Central Limit Theorem Comparsion

Christopher Hair

September 15, 2017

Overview

In this article we will discuss a comparison between the Central Limit Thereom and exponential distributions. Using simulated data we will look at how some of the properties of an exponential distribution will relate to the Central Limit Thereom. Specifically we will look at how the sample mean and sample variance relate to the theoretical mean and variance of the Central Limit Thereom and how the distribution of the simulated sample aligns with what is know about the Central Limit Thereom.

Build the Sample Distribution

So to start we will need to build the simulated data to use for comparing the distribution to the Central Limit Thereom. The data will be comprised of 1000 mean values of a random exponential value of 40 at a rate of 0.2. From there we use the cumulative sum of the data to use in the comparison to the theoretical mean.

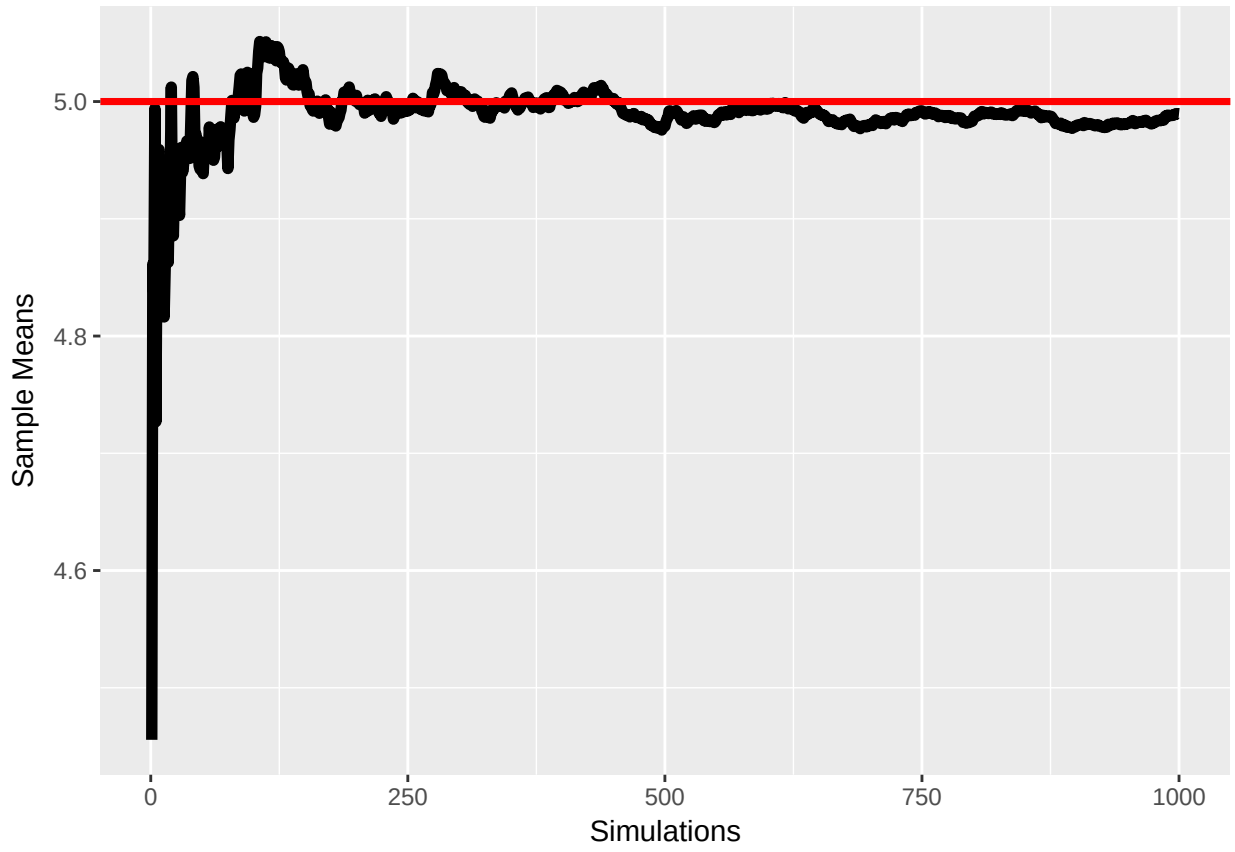
```
nosims <- 1000
n <- 40
lambda <- 0.2

sample_means <- NULL
for(i in 1:nosims) sample_means <- c(sample_means, mean(rexp(n,lambda)))

t_mean <- 1/lambda
cumulative_means <- cumsum(sample_means)/c(1:nosims)
sample_mean <- cumulative_means[nosims]

t_var <- (1/lambda)^2
sample_variance <- var(sample_means)*n
```

Comparison

Sample Mean vs. Theoretical Mean

After the calculations of the simulation data is complete we can see that the sample mean is ***4.9895908 ~ 5*** as the theoretical mean. We can visualize this approximation in using the following representation.

```
ggplot(data.frame(x = 1:nosims, y = cumulative_means), aes(x = x, y = y)) +
    geom_line(size = 2) +
    geom_abline(intercept = t_mean, slope = 0, color = "red", size = 1.25) +
    scale_y_continuous(limits = c(min(cumulative_means), max(cumulative_means))) +
    xlab("Simulations") +
    ylab("Sample Means")
```



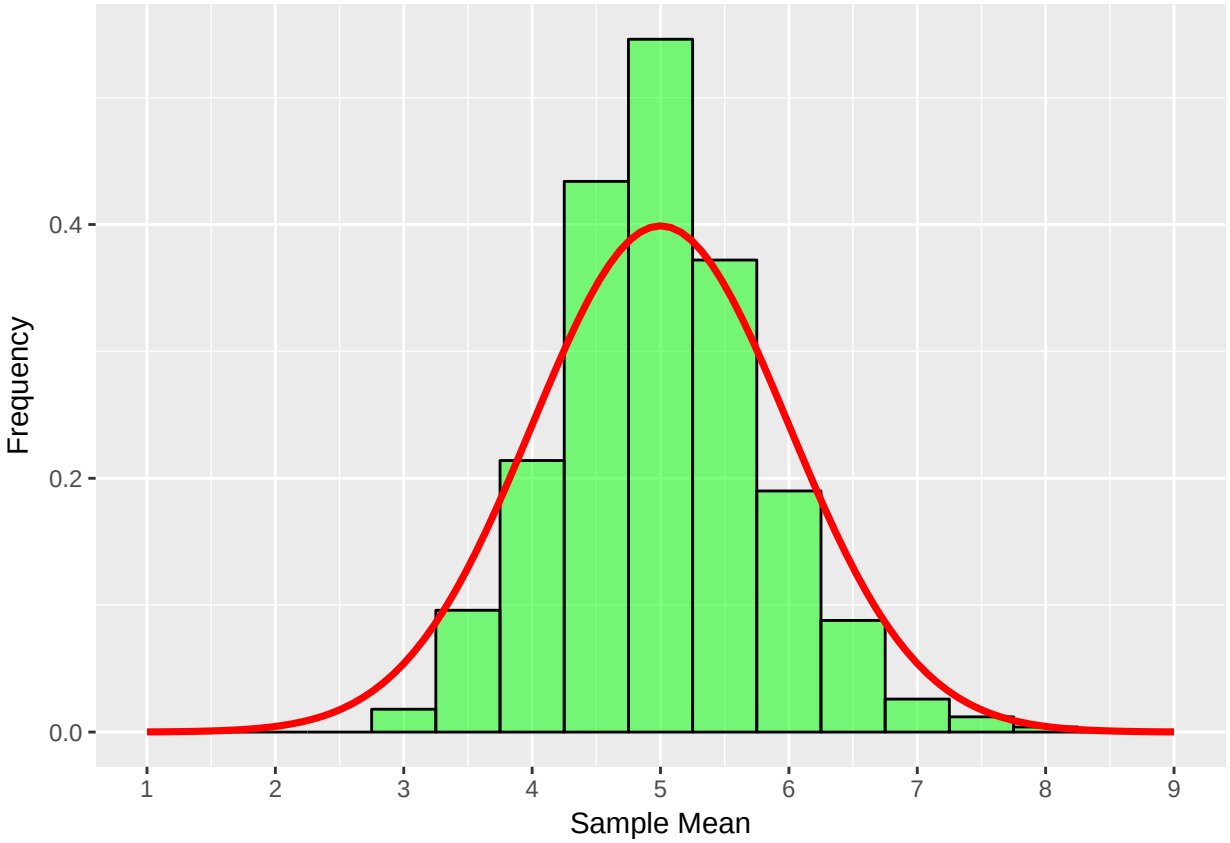
Sample Variance vs. Theoretical Variance

The comparison of the sample distribution variance to the theoretical variance is calculated as ***24.9881227 ~ 25***

Comparing the Sample Means to a Normal Distribution

The final comparison will be to show how the sample means of the exponential distribution are follow the rules of the Central Limit Theorem and follow the bell curve of a normal distribution.

```
ggplot(data.frame(x = sample_means), aes(x = x)) +
    geom_histogram(position = "identity",
                   binwidth = 0.5,
                   fill = "green",
                   color = "black",
                   alpha = 0.5,
                   aes(y = ..density..)) +
    stat_function(fun = dnorm, color = "red", size = 1.25, args = list(mean = t_mean)) +
    scale_x_continuous(breaks=c(1:9), limits=c(1,9)) +
    xlab("Sample Mean") +
    ylab("Frequency")
```


Frequency
0.4
0.2
0.0
1
2
3
4
5
6
7
8
9
Sample Mean