Statistical Inference Course Project - part 1

Marcelo Tibau

31 de agosto de 2016

Part 1: Simulation Exercise

Overview

In this project I investigate the exponential distribution in R and compare it with the Central Limit Theorem. I also provide the explanations and codes to assure the reproducibility of the investigation.

This report is based at the distribution of averages of 40 exponentials and a thousand simulations.

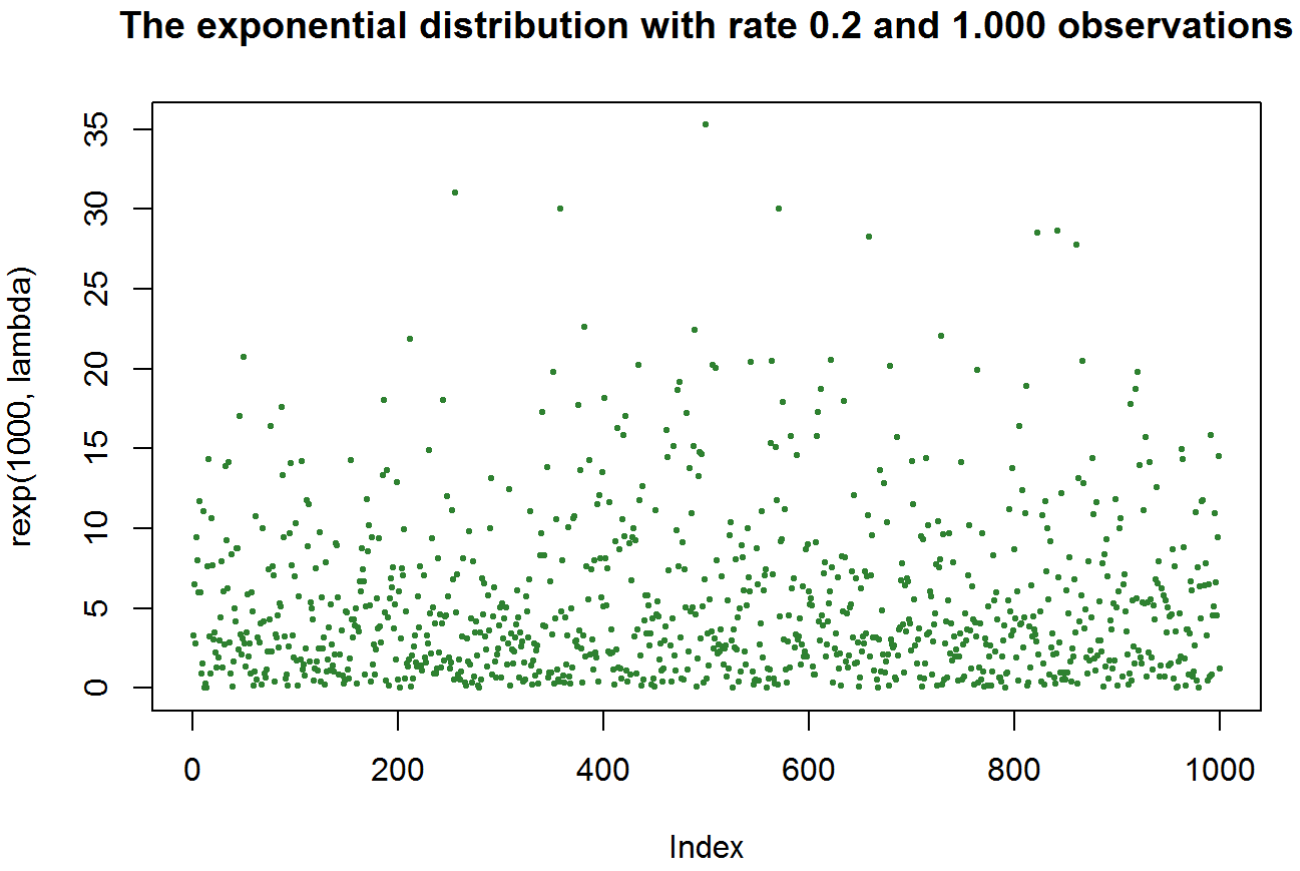
Simulation

Codes to set the chosen parameters. The parameters for lambda, number of exponentials and simulations were instructed.

```
set.seed(2016)
lambda <- .2
exponumber <- 40
simunumber <- 1000
data_simulation <- matrix(rexp(simunumber*exponumber, rate = lambda), simunumber, exponumber)
```

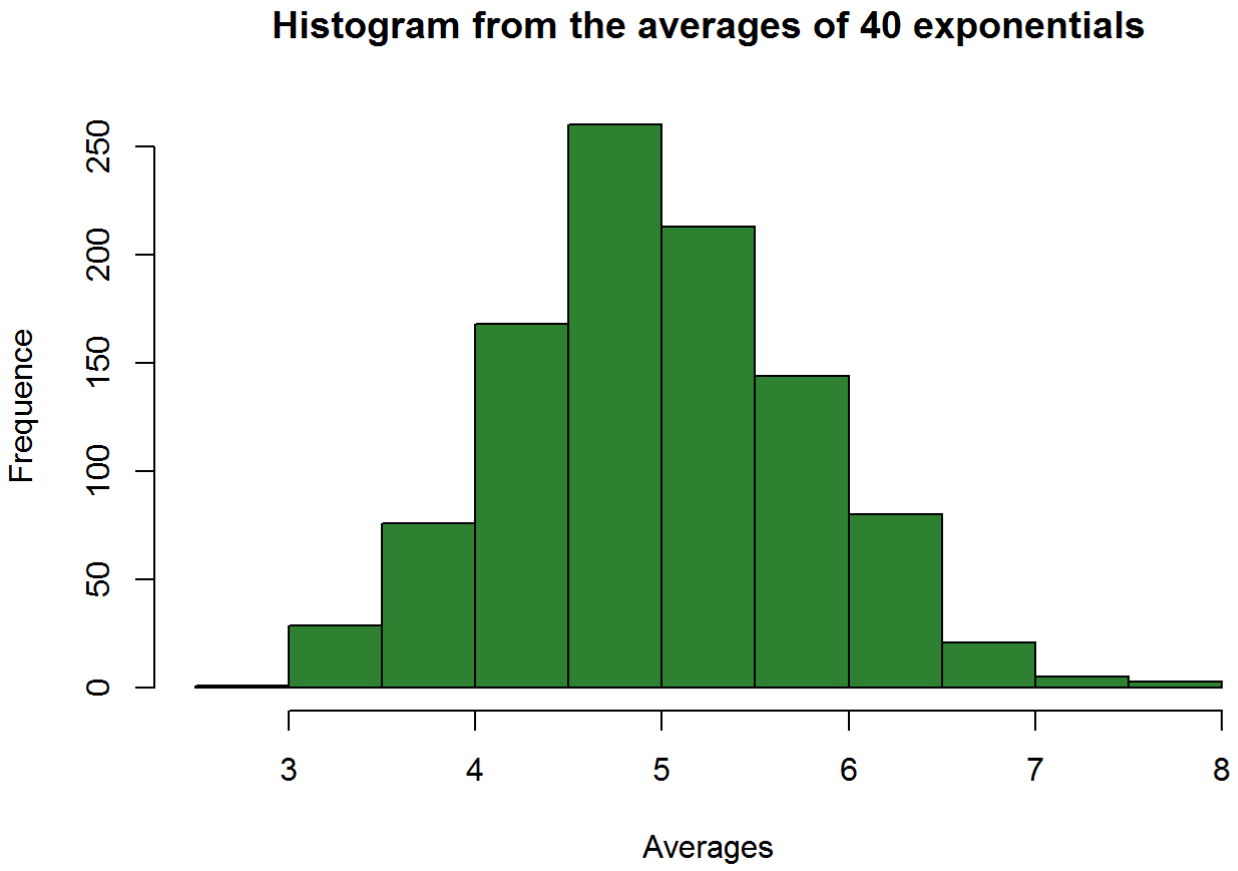
Code to plot the exponential distribution:

```
plot(rexp(1000, lambda), pch = 20, cex = .6, main = "The exponential distribution with rate 0.2
 and 1.000 observations",  col = "#2e8130")
```



Code to plot the averages of 40 exponentials in a histogram:

```
averages <- rowMeans(data_simulation)
hist(averages, xlab = "Averages", ylab = "Frequence", main = "Histogram from the averages of 40
 exponentials", col = "#2e8130")
```



1. Sample Mean versus Theoretical Mean

Codes to calculate the sample mean and the theoretical mean of the distribution and to plot the comparison:

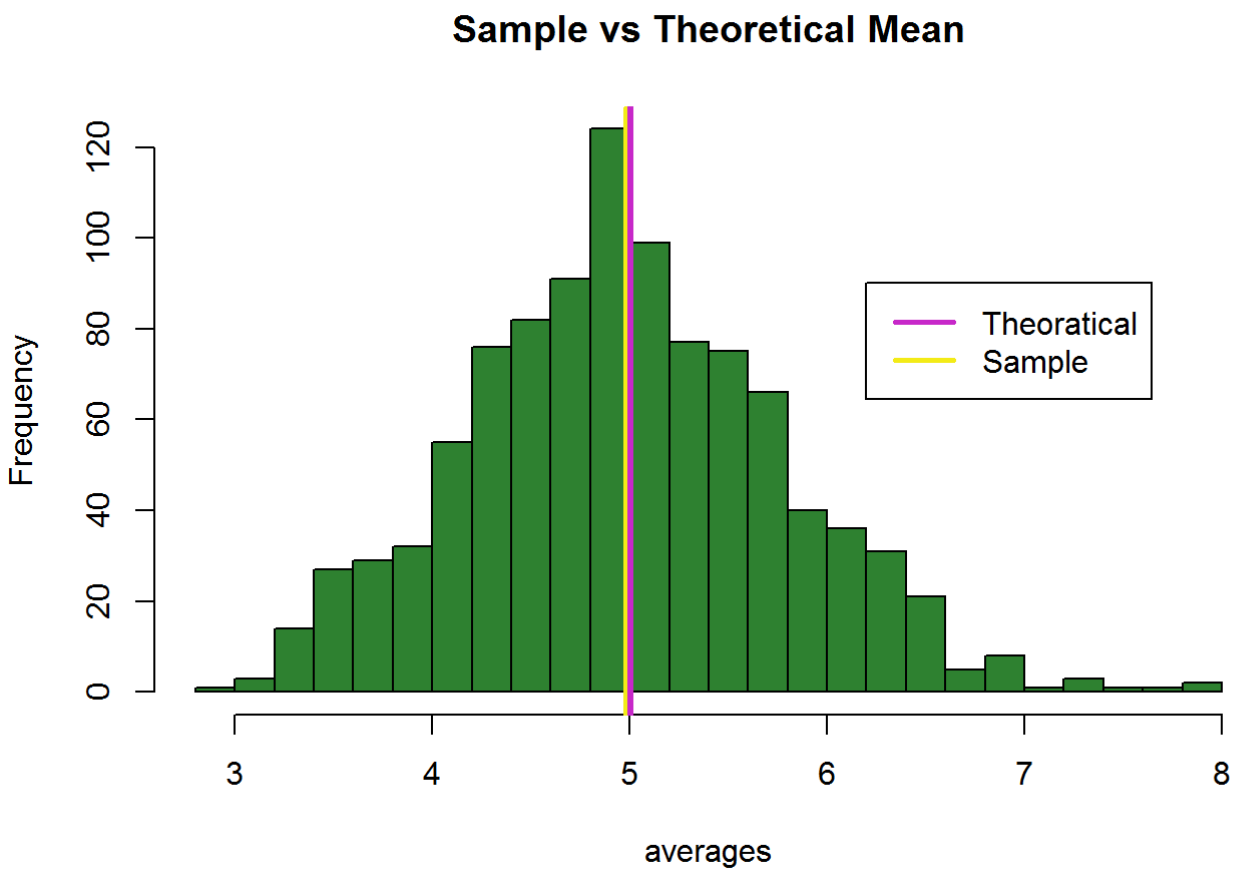
```
sample_mean <- mean(averages)
cat("Sample Mean: ", sample_mean)
```

```
## Sample Mean:  4.979186
```

```
theo_mean <- 1/lambda
cat("Theoretical Mean: ", theo_mean)
```

```
## Theoretical Mean:  5
```

```
hist(averages, col="#2e8130", main="Sample vs Theoretical Mean", breaks=20)
l1 <- abline(v=mean(theo_mean), lwd="4", col="#c728c9")
l2 <- abline(v=mean(sample_mean), lwd="2", col="#f3eb17")
legend(6.2, 90, c("Theoratical", "Sample"), lty = c(1, 1), lwd = c(2.5, 2.5), col =
c("#c728c9", "#f3eb17"))
```



2. Sample Variance versus Theoretical Variance

Codes to calculate the variances (both sample and theoretical):

```
sample_var <- var(averages)
cat("Sample Variance: ", sample_var)
```

```
## Sample Variance:  0.6379013
```

```
theo_var <- (1/lambda)^2/exponumber
cat("Theoratical Variance: ", theo_var)
```

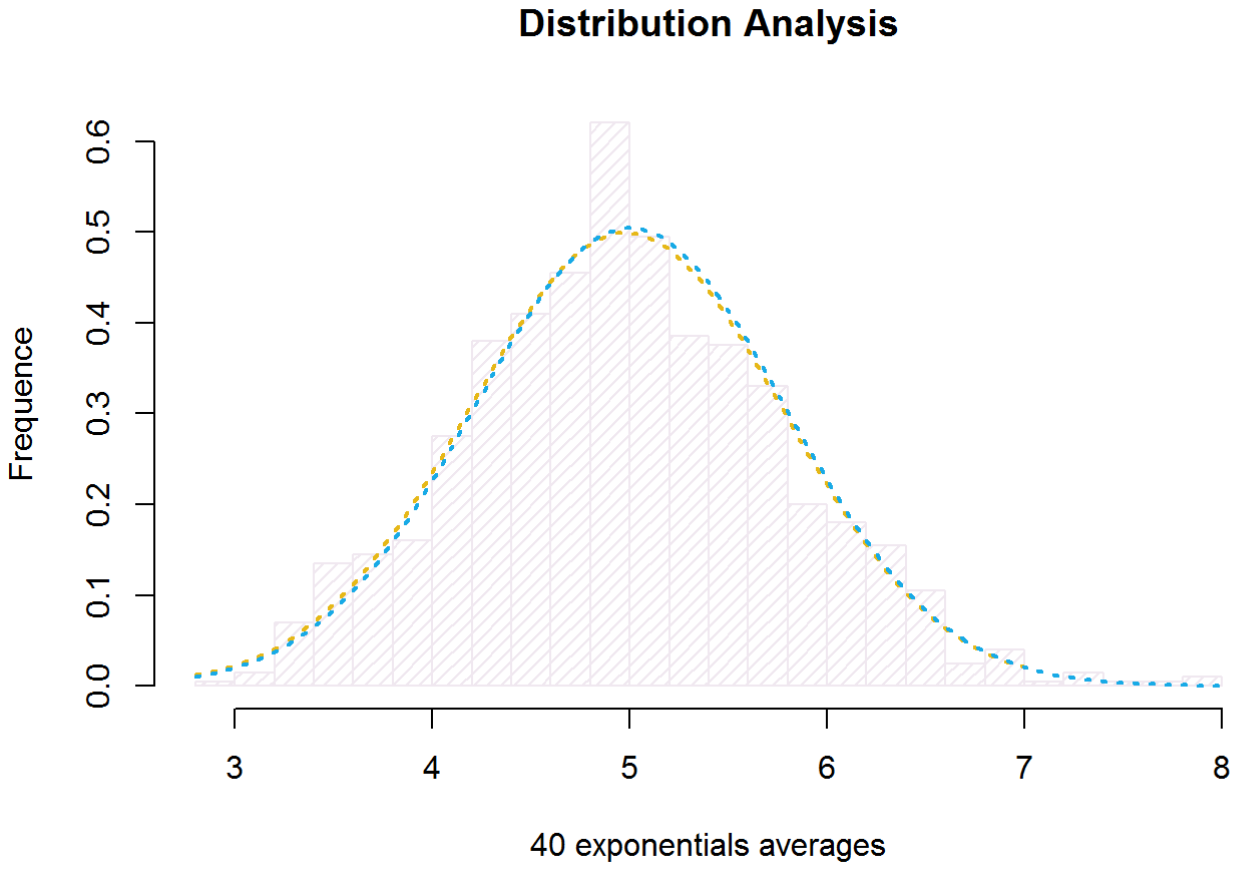
```
## Theoratical Variance:  0.625
```

As we can see above, the variances are very close.

3. Distribution analysis

Code to create an approximate normal distribution and aligns with the the sample:

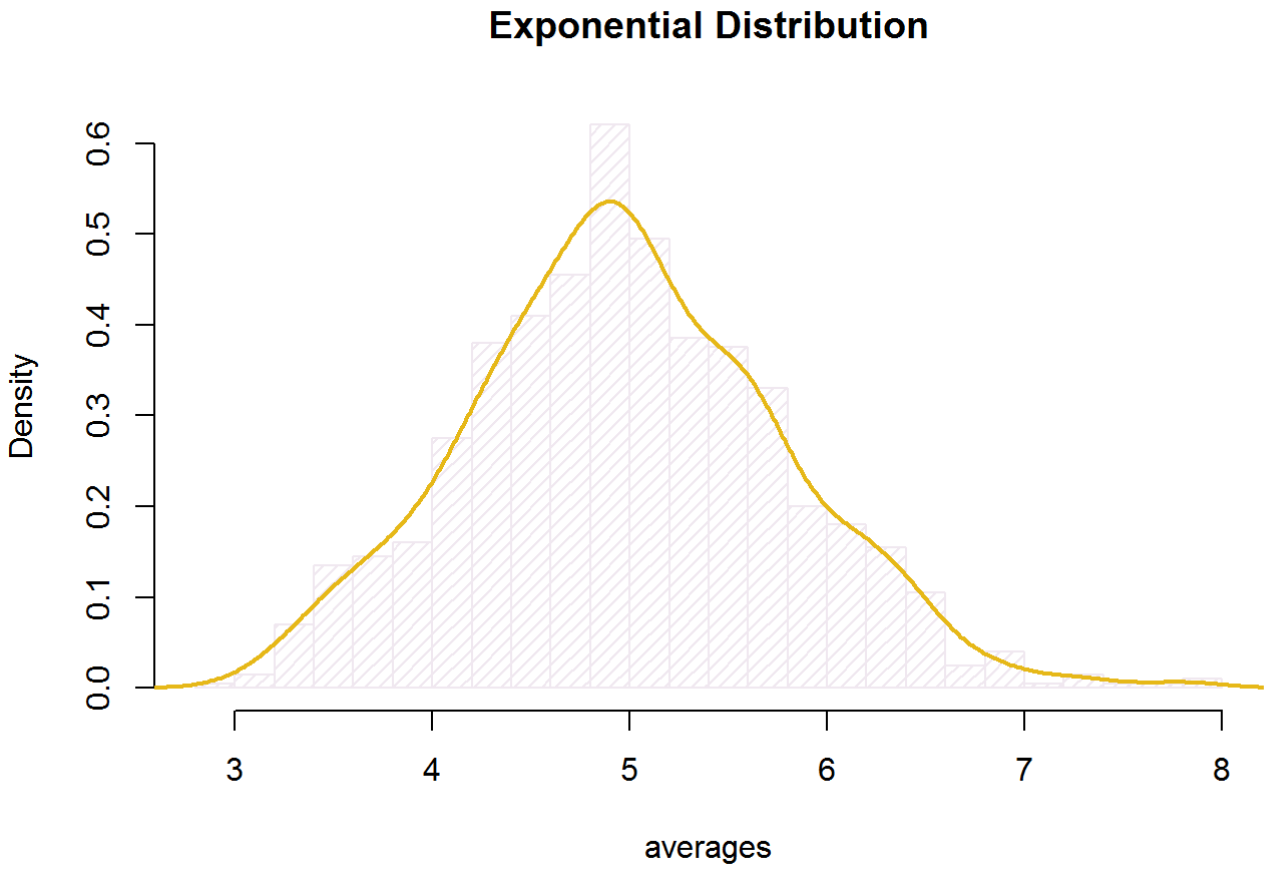
```
hist(averages, density = 20, breaks = 20, prob = TRUE, xlab = "40 exponentials averages", ylab
= "Frequence", main = "Distribution Analysis", col = "#f0e8f0")
curve(dnorm(x, mean = sample_mean, sd = sqrt(sample_var)), col = "#e6b818", lwd=2, lty = "dotte
d", add = TRUE, yaxt="n")
curve(dnorm(x, mean = theo_mean, sd = sqrt(theo_var)), col = "#18aae6", lwd=2, lty = "dotted",
add = TRUE, yaxt="n")
```



As seen above, the histogram can be approximated with the normal distribution. To be really sure, I propose to investigate if the exponential distribution is approximately normal. Due to the Central Limit Theorem, the averages of the sample mean should follow a normal distribution.

Code to plot the exponential distribution:

```
hist(averages, prob = TRUE, col = "#f0e8f0", main = "Exponential Distribution", density = 20, b
reaks = 20)
lines(density(averages), lwd = 2, col = "#e6b818")
```



The Distribution plot and the Exponential Distribution plot are similar. Therefore, I can infer that the distribution is approximately normal.