Demo 12: How Random Walks Go as Root *N*

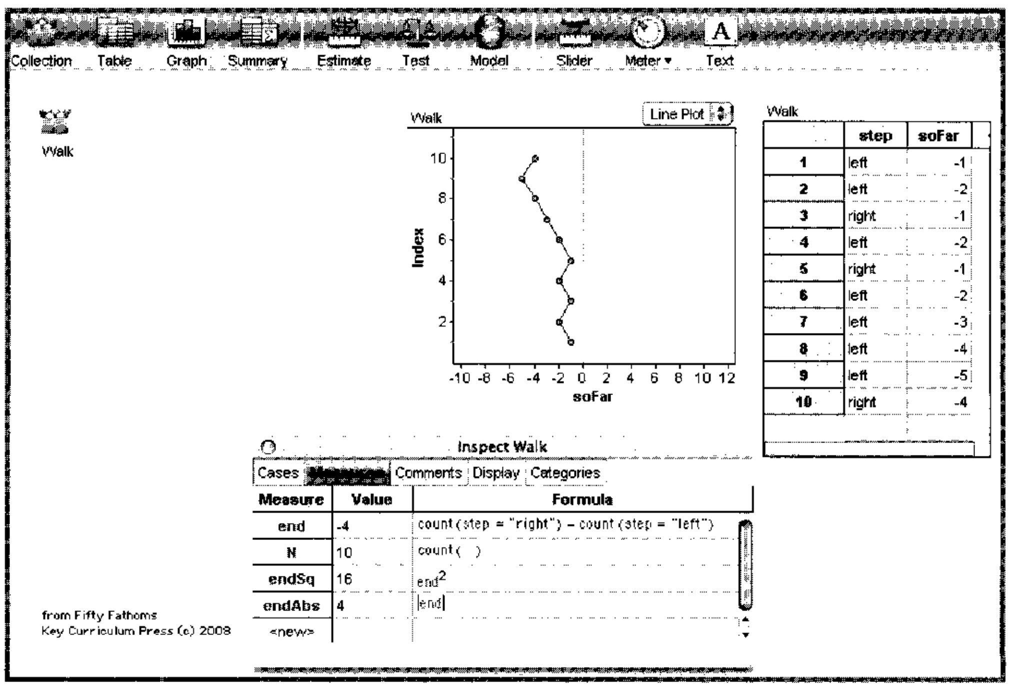
How the distance from the origin increases with the number of steps

The previous demonstration showed how the proportion of heads you flip approaches the "true" probability of heads. This demo shows how the number of heads gets farther and farther from the expected value. Learning why these two apparently contradictory statements are not in fact contradictory is one of the most important things you can do to understand statistics.

Instead of staying with flipping coins, we'll use a slightly different context—a random walk. The idea is that, for each step, you randomly choose whether to go east (to the right, positive) or west (to the left, negative). After *n* steps, how far have you gone?

On the average, you will not go anywhere. Your mean position will be where you started. But it is also true that after many steps, you're very unlikely to wind up *exactly* where you began. So how far—in absolute terms—are you likely to be from the origin after *n* steps? We will simulate this situation in this demo to find out. And the answer is that *the distance is proportional to the square root of n.*

Note: In this demo, we will turn the graphs on their sides so that the "time" axis will go up instead of to the right. This will fit with our left/right walk metaphor; it will also map well onto the graphs of the binomial distribution that we will create in Demo 13, "Building the Binomial Distribution."



What To Do

▹ Open **Random Walk Root n.ftm.** It will look something like the illustration.

Here, the **Walk** collection contains one random walk, shown in the graph. It starts at the bottom; it happens that after 10 steps, we've landed on 0. First, we'll see how the walks vary.

▹ The inspector should be open as shown; if not, double-click the **Walk** collection to open its inspector.

▹ Choose **Rerandomize** from the **Collection** menu.

⇨ The shortcut for **Rerandomize** is **⌘+Y** on the Mac or **Control+Y** in Windows.

▷ Rerandomize repeatedly to see how the walk changes. You can also see, in the **Measures** panel of the inspector (shown), how several quantities change: the **end** position of the walk, the square of that (**endSq**), and its absolute value (**endAbs**).

⇨ Note that you can also see their formulas.

Wouldn't it be nice to record all those numbers somewhere? That's a job for a measures collection—which has really been there all along, hidden.

▷ Close the inspector to make more screen space.

▷ Choose **Show Hidden Objects** from the **Object** menu. Now the screen should look something like the illustration below.

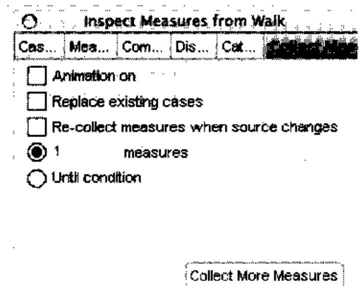
Three new objects appeared: a measures collection (**Measures from Walk**), a graph, and a summary table. The measures collection already holds the results from 10 walks.

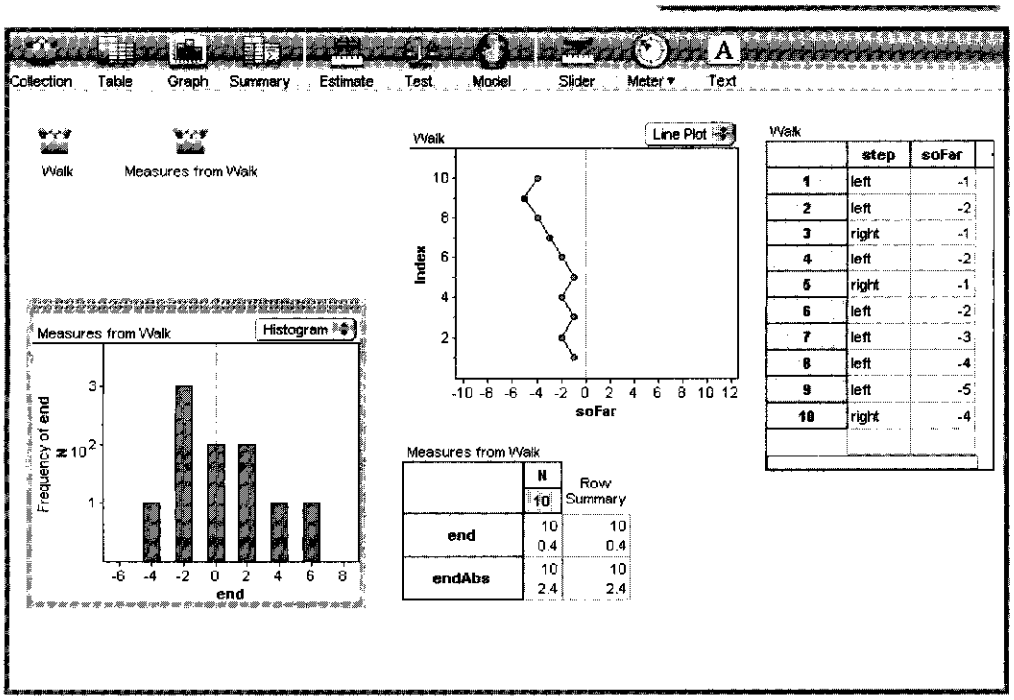
See in the graph that the ending positions (**end**) among the 10 walks ranged from −4 to +6. See in the summary table that the average position at the end is 0.4, but the average *absolute* position at the end is 2.4.

▷ Click the measures collection (it looks like the picture below) once to select it. Choose **Collect More Measures** from the **Collection** menu. Fathom will produce a new walk and add its results to the measures collection. The graph and the summary table will update to show the results from 11 walks instead of 10.

▷ Keep collecting more measures this way until you have at least 20 walks. Clearly we need a faster way to get these walks done.

▷ Double-click the measures collection to open its inspector. It will open to the **Collect Measures** panel, as shown. Edit the number so that you will collect 80 measures instead of one. (It doesn't have to be exactly 80—just a lot.)





Note: Depending on your screen, it may not be obvious where to put this. If you're lucky, it will appear directly over the graph of the single walk—top center. If it's not, move it there. We don't need that graph now.

▹ Click the **Collect More Measures** button. Fathom will update everything. Keep collecting more measures until you have at least 500 measures from walks of 10 steps each.

Having exactly 500 is not necessary, but it's helpful because then the decimal values in the summary table will be *short*—you won't have to scroll to see the data.

▹ Now we need a different number of steps. Click once on the **Walk** collection (the one in the upper left) to select it.

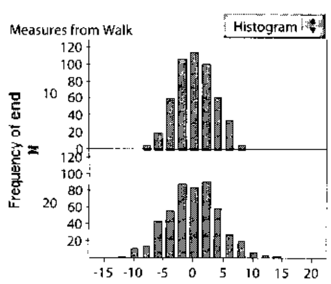
▹ Choose **New Cases** from the **Collection** menu. Give the collection 10 more cases (for a total of 20 steps).

▹ Set up the measures inspector to collect 500 measures, then click the **Collect More Measures** button. Again, Fathom updates, but this time you'll see data for 20 steps (**N = 20**) separate from the data for 10 steps. The summary table will look something like this illustration:

Measures from Walk

	N: 10	N: 20	Row Summary
end	500 0.176	500 -0.064	1000 0.056
endAbs	500 2.56	500 3.416	1000 2.988

The graph looks like this:



▹ Keep adding steps and collecting measures so that you have data for 10, 20, 40, 80, and 160 steps. Stretch the summary table as it grows, making it wider so that you can see the data. You may need to scroll.

Note: This is a little tricky. You're alternating between adding cases to the **Walk** collection (to make a walk of more steps) and adding cases to the **Measures from Walk** collection (to get the results of 500 walks). It's easy to make a mistake, but also easy to recover if you keep your eyes open. If you get too many cases in the **Walk** collection, that is, too many steps, select the excess in the graph or the case table and choose **Delete Cases** from the **Edit** menu. Don't worry about getting too many measures, but if you get some cases for an **N** you don't want, select them in the lower-left graph and, again, choose **Delete Cases** from the **Edit** menu.

▹ If you like, drag **endAbs** to the horizontal axis of the graph, replacing **end** to see what its distributions look like.

Questions

1 How does the average position at the end of the walk (called **end** in the summary table) change as you increase the number of steps?

2 How does the average absolute value at the end of the walk (**endAbs**) change? **Sol**

3 Why are these two quantities different?

4 How does the shape of the graph change as you increase the number of steps?

5 Judging from the summary table, is the mean absolute distance proportional to the number of steps? How do you know?

Onward!

Now we want to graph the numbers in the summary table. We have set this up for you and generally cleaned up the screen. We begin with a new file, based on the previous one:

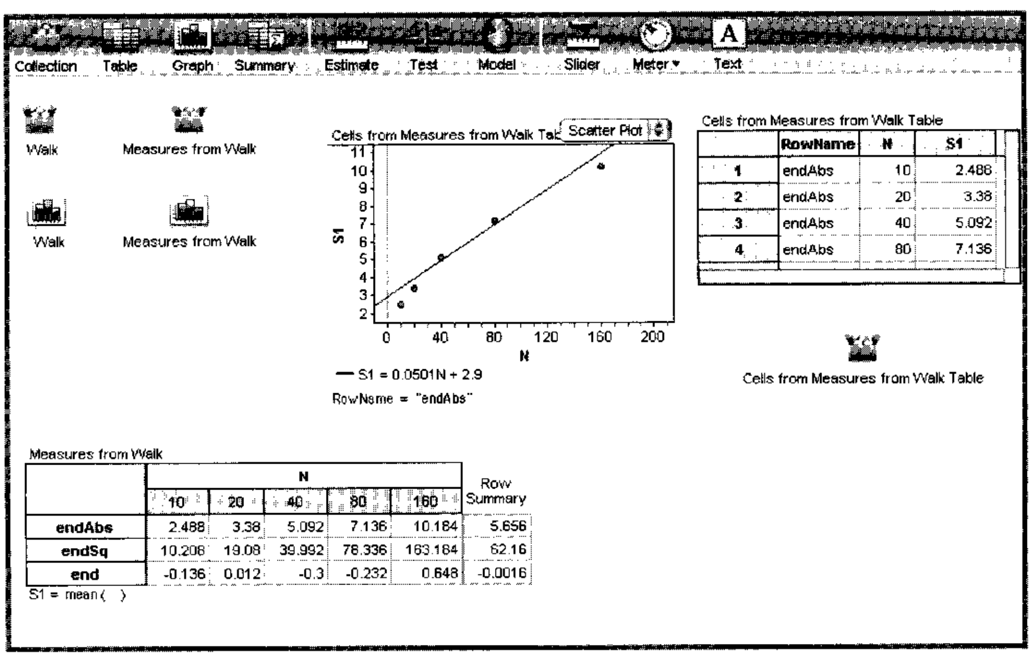
▹ Open **Random Walk Root n part 2.ftm.** It should look like the illustration.

The old graphs are collapsed into icons, and the summary table, at the bottom, is streamlined (we have removed the **count( )** formula—there are 500 cases in each column). And we have added the **Cells from Measures from Walk Table** collection, which is actually a measures collection from the table itself. The graph in the middle shows how the mean absolute distance changes as a function of the number of steps.

Note: We control what to plot with the *filter* (**RowName = "endAbs"**) at the bottom of the graph. The actual attribute is **S1** from the table at the bottom, which is the mean. But the filter tells the graph that you want to see only the absolute values. We can edit the filter by double-clicking it, for example, to **RowName = "endSq"**.

This whole setup is complex and layered. Be sure to understand that the three collections are linked left to right:

- **Walk** is a collection of steps—at this time, 160 of them, which is where we left off. Each case is one step, either left or right. To see the last walk, open the graph below it by dragging the lower-right corner. (Then shrink it again.)
- **Measures from Walk** is a collection of data from 2500 walks—500 at each of five different lengths. Each case is a walk, and its attributes are the number of steps, the position at the end, and that position's absolute value and square. Open the graph below it (drag the lower-right corner) to see the distribution of final positions.



Cells from Measures from Walk Table is a summary of the 2500 walks. It shows the mean of the absolute value (and square, and original **end** number) of the number of steps in the 500 walks at each of the five lengths.

▷ Optional: Add 40 cases to the **Walk** collection (for a total of 200) and **Collect More Measures.** The table and graph will update. You'll need to stretch the table.

Note: To add cases to the **Walk** collection, select it by clicking on it once, then choose **New Cases** from the **Collection** menu.

To collect more measures, first you need to select the **Measures from Walk** collection.

▷ Drag parts of the brown movable line in the graph to see if you can make it fit the data points there. It should be clear looking at the line that the pattern of the points is genuinely curved, which means that the mean absolute value of the distance is *not* proportional to the number of steps.

▷ Choose **Show Hidden Objects** from the **Object** menu. A new graph appears, showing the mean *square* distance as a function of the number of steps.

▷ Drag the line to fit the data. It does a lot better.

We have shown that, in a series of random walks, the number of steps you take is proportional to the mean of the square of the distance from where you started. In fact, the slope of the line is 1 and the intercept is 0, that is,

$$\textit{mean square distance} = N$$

Challenges

6 Edit the filter on one of the graphs to show how the mean of **end** does not change much as you change **N.**

7 Edit the formula in the table (or add a formula) to see how the *medians* (instead of the means) of these quantities change. Would you rather use median here? Why or why not?

8 Mean absolute deviation (which is what we were looking at, the mean of **endAbs**) ought to be a measure of spread, as is standard deviation. Figure out how to compare standard deviation to the mean absolute deviation. Are they proportional? If so, what is the constant of proportionality? Explain these results.

9 In what respects is this random-walk situation identical to (*isomorphic* to) the coin flipping we did in Demo 11, "Flipping Coins—the Law of Large Numbers"?

10 Explain clearly why, if these two demos are isomorphic, the spread decreased with n when we flipped coins, but increased with n when we did a random walk.

11 If the mean square distance is N, does that mean that the mean *absolute* distance is the square root of N? Exactly? Not at all? Only in the limit? (One way to look at it: In the case table, create a new attribute equal to the square root of **N;** then replace **N** with this new attribute in the first (**N-endAbs**) graph. The line is straight; what is its slope? How do you explain that it is not 1.0?) **Sol**

12 Suppose you took a long random walk and plotted the number of steps against the square of the distance. Would the graph approach a straight line? Explain.

Theory Corner

See "A Random Walk: Two Proofs That the Mean Square Distance Is N" in Appendix B.