# Validating the Central Limit Theorem on an Exponential Distribution

*Jonathan Bennett*

*Saturday, March 21, 2015*

## Overview

In probability theory, the central limit theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.

(Wikipedia) http://en.wikipedia.org/wiki/Central_limit_theorem (http://en.wikipedia.org/wiki/Central_limit_theorem), 3/21/2015

This analysis will sample 1000 different means of an exponential distribution of size 40, and calculate the simulated mean to show that it approaches the theoretical mean, that the distribution of those samples is approximately normally distributed, and that the standard deviation of the samples approaches the theoretical value of standard error.

## Simulations

```
    #generate 1000 means of 40 exponentially distributed numbers with lambda of 0.2
    #the for loop executes 1000 times
    #each execution generates 40 exponentially distributed numbers with lambda of .2
    #the mean of those 40 is appended to a running list of means.
    #by the end the means list will be 1000 long, each element representing a mean of 40
numbers
    #the same thing is done to generate a list of standard deviations
    set.seed(42)
    lambda <- 0.2
    samplesize <- 40
    totalsimulations <- 1000
    means <- NULL #initialize
    for(i in 1:totalsimulations) means=c(means,mean(rexp(samplesize,lambda)))
    simulationmean <- mean(means)
    simulationsd <- sd(means)
    simulationvar <- simulationsd^2
    theoreticalmean <- 1/lambda
    theoreticalsd <- (1/lambda)/sqrt(samplesize)
    theoreticalvar <- theoreticalsd^2
    meanaccuracy <- abs(round((simulationmean - theoreticalmean)/theoreticalmean*100,4))
    sdaccuracy <- abs(round((simulationsd - theoreticalsd)/theoreticalsd*100,4))
    varaccuracy <- abs(round((simulationvar - theoreticalvar)/theoreticalvar*100,4))
```

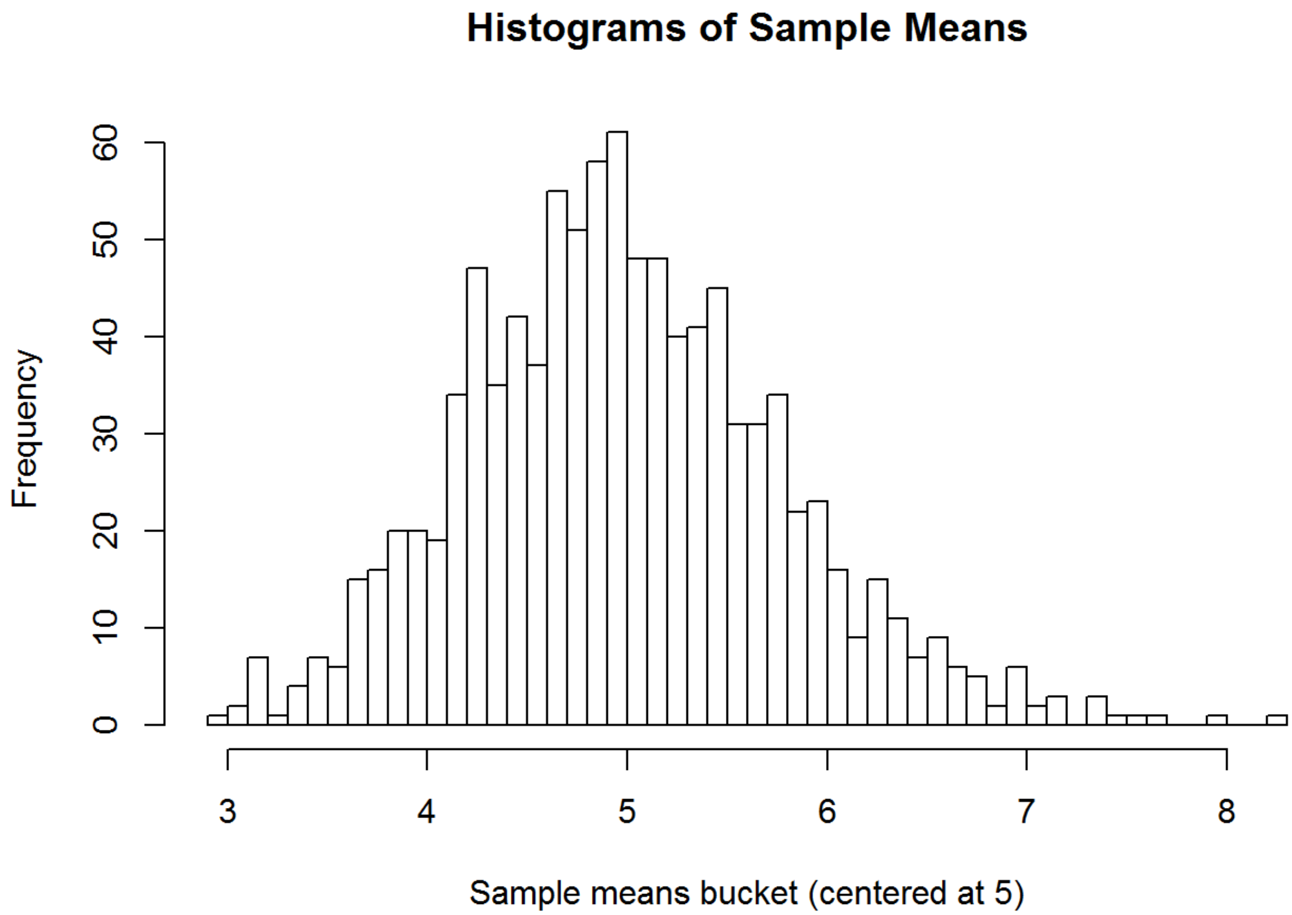
### The theoretical mean is 5 and the simulation mean is 4.9865083. The simulation mean is accurate within 0.2698%. (Validation of Central Limit Theorem)

### The theoretical sample standard deviation (standard error) is 0.7905694. The simulation standard deviation is 0.7965177. The simulated standard deviation is accurate within 0.7524%. (Validation of the standard error formula = sd(population)/sqrt(samplesize))

### The theoretical variance is 0.625. The simulation variance is 0.6344405. The simulated variance is accurate within 1.5105%.

The Central Limit Theorem says the sample will be approximately normally distributed. In order to visually validate this principle I plotted the distribution of the simulation means. It approximates a normal distribution. It is worth noting that increasing the number of simulations (say, to 100,000) greatly increases the accuracy of the approximations and creates a smoother, normalized curve approximation.

```
    hist(means,breaks=50, xlab = "Sample means bucket (centered at 5)", main = "Histogram
s of Sample Means")
```




```
print(summary(means))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.926   4.442   4.936   4.987   5.476   8.223
```