Coursera Data Scientist Track - Statistical Inference Course Project, Part One: Simulation Exercises

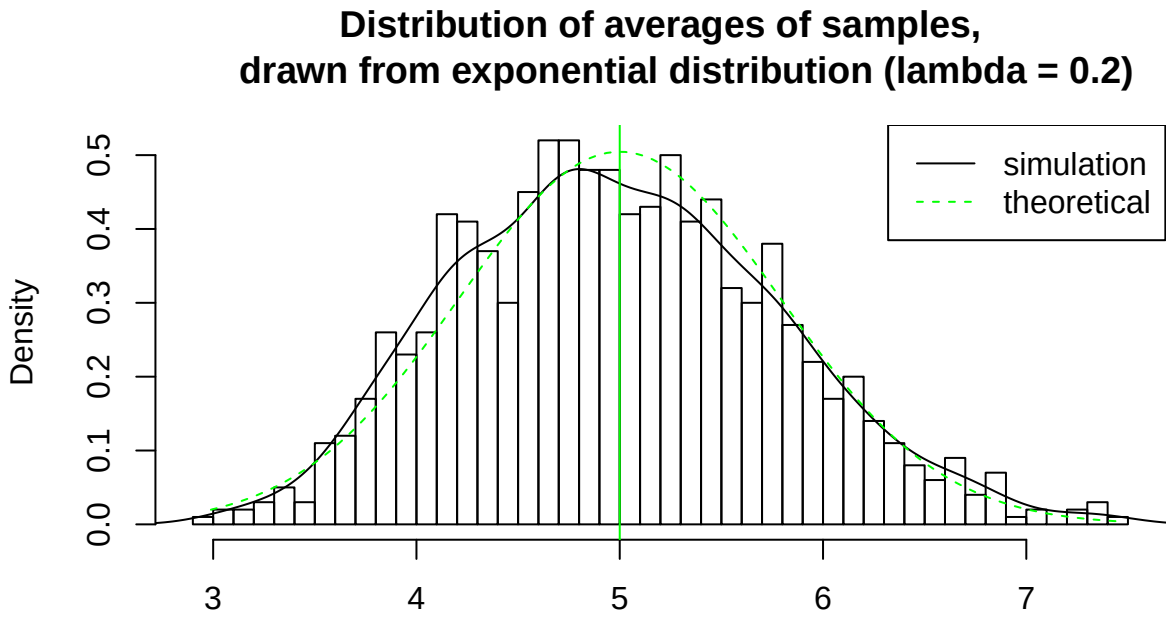
Pieter Jonckheere

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. Set lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials. We will do this based on a thousand simulations.

Perform a thousand simulated averages of 40 exponentials.

```
set.seed(3)
lambda <- 0.2
number_sim <- 1000
sample_size <- 40
sim <- matrix(rexp(number_sim*sample_size, rate=lambda), number_sim, sample_size)
row_means <- rowMeans(sim)
```

The distribution of sample means is as follows.



The distribution of sample means is centered at 4.9866197 and the theoretical center of the distribution is $\lambda^{-1} = 5$. The variance of sample means is 0.6257575, 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 a normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values.