Exponential Distribution Vs. the Central Limit Theorem

Overview

This project explores the contrast between the observed and estimated means and variances of a large collection of randomly simulated exponential variables. This is achieved by comparing mean and variance values estimated by the Central Limit Theorem (CLT) to those observed through random simulation.

Simulations

```r
library(grid)        # Plotting graphics (arrow)
library(ggplot2)     # Plotting (ggplot etc.)
set.seed(123)
n <- 40
lambda <- 0.2
nSim <- 1000
exps <- NULL
means <- NULL
vars <- NULL
for (i in 1:nSim){
    exp <- rexp(n, lambda)
    exps <- cbind(exps, exp)
    means <- c(means, mean(exp))
    vars <- c(vars, sd(exp) ^ 2)
}
tMean <- 1 / lambda
tSD <- 1 / lambda
tVar <- tSD ^ 2
```

The data were produced via 1,000 simulations (*nSim*) each generating 40 observations (*n*) of random exponentials using a rate of 0.2 (*lambda*). These values are applied via iteration of a loop to populate values of the expnonentials (*exps*), means (*means*) and variabilities (*vars*).

Per the CLT both the estimated mean (*tMean*) and the estimated standard deviaion (*tSD*) are $(1/\lambda) = (1/0.2) = 5.0$. Theoretical variance (*tVar*) is the square of the theoretical standard deviation ($5.0^2 = 25$).

Sample Mean Vs. Theoretical Mean

## Histogram of Exponential Sample vs. Theoretical Means
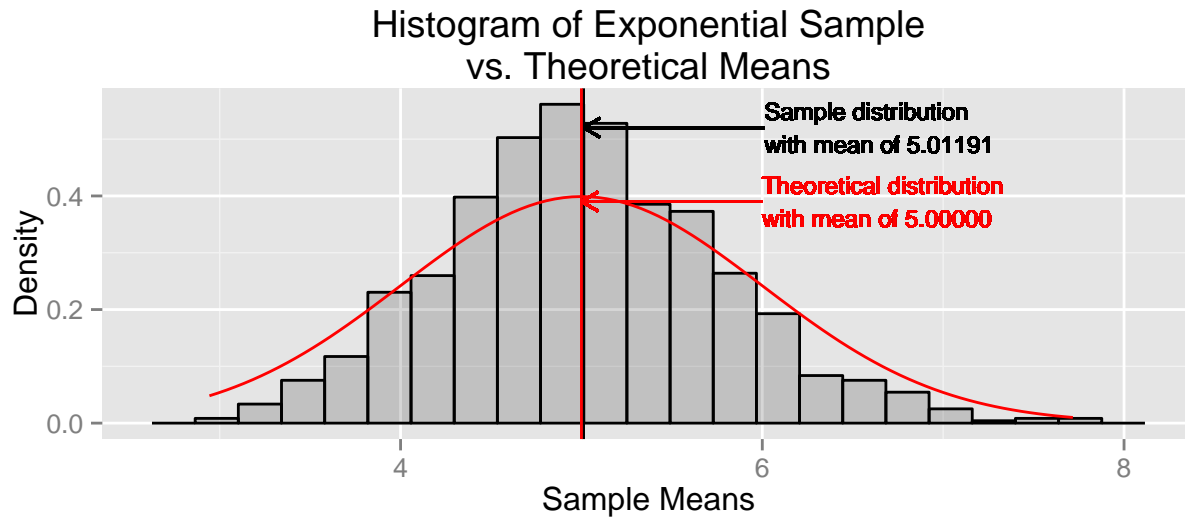
The base plot in this figure is comprised of a histogram of the sample means with overlayed data including the actual sample mean as well as the theoretical distribution and mean. This figure demonstrates that the sample mean (5.01191) is already closely approximated by the theoretical mean (5.0) as described above with a sample size of 1,000 simulations.

Sample Vaiance vs. Theoretical Variance

Histogram of Exponential Sample vs. Theoretical Variability
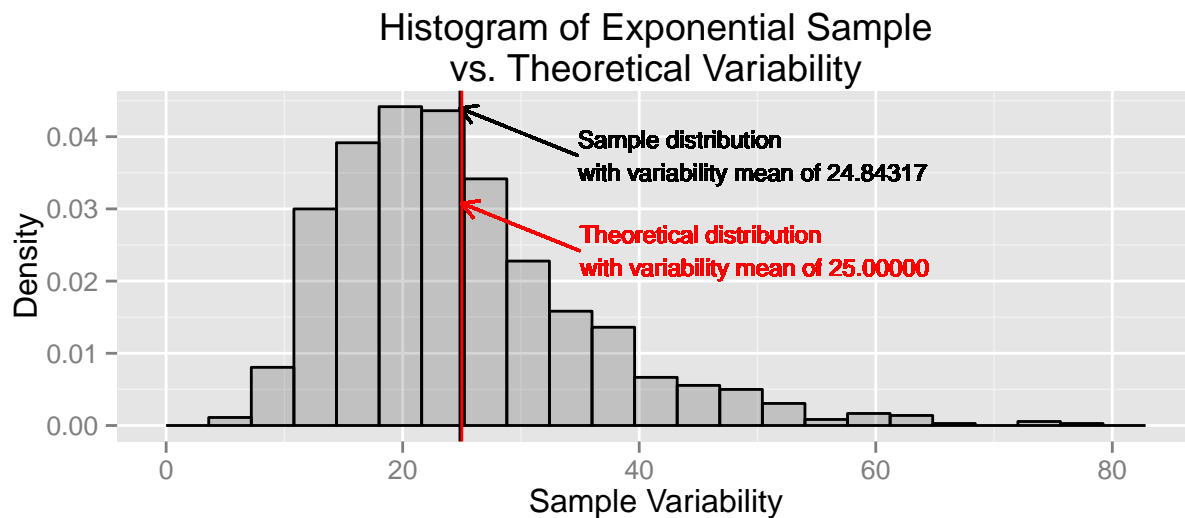
The base plot in this figure is comprised of a histogram of the variance of the sample with overlayed data including the actual sample variance mean and the theoretical variance mean. This figure demonstrates that the sample variance (24.84317) is already reasonably approximated by the theoretical variance (25.0) as described above with a sample size of 1,000 simulations.

Distribution

Density of Means for Exponential Variables vs. Density of Random Norma
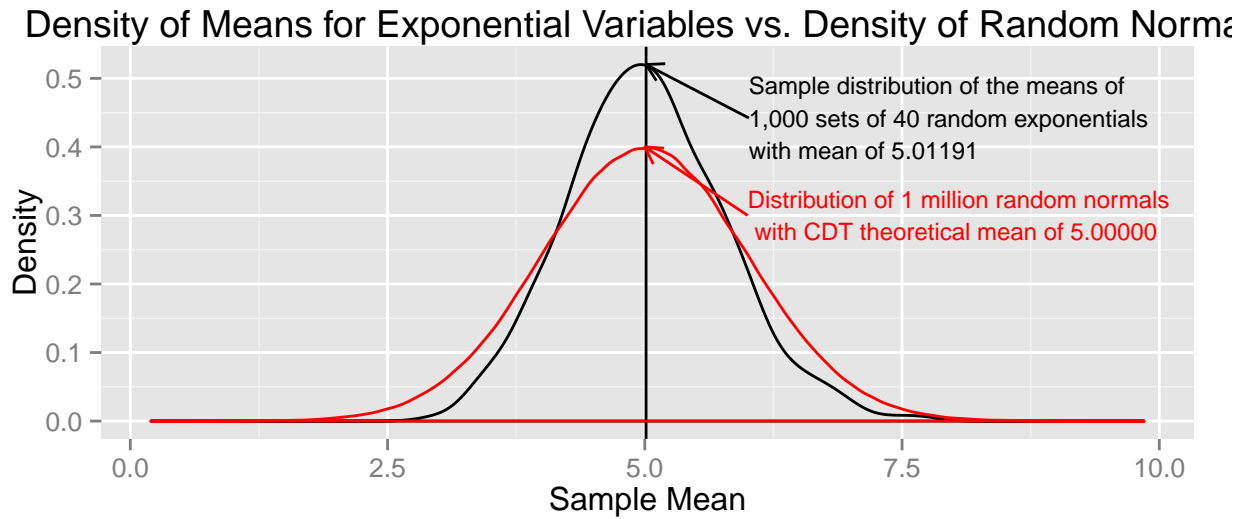
The base plot in this figure is comprised of the density of the means of the random exponential variables as describeda bove. The overlayed plot (in red) is the density of 1 million random normals utilizing the theoretical mean as estimated by the CDT: $(1/\lambda) = (1/0.2) = 5.0$ This demonstrates that the distribution of a large set of means of exponentials is approximately normal. As the sample size increases the distribution converges to the distribution of a large set of random normals.