Statistical Inference Project Part 1

Jing Wei Chan

23 August 2015

Overview

In this project, we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. lambda is set to 0.2 for all of the simulations. The distribution of averages of 40 exponentials will be investigated and a thousand simulations will be performed.

Simulations

We will first run 1000 simulations of 40 exponentials each and get the average of these 1000 simulations. These 1000 simulations will be stored in a matrix.

```
## Initialize values
set.seed(1)
lambda <- 0.2
sample_size <- 40
simulations <- 1000

## create a matrix with 1000 simulations of 40 exponentials each
sim <- matrix(rexp(simulations *  sample_size, rate = lambda), simulations)

## get the mean for the 1000 simulations
sim_mean <- rowMeans(sim)
```

Sample Mean versus Theoretical Mean

Let us first calculate the sample mean.

```
sample_mean <- round(mean(sim_mean),3)
```

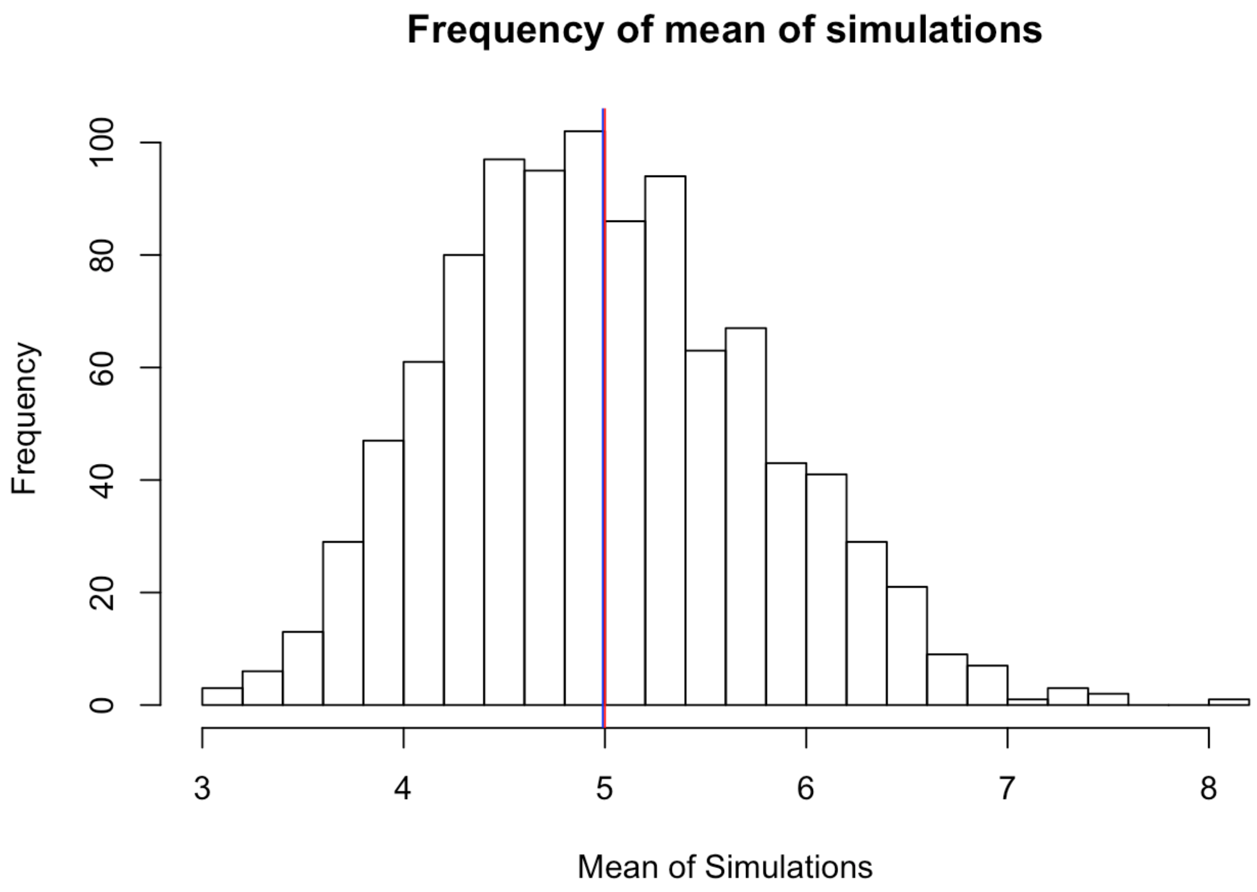
The sample mean is 4.99.

Next, we calculate the theoretical mean, 1/lambda.

```
theoretical_mean <- 1/lambda
```

The theoretical mean is 5.

```
hist(sim_mean,
     breaks = 20,
     xlab = "Mean of Simulations",
     ylab = "Frequency",
     main = "Frequency of mean of simulations")
abline(v = sample_mean, col = "blue")
abline(v = theoretical_mean, col = "red")
```



As shown in the diagram above, the sample mean 4.99 (in blue) is very close to the theoretical mean 5 (in red). The colour shown is in purple as the lines are very close to one another.

Sample Variance versus Theoretical Variance

Let us first calculate the sample variance.

```
sample_variance <- round(var(sim_mean),3)
```

The sample mean is 0.618.

Next, we calculate the theoretical mean, (1/lambda)^2 * (1/sample_size.

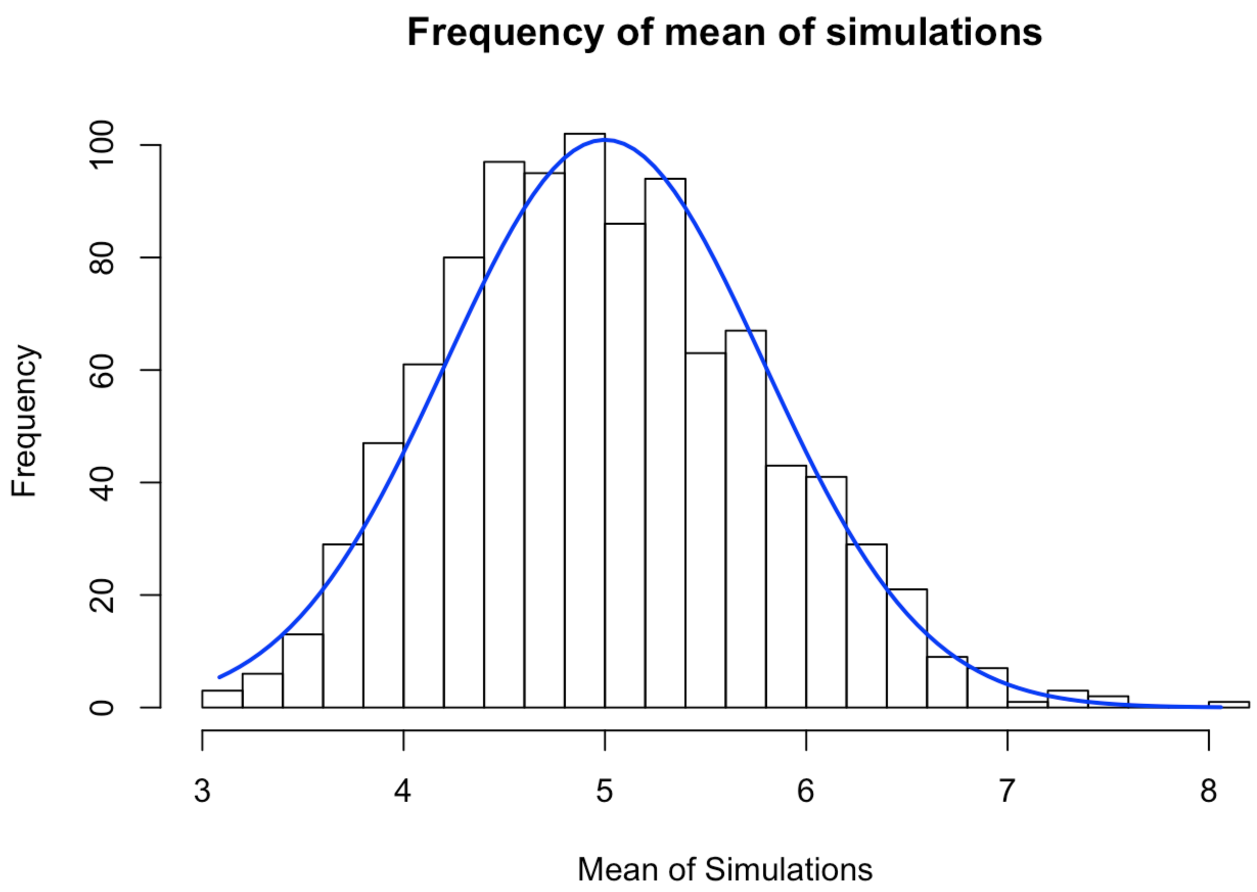
```
theoretical_variance <- (1/lambda)^2 * (1/sample_size)
```

The theoretical mean is 0.625.

Hence, we can see that the sample variance 0.618 is very close to the theoretical variance 0.625.

Distribution

```
h <- hist(sim_mean,
          breaks = 20,
          xlab = "Mean of Simulations",
          ylab = "Frequency",
          main = "Frequency of mean of simulations")
xfit <- seq(min(sim_mean), max(sim_mean), length = 100)
yfit <- dnorm(xfit, theoretical_mean, sqrt(theoretical_variance))
yfit <- yfit * diff(h$mids[1:2]) * length(sim_mean)
lines(xfit, yfit, col="blue", lwd=2)
```



As shown in the diagram, the distribution is approximately normal as shown by the normal distribution curve in blue.