Understanding the Central Limit Theorem Spreadsheet

The Excel "RAND()" function generates a random number drawn from a uniform distribution with values greater than or equal to 0 and less than 1. Therefore the Excel formula "=2*Rand()-1" generates a random number drawn from the uniform distribution with width 2 and lowest value -1.

This spreadsheet can generate draws from any uniform distribution of the form [a,b) by entering the value for a in Cell C2 and the value for b in Cell E2.

All uniform distributions [a,b) have a mean of the form a+b/2.
The uniform [-1, 1) distribution therefore has mean 0 [Cell H1].

All uniform distributions [a,b) have variance = ((b-a)^2)/12. The uniform [-1,1) distribution therefore has variance = 1/3 [Cell R1] and standard deviation = sqrt((b-a)^2)/12) = .577 [Cell J1].

Each column contains a "sample" of n = 100 values drawn at random from the uniform [-1,1) distribution [rows 16-115]. The "sample mean" for those 100 cells is given in row 6.

The *Central Limit Theorem* states that the distribution of a large number of sample means – the values found in row 6 – will be approximately Gaussian, with mean approaching the mean of the underlying distribution, namely 0, and standard deviation approaching the standard deviation of the original distribution divided by the square root of the sample size.

Since n = 100, and sqrt(100) = 10, for draws from [-1, 1) the standard deviation of the distribution of sample means should approach 1/(.577*10) = .0577

What do we mean by "approach"? As the number of sample means included grows large, the empirical distribution should get gradually closer to its theoretical Gaussian shape.

The mean of sample means is given in row 8. The difference between this value and the theoretical value of 0 is given in row 9. After 100 sample means, the difference is given in [Cell CW16]. After 250 sample means, the difference is given in [Cell IQ16 and Cell B9].

The standard deviation of sample means is given in row 12. The difference between this value and the theoretical value of .0577 is given in row 13. After 100 sample means, the difference is given in [Cell CW20]. After 250 sample means, the difference is still .001 [Cell IQ20 and Cell B10].

Note that every time you select one of the "Rand" cells and hit "Return" all the "Rand" cells generate new random values. To turn off automatic re-calculation of Rand() values to record a specific result, go to Excel/Preferences/Calculation and change "Automatically" to "Manually."

Note that the Excel formula for generating a random draw from a uniform distribution over [a, b) is always ((b-a)*Rand()) + a. This is pre-programmed into the Excel Spreadsheet.

Example

What are the mean, variance, and standard deviation of the uniform distribution over [-5, 8) ?

Answers: Change the interval [a,b) being studied to [-5, 8), by entering new values for "a" [Cell C2] and "b" [Cell E2]. The new Rand function is (13*Rand())-5, which is generated automatically.

The new mean = 1.5 [Cell H2]; variance = 14.08 [Cell N2]; and standard deviation = 3.753 [Cell N3].

What is the expected mean and standard deviation of sample means of sample size n = 50?

Answer: Change the sample size n to 50 [cell C3]. The new expected standard deviation is .5307 [Cell G6].