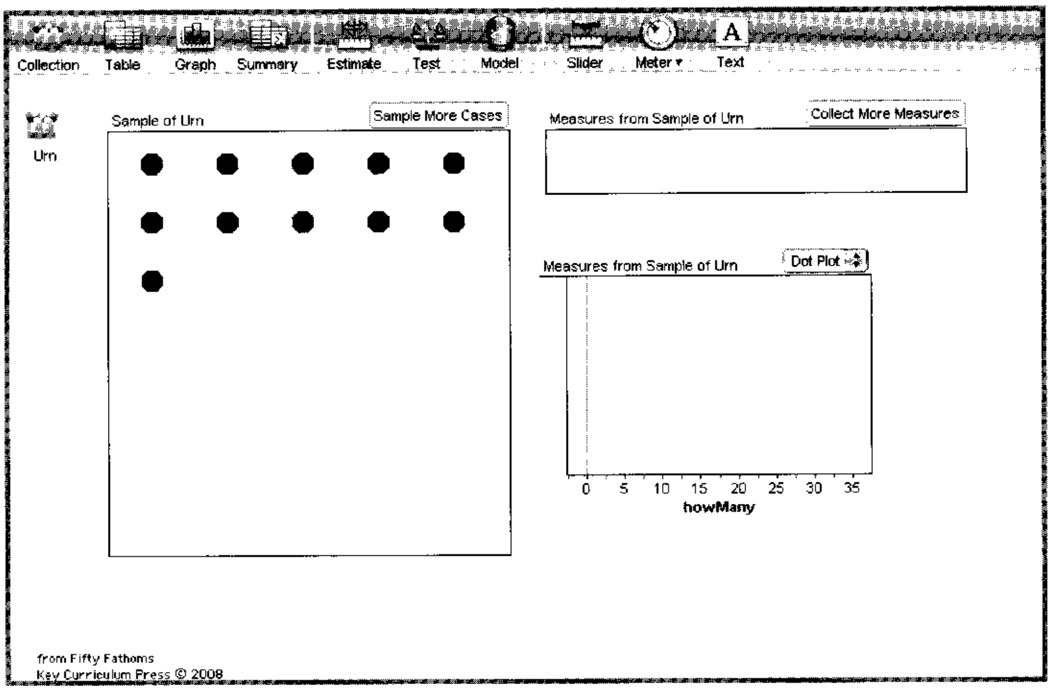
Demo 47: Wait Time and the Geometric Distribution

The distribution of times until something happens • How this is the geometric distribution

Urns are used for three things in our society: holding tea (rarely), holding the ashes of the deceased, and holding colored balls in probability problems from mathematics textbooks. In this demo, we will draw balls from an urn until we get a red one. Every time we draw out a ball, we will put it back before drawing the next one.

Here we are interested in the distribution of *wait times*—how long it takes (measured in numbers of balls) until we get the first red. We'll find it follows what's called the *geometric distribution,* so called because the counts at each successive draw follow a geometric sequence. It's also interesting to find the mean of this distribution, as that is the expected number of draws (in the "expected value" sense of that word).



What To Do

- ▷ Open **Wait Time.ftm.** It should look like the illustration.

Here you can see the **Urn**—the collection in the upper left. You can open it by dragging its lower-right corner. You'll see that it contains ten balls, only one of which is red. But then close it to save space.

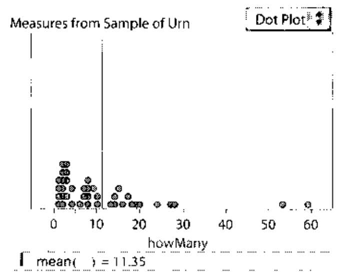
The next collection is the **Sample of Urn** collection; it contains the records of balls you draw—with replacement—from **Urn**. You draw until you get a red one. The other two objects we will get to shortly.

- ▷ Press the **Sample More Cases** button in the **Sample of Urn** collection. Fathom will draw new balls from **Urn** until you get a red one. Note how many draws it takes.
- ▷ Do this a few times, until you're sure you understand what's going on and you know what kinds of numbers of draws you see until there's a red ball.
- ▷ Let's get Fathom to record these automatically. Press the **Collect More Measures** button in the collection (empty box) in the upper right. Fathom goes to work, and you see measures (the number of balls, called **howMany**) appear in the collection and in the graph below.

▷ Every time you press that button, you get five more measures. Keep doing this until you start to get an idea of the shape of the distribution.

▷ Let's figure out the mean. Click once on the graph to select it, then choose **Plot Value** from the **Graph** menu. The formula editor opens.

▷ Enter **mean( )**, then press **OK** to close the editor. The mean will appear, and the graph will look something like the illustration.



By this time, you may be getting impatient with the flying gold balls. Let's get rid of them to speed things up:

▷ *Speed step 1:* Double-click the sample collection (the big one with the colored circles) to open its inspector.

▷ *Speed step 2:* Uncheck the **Animation** box in the **Sample** panel.

▷ *Speed step 3:* Double-click the measures collection (the one with green balls labeled **n = 3** and similarly) to open its inspector.

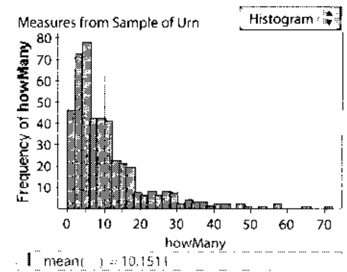
▷ *Speed step 4:* Uncheck the **Animation** box in the **Collect Measures** panel.

Now that we can go faster, let's add a lot of measures to that graph:

▷ In that **Collect Measures** panel, change the number of measures collected from 5 to 50. Then close the inspector.

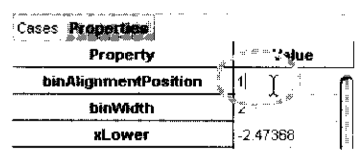
▷ Press the **Collect More Measures** button to add 50 cases to the dot plot. (You can make the measures collection go even faster if you make the **Sample of Urn** collection so small that it turns into an icon. That way Fathom doesn't have to redraw it.)

▷ Keep doing this until the mean seems to converge to a sensible value (the theoretical value is 10.0). If you get too many points in the dot plot, change to a histogram as shown in the illustration.



Note: If you make a histogram, it will probably have a bin width of 2, like the one in the illustration. This is a problem only in the first bin, which contains both 0 and 1. Since 0 is impossible (you can't draw a red ball on the zeroth try), that first bin has many fewer cases than the neighboring bin. This misrepresents the ever-decreasing distribution.

If this bothers you, force the binning to start at 1. Select the graph. From the **Object** menu, choose **Inspect Graph.** In the **Properties** panel, change the **binAlignmentPosition** (as in the illustration) to 1 and press **Tab** or **Enter.**



Questions

1 Did the mean converge to 10.0—or at least close to it?

2 How would you describe the shape of the distribution?

3 Why would each number be less likely than the one before it?

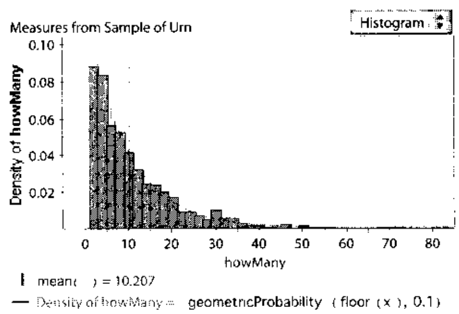
Extensions

Let's plot the theoretical distribution.

- ▹ Make sure your graph is a histogram. Then, with the graph selected, choose **Scale | Density** from the **Graph** menu. The vertical scale will change.
- ▹ Choose **Plot Function** from the **Graph** menu. The formula editor opens.
- ▹ Enter

  geometricProbability(floor(x), 0.10)

  (The 0.10 is the probability of success.)
- ▹ Press **OK** to leave the formula editor. The function appears as in the illustration.



Of course there are other interesting things to do. One of the best is to change the probability of getting a red ball. Do that by altering the original **Urn** collection. For example, to get rid of a black ball, select it in **Urn** and then choose **Delete Case** from the **Edit** menu.

If you want a new red ball, make a new case and give its attribute **tag** a value of **1**.

Then, before you **Collect Measures** to see the distribution of wait times, delete all of the cases in the **Measures from Sample of Urn** collection (they're from a different population). To do that, select them (select the graph and choose **Select All Cases** from the **Edit** menu); then **Delete Cases** in the **Edit** menu.

The theoretical probability function will probably not match any more—so you'll have to change it, altering the 0.10 probability to the new value.

Challenges

4 Make a conjecture about the mean of any geometric distribution.

5 Prove that conjecture. (It helps to know the sum of an infinite geometric series, but you can do without this if you know a trick.)

6 Is the *median* of a geometric distribution larger than, smaller than, or the same as the mean? Use Fathom to demonstrate that what you say is correct. Also, explain why it is correct based on the shape of the distribution. **Sol**

7 What is the *mode* of a geometric distribution? Explain. **Sol**

Theory Corner

We derive the theoretical expected value (that is, the mean) for the number of trials in "The Geometric Distribution: Proof That the Mean Is $(1/p)$" in Appendix B.