

Probability, the Law of Large Numbers, and Simulation

by Randy Gallaher

Elementary Statistics Module

The law of large numbers states that as the number of trials gets larger and larger, the relative frequencies with which certain outcomes occur approach the actual (or theoretical) probability of those outcomes. In this module, we will use some simple simulations to explore the law of large numbers.

Simulating Coin Tosses

If we flip a coin one time, what is the probability that it will land on heads? _____

What is the probability the coin will land on tails? _____

Because there are two possible outcomes and because each outcome has the same likelihood of occurring, the answer to both of these questions is of course $1/2 = 0.5$. Then, the law of large numbers says that if we were to flip a coin 1000 times, then about 500 of the 1000 flips should land on heads and about 500 of the 1000 flips should land on tails. We will use *Microsoft Excel* to simulate 1000 flips of a coin to demonstrate the law of large numbers. We will use *Excel* to randomly select values of 0 or 1; 0 will represent heads and 1 will represent tails. Because 0 and 1 will be equally likely to occur (as are heads and tails) we will have a simulation of the real thing.

Step 1: Simulate the flips. That is, generate the 0s and 1s.

Excel has a random number function, `RAND()`, that will generate random numbers from 0 to 1, including 0 but not including 1. We can generate random numbers from 0 to 2, including 0 but not including 2 by multiplying `RAND()` by 2. *Excel* also has an integer function, `INT()`, which will find the greatest integer value smaller than the argument value between the parentheses. For example, `INT(4.98)` will return a value of 4. Thus, the command `INT(2*RAND())` will return the integer values from 0 to 2, including 0 but not including 2. That is, `INT(2*RAND())` will return a random integer from 0 to 1, including both 0 and 1. In short, `INT(2*RAND())` will randomly return either 0 or 1. If we then count the number of 0s and 1s that are generated, we will have simulated the flips of the coin.

In cell A1, type in `=INT(2*Rand())`, and press **Enter**. Highlight cells A1 through A1000. Click **Edit**, and choose the **Fill** and **Down** options. When this is done, cells A1 through A1000 will be filled with 0s and 1s. Again, this represents our “heads” and “tails.” Thus, we now simply need to count the results.

Note: As further entries are made in this spreadsheet, the random values previously determined will be regenerated. That is, the values currently in cells A1 to A1000 will change as further entries are added to this spreadsheet.

Step 2: Tally the number of heads and tails (0s and 1s).

We will use *Excel's* COUNTIF() function to tally the results.

In cell C1, type in =COUNTIF(A1:A1000,0), and press **Enter**. The number of 0s that appear in cells A1 to A1000 now appears in cell C1.

In cell C2, type in =COUNTIF(A1:A1000,1), and press **Enter**. The number of 1s that appear in cells A1 to A1000 now appears in cell C2.

Step 3: Calculate the relative frequencies of heads and tails (0s and 1s).

In cell D1, type in =C1/1000. The relative frequency of 0s (heads) now appears in cell D1.

In cell D2, type in =C2/1000. The relative frequency of 1s (tails) now appears in cell D2.

Recall that when further entries are made in unused cells, the random numbers are regenerated. Click on an unused cell (such as E1). Press the **Delete** button. The random numbers are regenerated, and the tallies and relative frequencies are adjusted accordingly. Repeat this process as often as you like to see that the number of heads and tails usually remain near 500 each out of the 1000 flips and the relative frequencies remain near to 0.5.

Print page 1 of your spreadsheet to hand in: Click **File**, then **Print**. For the Print Range, click so that a bullet appears next to the “Page(s)” option. In the box next to “From:,” type in 1, and in the box next to “To:,” type in 1. Click the **OK** button and page 1 will print.

Repeat the procedure for a greater numbers of flips, say 5000 flips or 10,000 flips. Do you expect similar results?

Warning: *Excel* may respond more slowly as you increase to larger and larger number of trials. Also, be sure to use the new number of flips in the calculations. For example, if you are simulating 10,000 flips, in cell D1 (in Step 3), you would enter =C1/10000 instead of =C1/1000.

Simulating Rolls of a Single Die

If we roll a die one time, what is the probability that it will land on 1? _____
What is the probability the die will land on 2? _____ 3? _____ 4? _____
5? _____ 6? _____

Because there are six possible outcomes and because each of these six outcomes have the same likelihood of occurring, the answer to all of these questions is $1/6 \approx 0.167$. Then, the law of large numbers says that if we were to roll a die 600 times, then about 100 of the 600 rolls should land on 1, about 100 of the 600 rolls should land on 2, and so on. Let's use *Microsoft Excel* to simulate 600 flips of a coin. We will use *Excel* to randomly select values of 1 through 6.

Step 1: Simulate the rolls. That is, generate the 1s through 6s.

We can generate random numbers from 0 to 6, including 0 but not including 6 by multiplying `RAND()` by 6. As before, the command `INT(6*RAND())` will return the integer values from 0 to 6, including 0 but not including 6. However, we would like integers from 1 through 6, including both 1 and 6. To do this, we can add 1 to our function. The command `INT(6*RAND()+1)` will return integers from 1 through 7, including 1 but not including 7. In other words, `INT(6*RAND()+1)` will return a random integer from 1 to 6, including both 1 and 6. If we then count the number of 1s, 2s, etc. that are generated, we will have simulated the dice rolls.

Open a clean spreadsheet. In cell A1, type in `=INT(6*Rand()+1)`, and press **Enter**. Highlight cells A1 through A600. Click **Edit**, and choose the **Fill** and **Down** options. When this is done, cells A1 through A600 will be filled with 1s, 2s, 3s, 4s, 5s, and 6s. Again, these represent our rolls. Thus, we now simply need to count the results.

Note: Once again, as further entries are made in this spreadsheet, the random values previously determined will be regenerated. That is, the values currently in cells A1 to A600 will change as further entries are added to this spreadsheet.

Step 2: Tally the number of 1s, 2s, 3s, 4s, 5s, and 6s.

Once again, we will use *Excel's* `COUNTIF()` function to tally the results.

In cell C1, type in `=COUNTIF(A1:A600,1)`, and press **Enter**. The number of 1s that appears in cells A1 to A600 now appears in cell C1.

In cell C2, type in `=COUNTIF(A1:A600,2)`, and press **Enter**. The number of 2s that appears in cells A1 to A600 now appears in cell C2.

In cell C3, type in `=COUNTIF(A1:A600,3)`, and press **Enter**. The number of 3s that appears in cells A1 to A600 now appears in cell C3.

In cell C4, type in =COUNTIF(A1:A600,4), and press **Enter**. The number of 4s that appears in cells A1 to A600 now appears in cell C4.

In cell C5, type in =COUNTIF(A1:A600,5), and press **Enter**. The number of 5s that appears in cells A1 to A600 now appears in cell C5.

In cell C6, type in =COUNTIF(A1:A600,6), and press **Enter**. The number of 6s that appears in cells A1 to A600 now appears in cell C6.

Step 3: Calculate the relative frequencies of 1s, 2s, 3s, 4s, 5s, and 6s.

In cell D1, type in =C1/600. The relative frequency of 1s now appears in cell D1. Highlight cells D1 through D6. Click **Edit**, and choose the **Fill** and **Down** options. When this is done, cells D1 through D6 will contain the relative frequencies of 1s through 6s, respectively.

Once again, recall that when further entries are made in unused cells, the random numbers are regenerated. Click on an unused cell (such as E1). Press the **Delete** button. The random numbers are regenerated, and the tallies are adjusted accordingly. Repeat this process as often as you like to see that the number of 1s, 2s etc. remains near 100 each out of the 600 rolls.

Print page 1 of your spreadsheet to hand in: Click **File**, then **Print**. For the Print Range, click so that a bullet appears next to the “Page(s)” option. In the box next to “From:,” type in 1, and in the box next to “To:,” type in 1. Click the **OK** button and page 1 will print.

Repeat the procedure for a greater numbers of rolls, say 6000 rolls or 12,000 rolls. Do you expect similar results?

Warning: *Excel* may respond more slowly as you increase to a larger and larger number of trials. Also, be sure to use the new number of rolls in the calculations. For example, if you are simulating 12,000 rolls in cell D1 (in Step 3), you would enter =C1/12000 instead of =C1/600.

Simulating Rolls of a Pair of Dice

Suppose we roll two dice at the same time. What is the probability that the sum of the outcomes will be 2? _____ 3? _____ 4? _____ 5? _____
6? _____ 7? _____ 8? _____ 9? _____ 10? _____
11? _____ 12? _____

Because these eleven outcomes are not equally likely to occur, these probabilities are not as easy as to find as the previous ones. We will use *Excel* to find the relative frequencies of the outcomes for a large sample of rolls, say 1080. We can then estimate the actual probabilities.

Let's use *Microsoft Excel* to simulate 1080 rolls of the pairs of dice. We will use *Excel* to randomly select values of 1 through 6 for each die, and then add the results.

Step 1: Simulate the rolls of the pairs of dice and find the sum.

Open a clean spreadsheet. As before, the command `INT(6*RAND()+1)` will return a random integer from 1 to 6, including both 1 and 6. In cell A1, type in `=INT(6*Rand()+1)`, and press **Enter**. This simulates the roll of the first die. In cell B1, type in `=INT(6*Rand()+1)`, and press **Enter**. This simulates the roll of the second die. In cell C1, type in `=A1+B1`, and press **Enter**. This finds the sum of the two dice. Highlight cells A1 through C1080. Click **Edit**, and choose the **Fill** and **Down** options. When this is done, cells A1 through A1080 will be filled with 1s through 6s; cells B1 through B1080 will also be filled with 1s through 6s; and cells C1 through C1080 will be filled with sums of 2 through 12. Again, these represent our rolls and the sum of the outcomes of our rolls. Thus, we now simply need to count the results.

Note: Once again, as further entries are made in this spreadsheet, the random values previously determined will be regenerated. That is, the values currently in cells A1 to C1080 will change as further entries are added to this spreadsheet.

Step 2: Tally the number of each possible sum (2 through 12).

Once again, we will use *Excel's* `COUNTIF()` function to tally the results.

In cell D1, type in `=COUNTIF(C1:C1080,2)`, and press **Enter**. The number of sums of 2 that appears in cells C1 to C1080 now appears in cell D1.

In cell D2, type in `=COUNTIF(C1:C1080,3)`, and press **Enter**. The number of sums of 3 that appears in cells C1 to C1080 now appears in cell D2.

In cell D3, type in `=COUNTIF(C1:C1080,4)`, and press **Enter**. The number of sums of 4 that appears in cells C1 to C1080 now appears in cell D3.

In cell D4, type in =COUNTIF(C1:C1080,5), and press **Enter**. The number of sums of 5 that appears in cells C1 to C1080 now appears in cell D4.

In cell D5, type in =COUNTIF(C1:C1080,6), and press **Enter**. The number of sums of 6 that appears in cells C1 to C1080 now appears in cell D5.

In cell D6, type in =COUNTIF(C1:C1080,7), and press **Enter**. The number of sums of 7 that appears in cells C1 to C1080 now appears in cell D6.

In cell D7, type in =COUNTIF(C1:C1080,8), and press **Enter**. The number of sums of 8 that appears in cells C1 to C1080 now appears in cell D7.

In cell D8, type in =COUNTIF(C1:C1080,9), and press **Enter**. The number of sums of 9 that appears in cells C1 to C1080 now appears in cell D8.

In cell D9, type in =COUNTIF(C1:C1080,10), and press **Enter**. The number of sums of 10 that appears in cells C1 to C1080 now appears in cell D9.

In cell D10, type in =COUNTIF(C1:C1080,11), and press **Enter**. The number of sums of 11 that appears in cells C1 to C1080 now appears in cell D10.

In cell D11, type in =COUNTIF(C1:C1080,12), and press **Enter**. The number of sums of 12 that appears in cells C1 to C1080 now appears in cell D11.

Step 3: Calculate the relative frequency of each possible sum 2 through 12.

In cell E1, type in =D1/1080. The relative frequency of the sum of 2 now appears in cell E1. Highlight cells E1 through E11. Click **Edit**, and choose the **Fill** and **Down** options. When this is done, cells E1 through E11 will contain the relative frequencies of the possible sums 2 through 12, respectively.

The true probabilities for each possible sum 2 through 12 follow this section. Compare the relative frequencies you obtained with the actual probabilities. Are your relative frequencies close to the true probabilities?

Recall that when further entries are made in unused cells, the random numbers are regenerated. Click on an unused cell (such as F1). Press the **Delete** button. The random numbers are regenerated, and the tallies and relative frequencies are adjusted accordingly. Repeat this process as often as you like to see that the relative frequencies of sums 2 through 12 remain close to the true probabilities.

Print page 1 of your spreadsheet to hand in: Click **File**, then **Print**. For the Print Range, click so that a bullet appears next to the “Page(s)” option. In the box next to “From:” type in 1, and in the box next to “To:” type in 1. Click the **OK** button and page 1 will print.

Repeat the procedure for a greater numbers of rolls, say 3,600 rolls or 10,800 rolls. Do you expect similar results?

Warning: *Excel* may respond more slowly as you increase to a larger and larger number of trials. Also, be sure to use the new number of rolls in the calculations. For example, if you are simulating 3,600 rolls, in cell E1 (in Step 3), you would enter =D1/3600 instead of =D1/1080.

True Probabilities for Sums
When Rolling Two Dice.

Sum	Probability
2	$1/36 \approx 0.028$
3	$2/36 = 1/18 \approx 0.056$
4	$3/36 = 1/12 \approx 0.083$
5	$4/36 = 1/9 \approx 0.111$
6	$5/36 \approx 0.139$
7	$6/36 = 1/6 \approx 0.167$
8	$5/36 \approx 0.139$
9	$4/36 = 1/9 \approx 0.111$
10	$3/36 = 1/12 \approx 0.083$
11	$2/36 = 1/18 \approx 0.056$
12	$1/36 \approx 0.028$