Demonstration of Central Limit Theorem Using Computer Simulation Results

The Central Limit Theorem states that the arithmetic mean of a random sample is distributed normally, with a population mean of the mean of the distribution from which the sample came, and a variance of the variance of the variance of the distribution from which the sample came divided by the number of the sample. Three collections of samples are attached. The first collection (“Sample A” or “A”) contains 100 samples of 5. The second collection (“Collection B” or “B”) contains 100 samples of 15. The third collection (“Sample C” or “C”) contains 100 samples of 30. By varying the number in each sample, the effect of increasing sample size in relation to the Central Limit Theorem will be shown. The distribution of the sample variance will also be examined.

The population from which the data was collected was an exponential distribution with mean 10, thus a variance of 100 and standard deviation of 10. This means that the mean of the distribution of sample means is 10. So, for collections A, B, and C, the expected value of the sample means is 10. The variances of each collection have different distributions, but they have the same expected value, 100.

Collection A is a collection of 100 samples of 5 observations from the exponential distribution. The means have a mean of 10.053 and a variance of 16.417. The expected variance is 20. With such a small sample size, the different variances are not that surprising. The mean is surprisingly close to the theoretical mean. A histogram of the means shows a poorly formed bell curve, barely defined. A confidence interval analysis with 95% confidence places the mean between 9 and 11 (numbers rounded). Collection A variances have a mean of 87.84. The variances of the means have a X2 (4) distribution. An estimation interval with 95% confidence places the population variance between 73 and 102 (numbers rounded).

Collection B is a collection of 100 samples 15 observations from the exponential distribution. The sample means have a mean of 10.132 with a variance of 5.995. The expected variance for the means in this collection is 6.667. The increase in sample size drastically reduced the variance in the means and while keeping the means centered around 10. The histogram of the means shows a more defined bell curve. This reflects the increase in sample size. A confidence interval with 95% confidence places the population mean between 10.7 and 9.7. This reduction in interval length reflects the increase in sample size. The variances have a mean of 97.06. The variances are distributed X2 (14). A confidence interval with 95% confidence places the population variance between 112 and 83.

Collection C is a collection of 100 samples of 30 observations each. The sample means have a mean of 9.983 with a variance of 2.943. The expected variance for the means in this collection is 3.333. The increase in sample size further reduced the variance in the means while better centering the means around the population mean. A histogram of the means shows a clearly defined bell curve, reflecting the increase in sample size. Further increasing the sample size would further define the bell curve, reflecting the distribution of the means as normal. A confidence interval with 95% confidence places the mean between 10.33 and 9.65. This further reduction in interval is a reflection of the increase in sample size. The variances have a mean of 100.22. The variances are distributed X2 (29). A confidence interval with 95% confidence places the population variance between 109 and 91.

These results clearly show the effects of increasing sample size in relation to the Central Limit Theorem. With each increase in sample size, the histogram of the means becomes more defined as a bell curve. The means grow closer to the population mean, and their variance is reduced. With each increase in sample size, the mean of the variances grows closer to it's expectation. This trend would continue if the simulation were repeated with larger and larger sample sizes. Increasing the sample size results in more organized samples, which results in more organized samples of samples, thus it is intuitively reasonable to expect that increasing sample size would result in more precise analysis on the samples.