Comparing Exponential Distribution to the Central Limit Theorem

Travis Fell

Thursday, June 11, 2015

Overview

This analysis compares the distributions, means and variances of an exponential distribution to the same for the distribution of sample means from that exponential distribution. The end goal is to show the central limit theorem in action.

Simulations

This following R code creates the exponential distribution, finds the mean and variance for the distribution, pulls the requisite number of samples, finds the distribution of all those sample means and variances for comparison to the theoretical mean and variances later.

First, let's create the exponential distribution as well as its mean and variance.

```
exp <- rexp(1000, .02)
exp_mu <- (1/.2)
exp_var <- (1/.2)
```

The overall population mean, exp_mu equals 5 and the population variance, exp_var equals 5.

With the parameters of the overall population established, let's take 1000 averages from 40 exponential samples then calculate the mean and variance of the distribution of those sample means.

```
exp_samp <- NULL
for(i in 1:1000) exp_samp <- c(exp_samp, mean(rexp(40, .2)))
exp_samp_mean <- mean(exp_samp)
exp_samp_var <- exp_var/40
```

The collection of sample means equals 4.9996758 and the sampling variance equals 0.125.

Sample Mean Vs. Theoretical Mean

The collection of sample means of 4.9996758 tracks relatively closely to the the mean of the exponential distribution, which is 5. This close correspondence supports the property of the Central Limit Theorem that the expected value of the sample means should be the same or close to the mean of the original distribution.

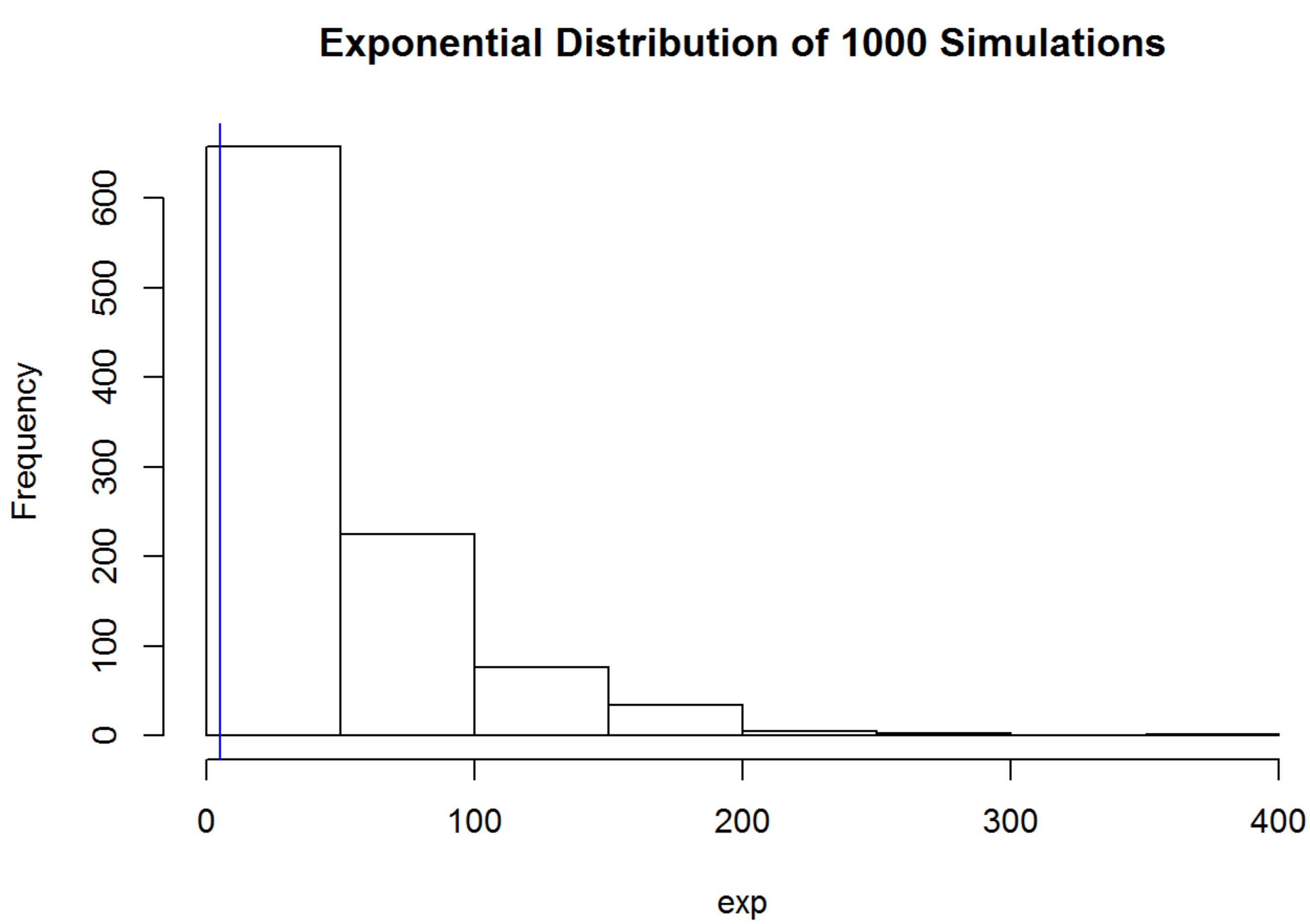
Sample Variance Vs. Theoretical Variance

The variance of the sample means of 0.125 is dramatically smaller than that of the theoretical variance of the exponential distribution of 5. Given that 1000 samples were used in the simulation, this narrow sample variance demonstrates that as sample sizes increase, the distribution tends to “tighten up” around the sample mean.

Distribution

The code and plot below describe the exponential distribution.

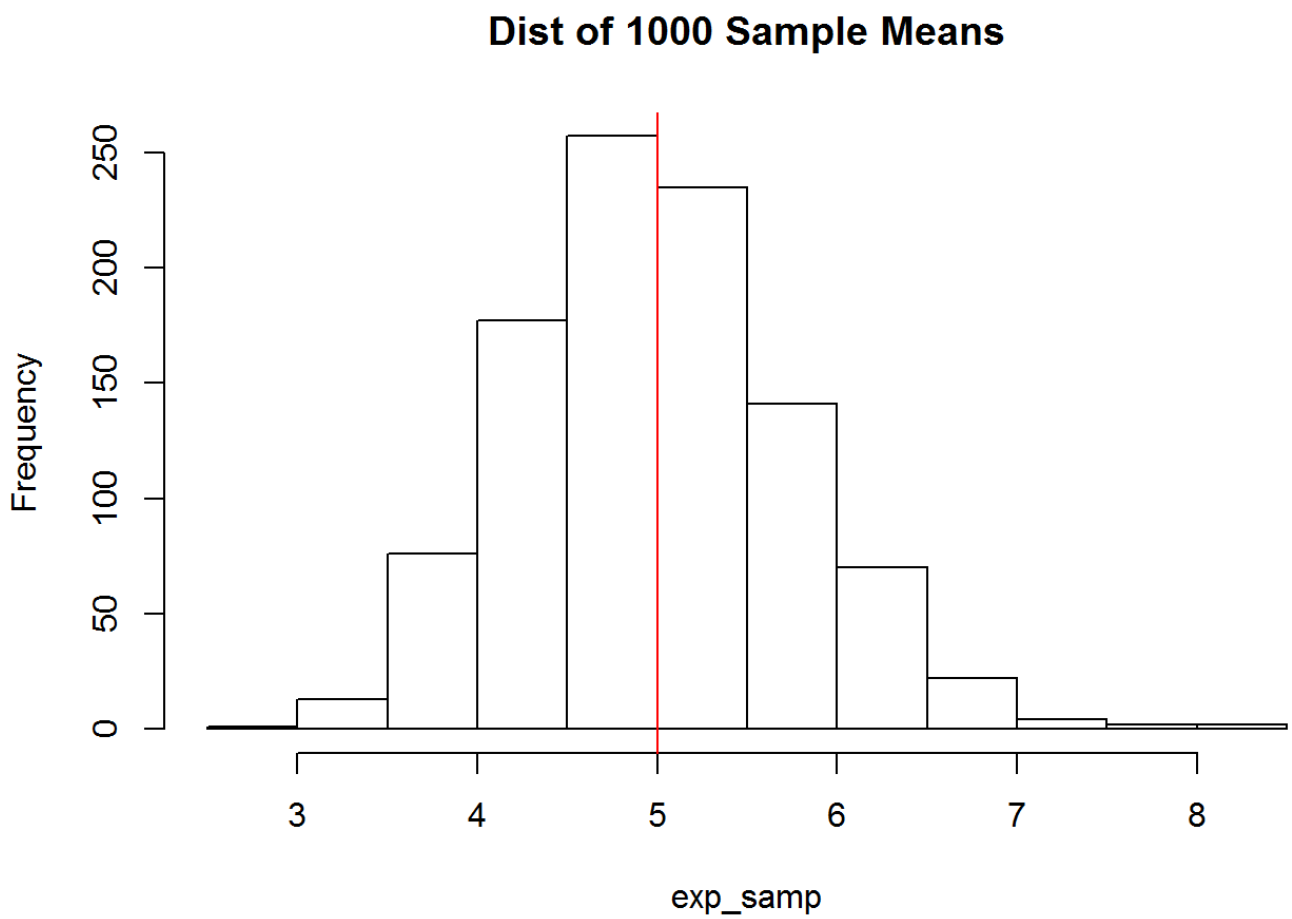
```
hist(exp, main = "Exponential Distribution of 1000 Simulations")
abline(v=exp_mu, col = "blue")
```



This distribution starts high and “ski slopes” down toward zero.

Now let’s look at the distribution of sample means from 40 exponential samples.

```
hist(exp_samp, main = "Dist of 1000 Sample Means")
abline(v=exp_samp_mean, col="red")
```



The distribution of the sample looks more like a normal distribution. These two plots put together illustrate why the distribution of sample means for any distribution will tend to be a normal looking distribution: For a single sample, the mean will be the “balancing point” for the sample. Because the mean will vary from sample to sample but should more often be close to the true mean of the population than far away, the collection of balancing points from many samples will create a normal looking distribution.