THE LAW OF LARGE NUMBERS
EXCEL LAB #5

BUSN/ECON/FIN 130: Applied Statistics
Department of Economics and Business
Lake Forest College
Lake Forest, IL 60045

Overview

This lab is written for Excel 2010, which is available to students in the library. The notation => can be read as "go to" or "click on." This notation will most often be used when navigating the menu or toolbars in Excel. To indicate a command or icon that you might click on or search for in Excel, **bold** will be used. Likewise, anything that you are to type into Excel will be **bolded** in the instructions. Do not enter such text as bolded text unless the instructions ask you to do so.

In this tutorial we will explore the Law of Large Numbers portion of the Central Limit Theorem by creating three underlying distributions, taking random samples from each, and showing how the sample mean converges to the population mean with ever larger samples (and showing how the standard deviation of the sample averages converges to zero).

Tutorial

Open **Central Limit Theorem Data.xlsx**. There is a single column of data, in cells A2 through cells A5001. In particular, the data contains 5,000 random draws from a uniform random variable distributed from zero to one.

1. For columns A through N, set the width to 10 and center-justify the cells with 9 point Arial font.

2. In cell E2 enter **Average**. In cell E3 enter **Variance**. In cell F1 enter **Unif(0,1)**. In cell J1 enter <u>**Uniform**</u> (bold and underlined).

3. In cell F2 enter **=AVERAGE(A2:A5001)**. In cell F3 enter **=VAR(A2:A5001)**.

Recall that the expected value of a Unif[0,1] is 0.5 and its variance is 1/12 = 0.0833. Cell F2 shows the average of the 5,000 random draws on the Unif[0,1] to be 0.4996, which is very close to 0.5. Cell F3 shows the variance of the 5,000 random draws on the Unif[0,1] to be 0.0820, which is fairly close to 0.0833.

4. Make cells F2 and F3 report to 4 decimal places.

5. To demonstrate the Law of Large Numbers we are going to take several samples of various sizes from the 5,000 random draws in column A. To help keep track of things, in cell G3 enter **2**, in cell G4 enter **=G3+500**. Copy and paste cell G4 into cells G5 through G12. Cells G3 through G12 should now have the numbers 2, 502, 1002, 1502, …, 4002, 4502. These numbers represent the first row in each sample.

6. In cell J13 enter **Average**, and in cell J14 enter **St. Dev.**. In cells J3 through J12, enter, in order, **Sample 1**, **Sample 2**, …, **Sample 10**.

7. In cell K2 enter **N=10**, in cell L2 enter **N=100**, in cell M2 enter **N=250**, and in cell N2 enter **N=500**.

In column K, we will enter the average of 10 draws on the Unif[0,1] ten times. In column L, we will enter the average of 100 draws on the Unif[0,1] ten times. In column M, we will enter the average of 250 draws on the Unif[0,1] ten times. And in column N, we will enter the average of 500 draws on the Unif[0,1] ten times. Each set of draws will begin in the row as indicated in column J. Therefore:

8. In cell K3 enter **=AVERAGE(A2:A11)**, which is a sample of 10 draws. However, for reasons that will appear clear shortly, it will save time to enter in cell K3 **=AVERAGE($A$2:$A$11)**. Using column J as a guide, in cells K4 through K12, continue to enter in order:

   In cell K4: **=AVERAGE($A$502:$A$511)**
   In cell K5: **=AVERAGE($A$1002:$A$1011)**
   …
   In cell K12: **=AVERAGE($A$4502:$A$4511)**

   Each of the above averages, therefore, is an average over 10 draws.

9. In cells L3 through L12, repeat what was just entered in column K, but extend each sample for 100 draws.

   In cell L3: **=AVERAGE($A$2:$A$101)**
   In cell L4: **=AVERAGE($A$502:$A$601)**
   …
   In cell L12: **=AVERAGE($A$4502:$A$4601)**

   You might find it quicker to copy and paste cells K3 – K12 into cells L3 – L12, and then go back through each cell and change only the number associated with each formula.

10. In cells M3 through M12, repeat what was just entered in column K, but extend each sample for 250 draws.

    In cell M3: **=AVERAGE($A$2:$A$251)**
    In cell M4: **=AVERAGE($A$502:$A$751)**
    …
    In cell M12: **=AVERAGE($A$4502:$A$4751)**

11. In cells N3 through N12, repeat what was just entered in column K, but extend each sample for 500 draws.

    In cell N3: **=AVERAGE($A$2:$A$501)**
    In cell N4: **=AVERAGE($A$502:$A$1001)**
    …
    In cell N12: **=AVERAGE($A$4502:$A$5001)**

12. Now that you are done entering these formulas, delete cells G3 through G12: **left click on G3 and drag through G12** => **press the Delete key on the keyboard**.

13. In cell K13 enter **=AVERAGE(K3:K12)** and in cell K14 enter **=STDEV(K3:K12)**. Copy and paste cells K13 and K14 to columns L, M, and N.

14. Make the number formatting of cells K3 through N14 each have 4 digits after the decimal point.

At this point, this part of your spreadsheet should look like the following:

Uniform				
	N=10	N=100	N=250	N=500
Sample 1	0.4682	0.5088	0.4924	0.5080
Sample 2	0.5482	0.5101	0.5030	0.4894
Sample 3	0.4027	0.4951	0.4837	0.4785
Sample 4	0.3471	0.5061	0.5239	0.5070
Sample 5	0.4227	0.5121	0.4864	0.4951
Sample 6	0.5541	0.4723	0.5025	0.5053
Sample 7	0.6037	0.5419	0.4991	0.4965
Sample 8	0.4349	0.4873	0.4881	0.5051
Sample 9	0.5071	0.4948	0.4950	0.4954
Sample 10	0.4988	0.5606	0.5153	0.5156
Average	0.4788	0.5089	0.4989	0.4996
St. Dev.	0.0786	0.0258	0.0128	0.0107

The idea of the Law of Large Numbers is represented in the Average and Standard Deviation rows (J13 – N13 and J14 – N14). Specifically, in ten samples of 10 draws each, the average draw is 0.4788. This is 0.0212 away from the population mean of 0.5 for a Unif[0,1]. Ten samples of 100 draws each, however, average 0.5089, which is even closer to the population mean as it is only 0.0089 away from 0.5. Ten samples of 250 draws each is even closer at 0.4989, which is only 0.0011 away from the population mean. And, ten samples of 500 draws each is closer still at 0.4996, which is only 0.0004 away.

The standard deviation row shows that the standard deviation of sample means is going to zero as sample sizes increase. When only 10 draws are taken, the sample standard deviation is 0.0786. As the sample size increases, the sample standard deviation steadily falls (to 0.0107 when 500 draws are taken). Although not readily apparent in the four columns of data here, the sample standard deviation goes to zero as the sample size (N) approaches infinity.

This convergence of sample means to the population mean (and hence the convergence of the sample standard deviation to zero) as samples get larger is the Law of Large Numbers. (The Central Limit Theorem is a statement that the convergence of these statistical properties occurs according to an approximate normal distribution when the sample size gets large.)

15. Now we want to repeat what we have done above for a Bernoulli distribution with probability of success of 15%. In cell B1 enter **Bern(0.15)**.

A Bern(0.15) has a 15% chance of returning a success, and an 85% chance of returning a failure. It’s expected value, therefore, is 0.15 and its variance is 0.15 * 0.85 = 0.1275.

16. To start we want to enter 5,000 random draws on a Bernoulli distribution with 15% probability of success in column B. To enter these random draws, we will use the random draws from a Unif[0,1] in column A. In particular, we will associate with any value in column A that is under 0.15 to be a success (Bernoulli value = 1) and any value not under 0.15 to be a failure (Bernoulli value = 0). To do this, in cell B2 enter **=IF(A2<0.15,1,0)**.

17. Copy and paste this formula to cells B3 through B5001. Notice that cell B2 and B3 have 0's in them, as cells A2 and A3 are both above 0.15. The first cell in column B to report a success is cell B4 (which has a 1), as cell A4 is under 0.15 at 0.0754.

18. Enter **<u>Bern(0.15)</u>** in cell G1. Copy and paste cells F2 and F3 into cells G2 and G3. Cells G2 and G3 should report out to 4 decimal places. If this is not the case, change the cell formatting to make it so.

Notice that the average value of the 5,000 random draws on the Bern(0.15) has an average value of 0.1500, which is precisely the expected value of the random variable. Moreover, the variance of the 5,000 random draws is 0.1275, which is also precisely what the variance of the random variable is.

19. In cell J16 enter **<u>Bernoulli</u>**. In this case, enter the text as bold and underlined.

20. ***Find and Replace.***

    At this point we aim to reproduce in cells J18 through N34 exactly the same information for the Bern(0.15) contained in column B that we recorded from the Unif[0,1] in column A in cells J3 through N14. Copy and paste cells J2 through N14 to cells J17 through N29.

    You will notice that everything looks exactly the same as it did for the uniform. This is because all of the formulas included dollar signs before the letters and numbers. You now must proceed through all of the formulas in cells K18 through N27 and change all **A** to **B**. Rows 28 and 29 are correct, so there is no need to change them.

    Alternatively, rather than going through all 40 formulas and changing **A** to **B**, you can use Excel's find and replace tool, which works inside of formulas: **left click on cell J17** => **press CTRL-H** => **enter A under Find** and **enter B under Replace with** => **make sure you are searching by rows and in formulas** => **Find Next** => **Replace all such finds in rows 18 through 27**. Once all formulas have been changed, then the numbers in the cells K18 through N27 will be correct, and the averages and standard deviations in rows 28 and 29 will be correct as well.

21. If it is not already so, make the number formatting of cells K18 through N29 have 4 digits after the decimal point.

At this point, this part of your spreadsheet should look like the following:

Bernoulli

	N=10	N=100	N=250	N=500
Sample 1	0.1000	0.1200	0.1320	0.1340
Sample 2	0.2000	0.1300	0.1240	0.1400
Sample 3	0.5000	0.1700	0.1480	0.1700
Sample 4	0.3000	0.1100	0.1040	0.1260
Sample 5	0.2000	0.1600	0.1560	0.1520
Sample 6	0.1000	0.2400	0.1880	0.1660
Sample 7	0.1000	0.1100	0.1600	0.1640
Sample 8	0.2000	0.1500	0.1880	0.1500
Sample 9	0.1000	0.2100	0.1800	0.1560
Sample 10	0.0000	0.1300	0.1400	0.1420
Average	0.1800	0.1530	0.1520	0.1500
St. Dev.	0.1398	0.0435	0.0281	0.0145

Once again the idea of the Law of Large Numbers is represented in the Average and Standard Deviation rows (J28 – N28 and J29 – N29). Specifically, in ten samples of 10 draws each, the average draw is 0.1800. This is 0.03 away from the population mean of a Bern(0.15) which is 0.15. Ten samples of 100 draws each, however, average 0.1530, which is even closer to the population mean of 0.15 as it is only 0.0030 away. Ten samples of 250 draws each is even closer at 0.1520, which is only 0.0020 away. And ten samples of 500 draws each is closer still at 0.1500, precisely the population mean.

The standard deviation row shows that the standard deviation of sample means is going to zero as sample sizes increase. When only 10 draws are taken, the sample standard deviation is 0.1398. As the sample size increases, the sample standard deviation steadily falls (to 0.0145 when 500 draws are taken).

Now we want to repeat what we have done above for a trivariate distribution. In cell C1 enter <u>**Triv(0.2,0.65)**</u>. In column C we want to enter 5,000 random draws on a trivariate distribution that returns a 1 in 20% of the draws, returns a 2 in 45% of the draws, and returns a 3 in 35% of the draws. Thus, the expected value for this distribution is 1(0.2) + 2(0.45) + 3(0.35) = 2.15. Although we will not show it here, the variance of this trivariate distribution is 0.5275. To enter these random draws, we will use the random draws from a Unif[0,1] in column A. In particular, we will associate with any value in column A that is under 0.20 with a 1, any value in column A that is between .20 and .65 with a 2, and any value in column A above .65 with a 3.

22. In cell C2 enter **=1+IF(A2>0.2,1,0)+IF(A2>0.65,1,0)**. Copy and paste this value to cells C3 through C5001.

23. Enter <u>**Tri(0.2,0.65)**</u> in cell H1. Copy and paste cells F2 and F3 into cells H2 and H3.

24. Cells H2 and H3 should report out to 4 decimal places. If this is not the case, change the cell formatting to make it so.

Notice that the 5,000 random draws on the Triv(0.2,0.65) have an average value of 2.1498, which very close to the expected value of 2.15. Moreover, the variance of the 5,000 random draws is 0.5207, which is also close to the population variance of 0.5275.

25. In cell J31 enter **Trivariate** (in bold and underlined).

26. At this point we aim to reproduce in cells J33 through N44 exactly the same information for the Triv(0.2,0.65) contained in column C that we recorded from the Unif[0,1] in column A in cells J3 through N14. Copy and paste cells J2 through N14 to cells J32 through N44. You will notice that everything looks exactly the same as it did for the uniform. This is because all of the formulas included dollar signs before the letters and numbers. You now must proceed through all of the formulas in cells K33 through N42 and change all **A** to **C**. Rows 43 and 44 are correct, so there is no need to change them.

    Alternatively, rather than going through all 40 formulas and changing **A** to **C**, you can **find and replace** the **A** with **C** for rows 32 through 44. Once all formulas have been changed, then the numbers in the cells K33 through N42 will be correct, and the averages and standard deviations in rows 43 and 44 will be correct as well.

27. If it is not already so, make the number formatting of cells K33 through N44 have 4 digits after the decimal point.

At this point, this part of your spreadsheet should look like the following:

Trivariate				
	N=10	N=100	N=250	N=500
Sample 1	2.2000	2.2000	2.1440	2.1820
Sample 2	2.2000	2.1400	2.1440	2.1260
Sample 3	1.8000	2.1300	2.0960	2.0780
Sample 4	1.9000	2.1800	2.2200	2.1680
Sample 5	1.9000	2.1900	2.1320	2.1400
Sample 6	2.4000	2.1200	2.1560	2.1660
Sample 7	2.5000	2.2700	2.1640	2.1480
Sample 8	1.8000	2.1000	2.1080	2.1680
Sample 9	2.1000	2.1200	2.1320	2.1380
Sample 10	2.0000	2.2700	2.1880	2.1840
Average	2.0800	2.1720	2.1484	2.1498
St. Dev.	0.2440	0.0612	0.0364	0.0318

Once again the idea of the Law of Large Numbers is represented in the Average and Standard Deviation rows (J43 – N43 and J44 – N44). Specifically, both the average value is converging to the population mean of 2.15 and the sample standard deviation is converging to zero as N gets larger.

Exercises

In this exercise, we want to demonstrate the Law of Large Numbers one more time.

1. Repeat what you have done above by inserting a distribution (see point 2 below) in column D, provide the average and standard deviation in Column I, and provide samples and statistics in cells J46 through N59.

2. The distribution you are to use is uniform, specifically Unif[8,33]. Thus, you need to think about what formula to enter in column D, using the values in column A, to replicate 5,000 random draws on a Unif[8,33]. (Hint: a previous Excel lab required you to “stretch” a Unif(0,1). The same technique should be employed here.)

3. Fill in the table on the answer sheet to be turned in.

4. When you are done, save your file as YourName_Lab5.xlsx.

Turning in your work

Email YourName_Lab5.xlsx to your professor as a file attachment to an email with the subject heading Excel Lab 5: Your Name. Also print and turn in your filled-in answer sheet.

Answer Sheet for Lab #5: Law of Large Numbers

Name: ______________________

Fill in the following table with your results from the Unif[8,33].

Uniform[8,33]				
	N=10	N=100	N=250	N=500
Sample 1				
Sample 2				
Sample 3				
Sample 4				
Sample 5				
Sample 6				
Sample 7				
Sample 8				
Sample 9				
Sample 10				
Average				
St. Dev.				