Part 1 of the Statistical Inference Project

Andy Tse

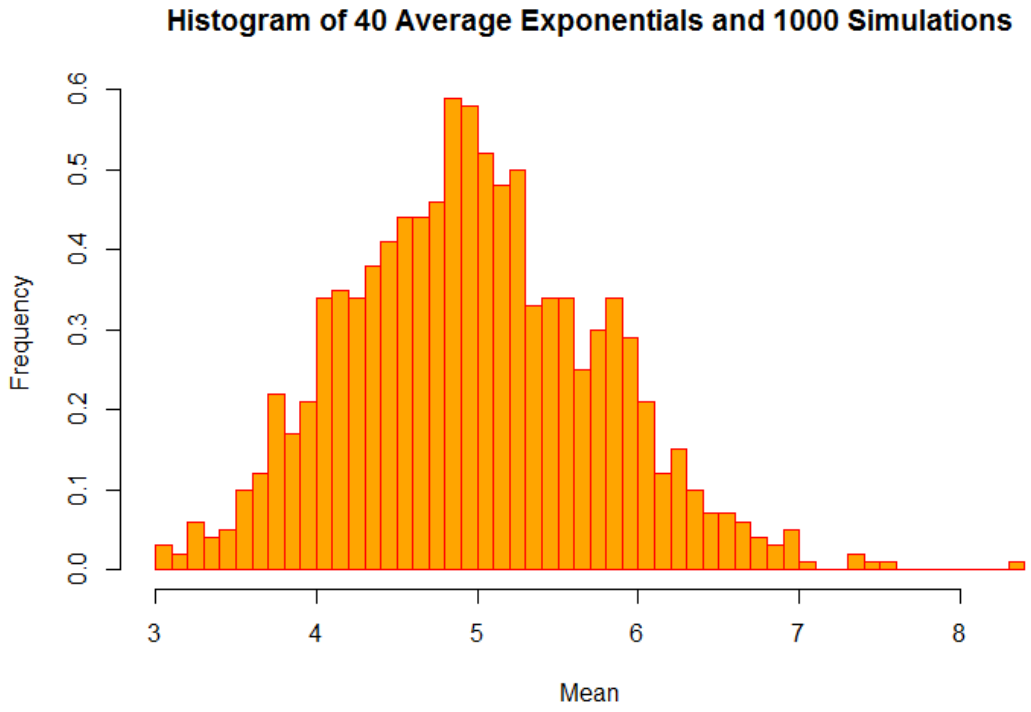
August 7, 2016

Overview

For the first part of the Statistical Inference Project, there will be 40 random exponentials and 1000 stimulations that will be analyzed. It is going to show comparisons between the sample mean versus the theoretical mean. It is also going to illustrate the differences between the sample variance and the theoretical variance.

Simulations:

```
lambda <- 0.2
sample_size <- 40
sims <- 1000
set.seed(12345)
sim_matrix <- matrix(rexp(sims*sample_size, rate=lambda), sims,
sample_size)
sim_mean <- rowMeans(sim_matrix)
hist(sim_mean, breaks=40, prob=TRUE,main="Histogram of 40 Average
Exponentials and 1000 Simulations", xlab="Mean", ylab="Frequency",
col="orange", border="red")
```



Sample Mean Versus Theoretical Mean Comparison:

In this section, I am going to analyze the differences in order to determine the sample versus the theoretical mean in this project.

```
sample_mean <- mean(sim_mean)
sample_mean

[1] 4.971972

theoretical_mean <- 1/lambda
theoretical_mean

[1] 5
```

As it shows, both comparisons between the sample mean and theoretical mean are both close to each other statistically.

Sample Variance Versus Theoretical Variance Comparison:

For this section, there will be an analysis for the differences for the 1000 stimulations on the sample variance in comparison to the theoretical variance comparisons.

```
variable_sample_mean <- var(sim_mean)
variable_sample_mean
```

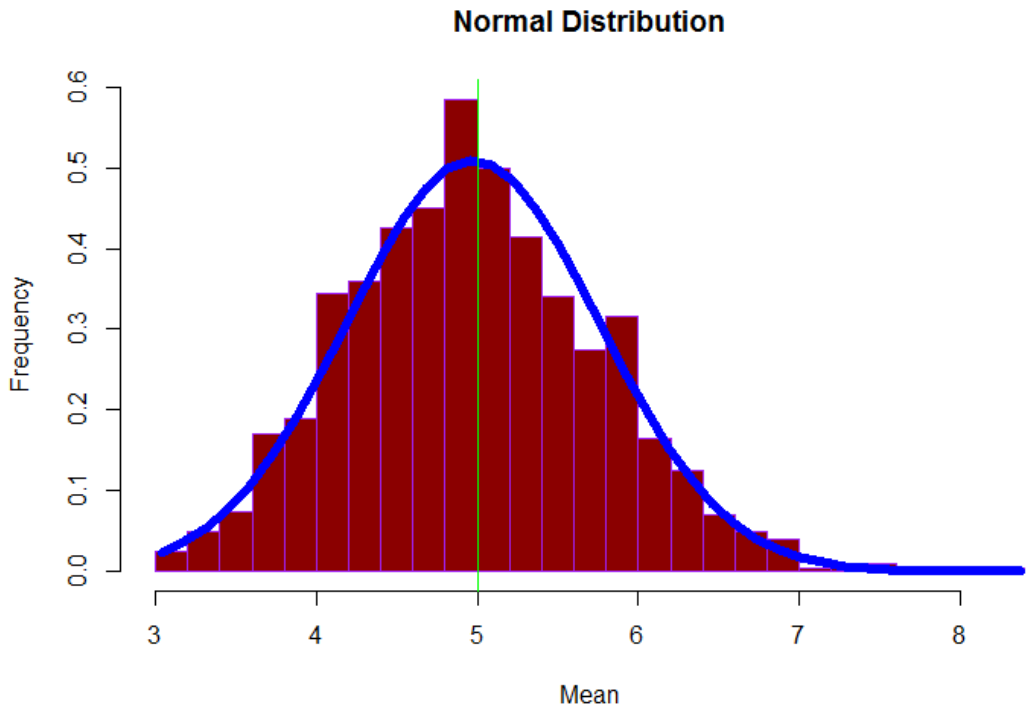
```
[1] 0.6157926

theoretical_var_mean <- ((1/lambda)/sqrt(sample_size))^2
theoretical_var_mean

[1] 0.625
```

As it shows, the sample variance and theoretical variance are almost close to each other.

Distribution:

In this part of the project, this is the graph demonstrating the information on the 1000 simulations in a normal distribution.

```
hist(sim_mean, breaks=30, prob=TRUE,main="Normal Distribution",
xlab="Mean", ylab="Frequency",
col="darkred", border="purple")
xfit <- seq(min(sim_mean), max(sim_mean), length = 40)
yfit <- dnorm(xfit, mean = mean(sim_mean), sd = sd(sim_mean))
lines(xfit, yfit, lwd=6, pch=20, col="blue")
abline(v=1/lambda,col="green")
```



Based on the comparisons from the first graph and the second graphs, it has been determined that the graph is distributed normally.