Statistical Inference Project

Marcelo Nobrega

Part 1: Simulation Exercise Instructions

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

```r
set.seed(2017)
lambda <- 0.2
n <- 40 # number of exponentials
nosim <- 1000 # number of simulations

# The exponential distribution can be simulated in R with rexp(n, lambda)
# where lambda is the rate parameter
exp_sim <- matrix(rexp(n * nosim, rate = lambda), nosim) # simulating data
means <- rowMeans(exp_sim) #calculating the means
mean_sim <- mean(means)
mean_sim

## [1] 4.982863
```
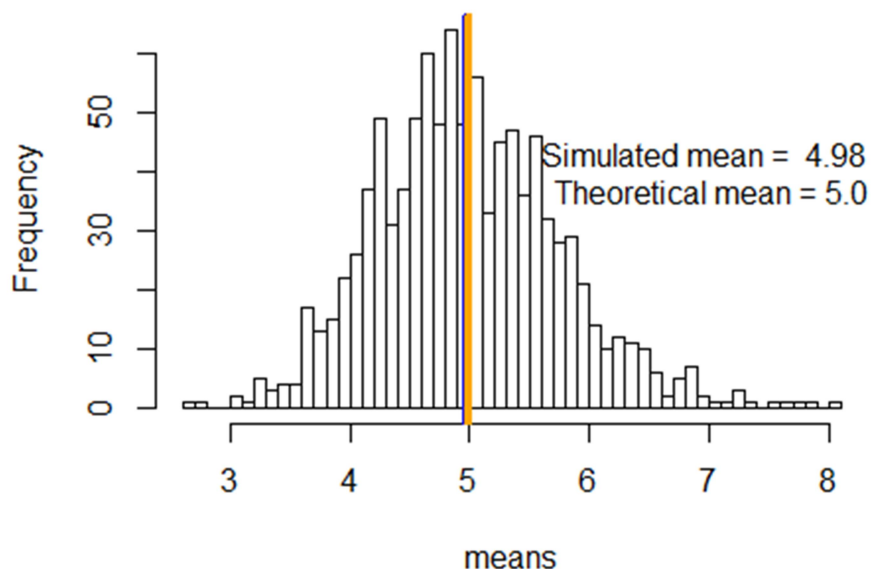
Theoretical mean

```r
T_mean <- 1/ lambda
T_mean

## [1] 5

hist(means, col = "white", main = "rexp mean distribution x Theoretical Mean", breaks = 40)
abline(v=mean_sim, lwd = "4", col = "blue")
abline(v=5.0, lwd = "4", col = "orange")
text(7, 40, paste("Simulated mean = ", round(mean_sim, 2), "\n Theoretical mean = 5.0"), col = "black")
```

rexp mean distribution x Theoretical Mean

The simulated mean is 4.9828 and the theoretical mean is 5, showing that the center of distributuion of the simulated data is very close to the theoretical center of distribution.

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
Expvar <- var(means)
Expvar
```
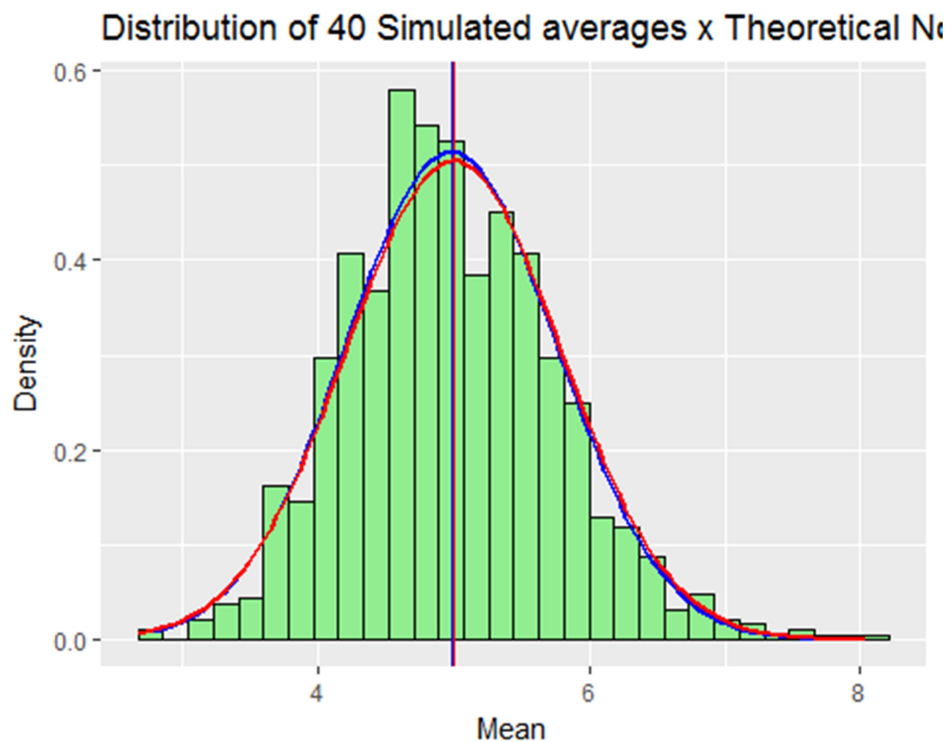
```
## [1] 0.6035768
```

```
# Theoretical Var
T_var <- (1/lambda)^2/n
T_var
```

```
## [1] 0.625
```

Variance of the sample means is 0.6035 while the theoretical variance is 0.625, very close values.

3. Show that the distribution is approximately normal.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Distribution of 40 Simulated averages x Theoretical N

As presented the the simulatd mean and the theoretical one are vey similar. Additionally, the simulated data has a distribution that resembles a normal distrib (blue trace), presenting a great overlap with the normal distribution (red trace).