Stastistical Inference Course Project

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be expDistributionulated 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. Set lambda = 0.2 for all of the simulations.

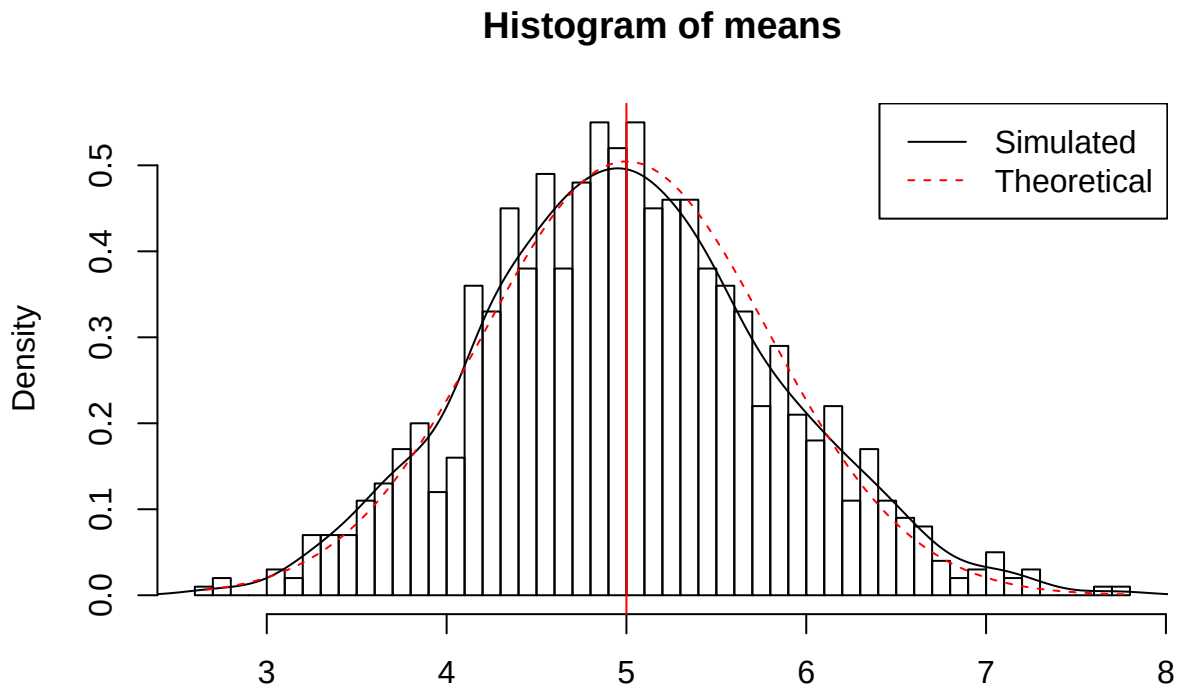
We will investigate the distribution of averages of 40 exponentials. Lets do the 1000 simulations.

```
#Load the libraries
library(ggplot2)

#set the seed for reproducability
set.seed(31233)

#set-up the simulaiton
lambda <- 0.2
num_sims <- 1000
sample_size <- 40
expDistribution <- matrix(rexp(num_sims*sample_size, rate=lambda), num_sims, sample_size)
means <- apply(expDistribution, 1, mean)
```

1. Sample Mean vs Theoretical Mean

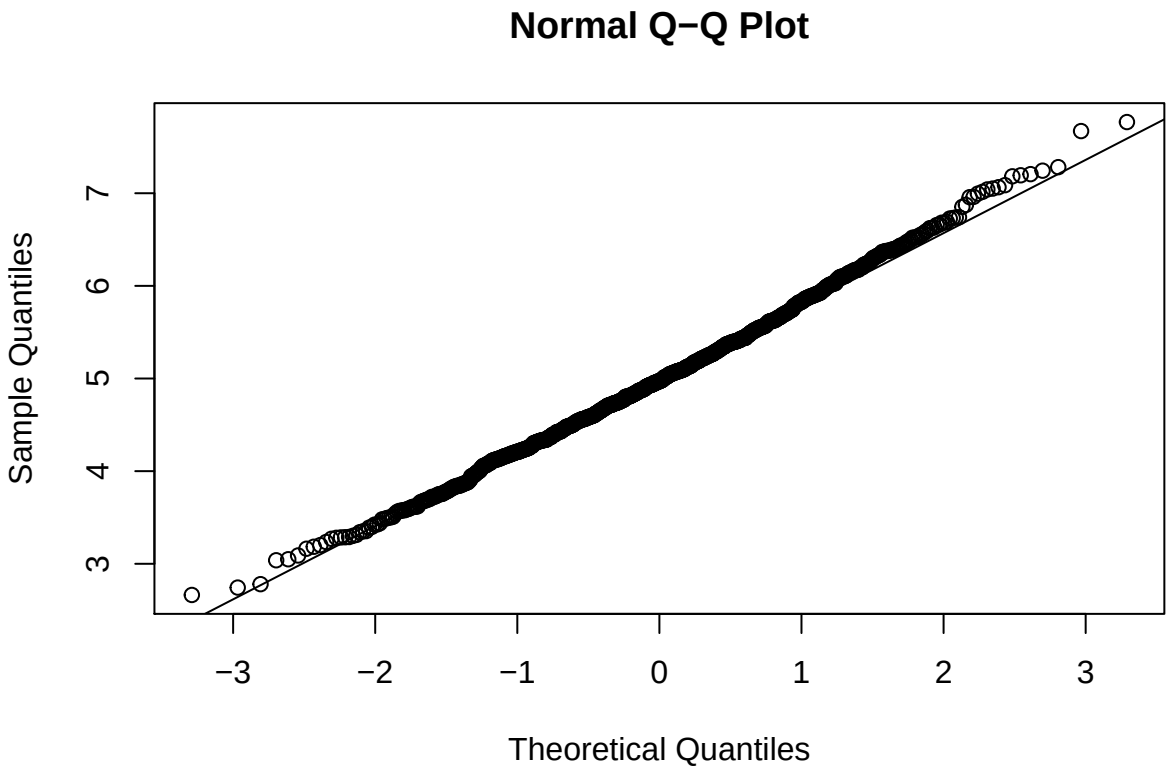


The distribution of sample means is centered at 5.0030944 and the theoretical center of the distribution is 5.

2. Sample Variance vs Theoretical Variance

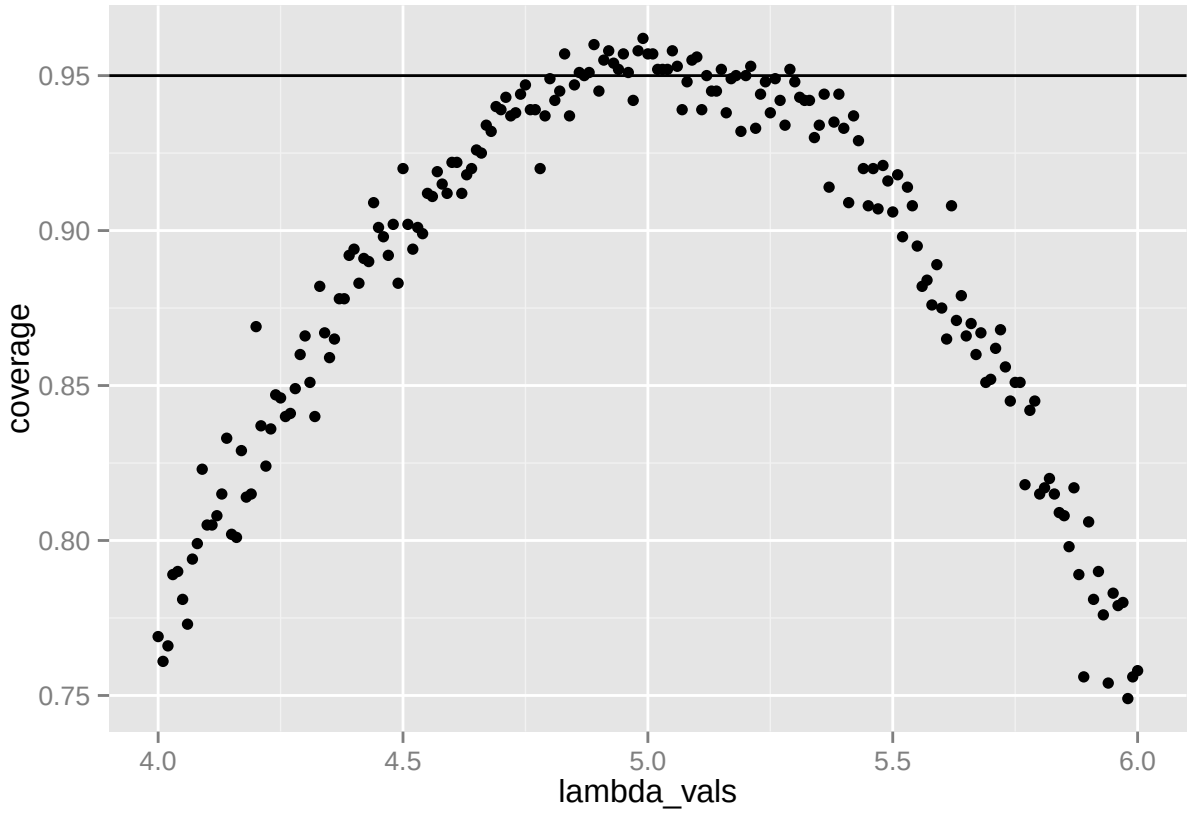
The variance of sample means is 0.6615225 where the theoretical variance of the distribution is $\sigma^2/n = 1/(\lambda^2 n) = 1/(0.04 \times 40) = 0.625$.

Due to the central limit theorem, the averages of samples follow normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. Also, the q-q plot below suggests the normality.



3. Distribution

Let's evaluate the coverage of the confidence interval.



As you can see from the plot above, for lambda around 5, the average of the sample mean falls within the confidence interval, at least 95% of the time. The true lambda is 5.