Dice

Now you'll design a simulation in Fathom.

5. Open the document **RollingDice.ftm.** It has two collections: Die and Sample of Die. Make sure you can see the case table or the collection for Sample of Die.



6. Open the sample collection's inspector. On the **Sample** panel, you want to set up the simulation to roll the die until it shows a six. Click the **Until condition** button. Enter the expression at left. This expression tells Fathom when to stop collecting samples.

Q1 Click **Sample More Cases** repeatedly. Describe what happens to the case table. What was the greatest number of rolls required to get a six? About how many times did you have to roll to get a six?

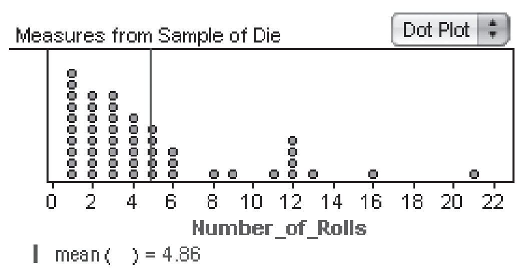
Now you will collect the numbers of rolls that it takes to get a six. Counting the number of rolls will be the same as counting the number of cases in the Sample of Die collection.

7. On the **Measures** panel in the inspector, define a new measure, *Number_of_Rolls*, with the formula count().

Select the Sample of Die collection and choose **Collection | Collect Measures.**

8. Collect 5 measures from the sample collection.
9. Make a dot plot of *Number_of_Rolls* and plot the mean.
10. Show the measures collection's inspector and go to the **Collect Measures** panel. Change the number of measures to 45. Click **Collect More Measures.** The simulation runs 45 times, collecting the numbers of rolls it took to get a six and recording them in the measures collection.

Q2 Describe your distribution of *Number_of_Rolls* in terms of shape, center, and spread. Which number of rolls is the most likely? What is the maximum number of rolls your simulation took to get a six?



If you decide to collect a large number of measures, it will speed things up to uncheck Animation on. Also, you might want to change your dot plot to a histogram.

Q3 Collect measures again, watching the graph. Continue to collect measures to answer the original question: What is the mean number of rolls you need to get a six?

Q4 Based on the simulation, what do you think is the *most likely* number of rolls you need to get a six (that is, the number that happens most frequently)? Explain. Sketch or refer to a graph if you need to.

Waiting for Type A Blood

Now you'll set up the situation about waiting for type A blood. This time, the basic idea is to use a collection with one randomly generated donation, then sample from the collection until you get a donation of type A blood.

You can either scroll down to an empty space in the document or create a new document by choosing **File | New.**

11. Create a collection named Donor with one case. Define an attribute named *BloodType* using the formula here.

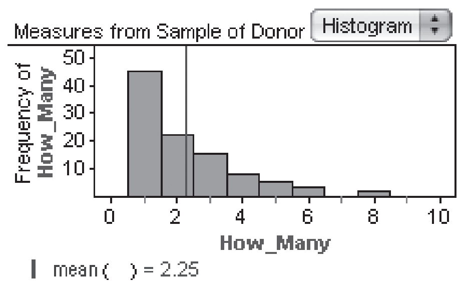
BloodType	Not A	if (random ( ) < 0.4) {"A" "Not A"

You can also rerandomize by pressing Ctrl+Y (Win) ⌘+Y (Mac).

12. Repeatedly choose **Collection | Rerandomize.** Observe the effect on *BloodType.* How many times do you have to rerandomize to get a type A donation?
13. Select the Donor collection and choose **Collection | Sample Cases.** Show the sample collection's inspector. On the **Sample** panel, use what you learned in the first part of this activity to write the formula that tells Fathom to sample cases *until* the first "A."

To collect measures, select the sample collection and choose **Collection | Collect Measures.**

14. As you did in step 7, define a new measure, named *How_Many,* that counts the number of samples until there is a donation of type A blood. Collect 5 measures.
15. Make a dot plot or histogram of *How_Many* and plot the mean.
16. Use the measures collection's inspector to collect at least 100 measures.



Q5 Describe the shape, center, and spread of the distribution of measures.

Q6 Estimate the probability that the first donation with type A blood is the second one checked.

Q7 What is your estimate of the probability that at most five donations are checked to get one that is type A? At least two donations? At most four donations?

Q8 Which donation is the most likely to be the first one that is type A?

EXPLORE MORE

1. Repeat the simulation for the 7% of donors having type O− blood. What is the new mean? Repeat the simulation with other probabilities. Make a conjecture about the relationship between the probability and the mean of the distribution.
2. Fathom has a function for generating random numbers from a geometric distribution. Investigate the distribution of such numbers and see if there is a connection to the blood-donation simulation.
3. On the average, how many rolls does it take to get a five *and* a six? A five *or* a six? Make predictions, then simulate both situations.
4. On the average, how many rolls does it take to get all six values: 1, 2, 3, 4, 5, and 6?

Objectives

- Understanding the nature of the geometric distribution and distinguishing the geometric distribution from other distributions (particularly being skewed right)
- Understanding that even though the results (measures) vary from one iteration to the next when a random process is repeated over and over, knowing the distribution of the measures enables you to decide whether any *single* observation is probable or improbable
- Using simulation to calculate empirical probabilities
- Using the geometric probability model as a theoretical model for a real-life situation

Activity Time: 35–50 minutes

Setting: Paired/Individual Activity (collect data, use **RollingDice.ftm**, then build simulation or use **TypeADonor.ftm**)

Optional Document: RollingDiceExp.ftm (Explore More 3 and 4 solutions)

Materials

- One six-sided die for each student

Statistics Prerequisites

- Definition of probability
- Familiarity with the definition of independent events
- Describing distributions in terms of shape, center, and spread
- Familiarity with sampling

Statistics Skills

- Geometric distribution and its probabilities
- Setting up a simulation of a geometric situation
- Doing a simulation to compare experimental results to a theoretical model to see if the results are reasonable
- Calculating probabilities using the plot of a simulation of the geometric situation
- Setting up and doing a simulation using randomization
- Sampling and collecting measures

AP Course Topic Outline: Part II A (4), B; Part III A, B (1), D (6)

Fathom Prerequisites: Students should be able to make a new collection, define attributes and measures, make histograms, and collect measures.

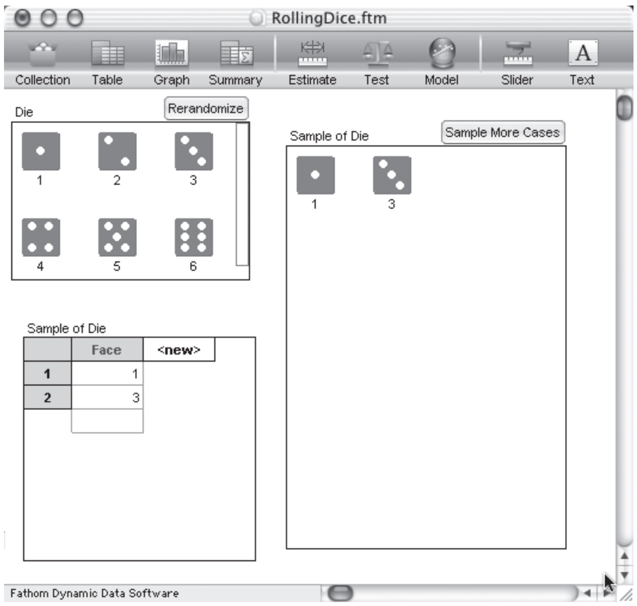
Fathom Skills: Students set up a geometric situation to sample from, compare results using plots and a distribution function, sample until a criterion is met, and use an "if" statement to generate a population. *Optional:* Students work with the geometric distribution function (Explore More 2).

General Notes: This activity is essential so that students can understand the nature of the geometric distribution. It helps students understand what distinguishes a geometric distribution from other distributions. Using Fathom, students can easily simulate waiting-time problems and build the sampling distributions.

Procedure: The first part of this activity has students roll a die until they roll a six, recording the number of trials it took to roll a six. They then repeat this process ten times. You can have your class plot the distribution on the board or on an overhead. Usually, this part of the activity goes quickly and they then have a better understanding of what they do with Fathom.

The Rolling Dice section uses Fathom to explore how many trials it will take to roll a six on a six-sided die. If the probability of an event is p, the expected number of trials you need to get a success is $1/p$. In this case the probability of rolling a six is 1/6, so the expected number of trials to roll a six is 6.

When the students open the document **RollingDice.ftm,** their screen will look like this.



The collection Die is the population of possible rolls when rolling a six-sided die. The sample collection at this point is set up to roll a die twice. The sample collection window is large so that when the students do sample until they get a six, they can see most of the rolls in their samples if the samples end up being large.

In the next part of the activity, students use Fathom to simulate how many blood donations a technician might have to process before she or he finds the first donation that is type A, given that in the general population, about 40% of people have type A blood (Source: www.redcross.org/services/biomed/blood/supply/usagefacts.html).

You can change this to any waiting-time simulation because Fathom makes it easy to change the probability of an event, thereby allowing students to explore the effects of different probabilities on the distribution.

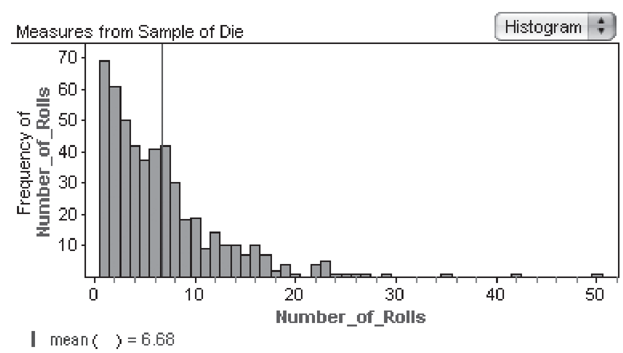
If you do not have time to do this part of the activity in full, your students can use the document **TypeADonor.ftm** instead. They can then resume the activity at Q5.

INVESTIGATE

Q1 As the sample is collected, the number of cases in the case table increases or decreases to accommodate the number of rolls. The last number in the case table is always 6. The greatest number of rolls required will vary; some students have gotten values as large as 36.

Q2 This situation gives rise to a skewed distribution with the skew toward the larger values. With this kind of distribution, mean > median > mode. The first bin will probably be the tallest, although with only 50 measures any of the bins from 1 to 6 could be the tallest. The bins probably (but not necessarily) decrease in size moving to the right (it is a geometric distribution). Again, the maximum will vary. The sample dot plot in the student activity has mean 4.86 and standard deviation 4.45. The median is 3, with an *IQR* of 4 ($6 - 2$). The maximum is 21.

Q3 The mean is 6, although depending on the number of measures students collect, their answers will vary. Here is a histogram of 500 measures with mean 6.68.



Q4 The most likely number of rolls to get a six is the mode, which is 1.

Q5 The histogram in the activity shows a sample distribution of 100 measures. The mean of the 100 waiting times is 2.25, slightly smaller than the expected waiting time of 2.5. The estimated mean typically will be too small because a small simulation usually won't produce any of the large possible values for X. The distribution is skewed right toward the larger values, and the standard deviation is about 1.6. (For $p = 0.4$, the standard deviation is about 1.94.) The median is 2 and the *IQR* is $3 - 1$, or 2.

Note: If your students would like their captions to show "A" or "Not A" as in **TypeADonor.ftm,** show the inspector for the Donor collection and go to the **Display** panel. Double-click in the formula cell for the attribute *caption* and enter BloodType.

Q6–Q7 Simulated answers will vary. Theoretical answers: $P(X = 2) = 0.24$, $P(X \leq 5) = 0.92224$, $P(X \geq 2) = 1 - P(X = 1) = 0.6$, and $P(X \leq 4) = 0.8704$. For the above histogram, these would be 0.22, 0.95, 0.55, and 0.9, respectively. You might need to remind students that the easy way to find these is to select the bars in the histogram that they want to include (hold down the Shift key to select more than one). Move the cursor over the measures collection to see the total in the lower-left corner of the document window.

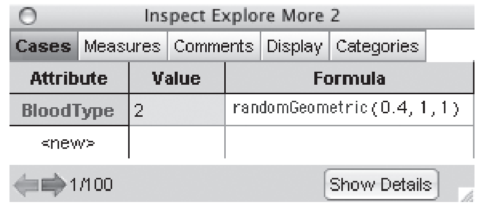
Q8 The first checked donation is the donation most likely to be the first one that is type A. That probability is 0.4. Each subsequent checked donation has the same chance of being type A itself, but a number of failures have to occur before it even has the chance of being the first one that is type A. You can use the discussion questions to help students make sense of this.

DISCUSSION QUESTIONS

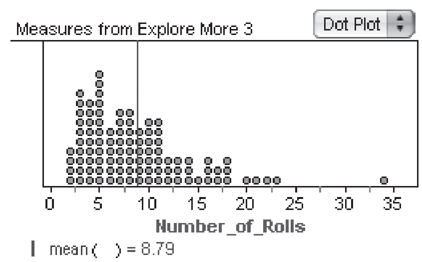
- Explain how the "if" statement chooses a blood type for a single donation.
- What was the most frequent number of donations to wait for? Explain why that makes sense.
- Generally, the frequency of each higher number of donations to wait for is less than the previous number. Why does this make sense?
- If the probability that a person has type A blood is much less than 0.4, what do you think would happen to the distribution of measures? Why? (Use Explore More 1 to test your conjecture.)

EXPLORE MORE

1. Students may conjecture that the mean (expected value) of the distribution is 1 divided by the probability of the event, or $1/p$.
2. Students should use Fathom Help to learn about the randomGeometric function. Using this random number generator to define an attribute for a collection of many cases enables students to redo the activity without sampling or collecting measures. Encourage students to change the probability value, possibly by using a slider.



3. To sample until you get a five *and* a six, use the formula (count(Face=6)>0) and (count(Face=5)>0), then collect measures. For 100 such measures, the mean number of trials to get a five and a six was 8.79, as shown below. The average number of trials to get a five or a six was 2.4 (the theoretical answer is 1/(1/3), or 3). See the document **RollingDiceExp.ftm** for these setups.



4. In one experiment, the average number of trials to get all six values was 14.81. Students can use the until condition uniqueValues(Face)=6. See the document **RollingDiceExp.ftm.**