Statistical Inference Course Project

Masha Maziuk

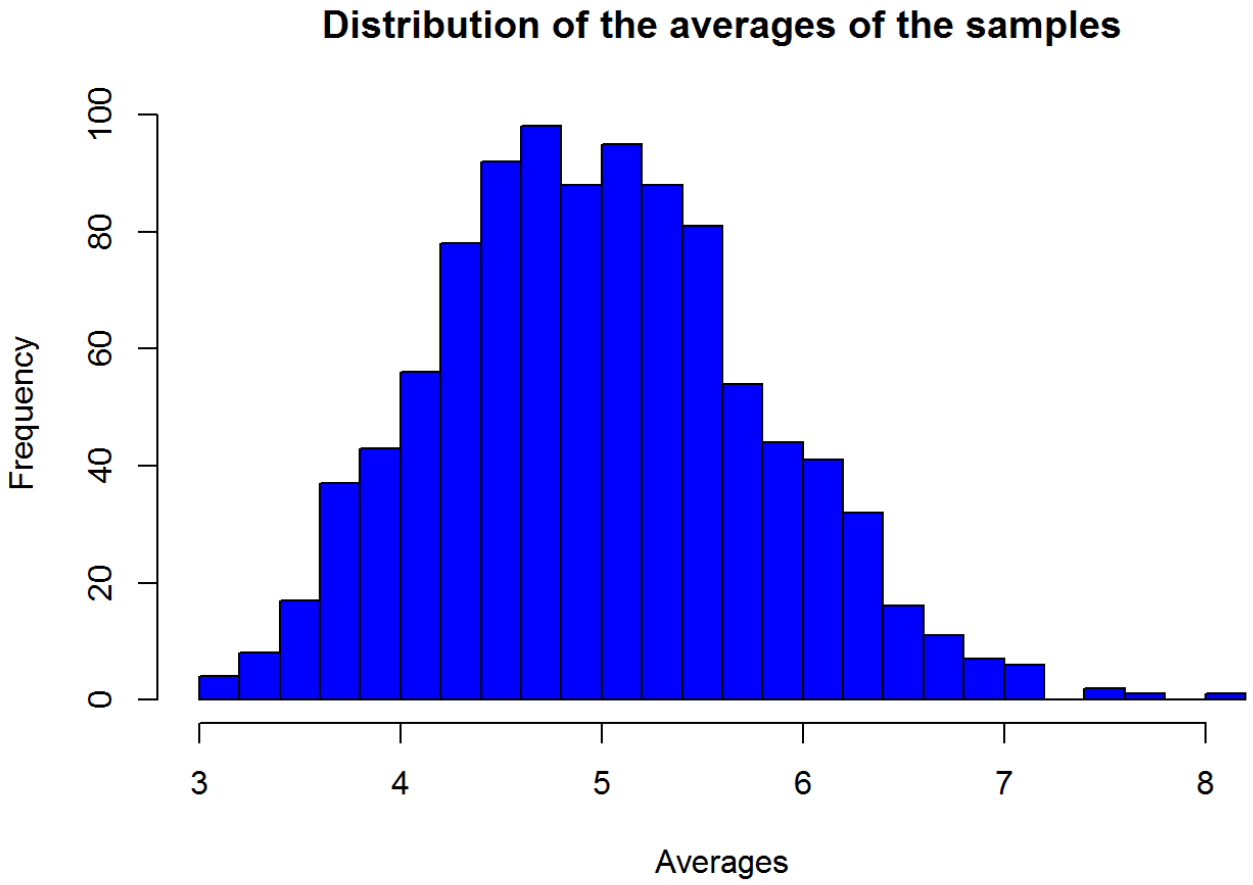
Overview

In this project I am investigating the exponential distribution in R and comparing it with the Central Limit Theorem. The exponential distribution is simulated 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`. I'm setting `lambda = 0.2` for all of the simulations. I am investigating the distribution of averages of 40 exponentials. I'm going to do a thousand simulations.

Simulations

```
# Loading libraries
library(ggplot2)

# Assigning given values
lambda <- 0.2
n <- 40
N.sim <- 1000
mean <- 1/lambda
sd <- 1/lambda

# Setting seed
set.seed(40)

# Creating a matrix of simmulated values with N.sim rows and n columns
sim.matrix <- matrix(rexp(N.sim * n, rate = lambda), N.sim, n)

# Calculating means accross the rows, i.e. averages of the samples:
sim.avg <- rowMeans(sim.matrix)
```

Plotting the distribution of the averages of the samples:



Sample mean vs Theoretical mean:

```
# Calculating the sample mean
smean <- mean(sim.avg)
round.smean <- round(smean, 2)
```

The sample mean is 4.99

The theoretical mean is `1/lambda = 1/0.2 = 5`

The theoretical mean and the sample mean are very close to each other.

Sample variance vs Theoretical variance

```
# Calculating the sample variance
svar <- var(sim.avg)
round.svar <- round(svar, 3)
```

The sample variance is 0.643

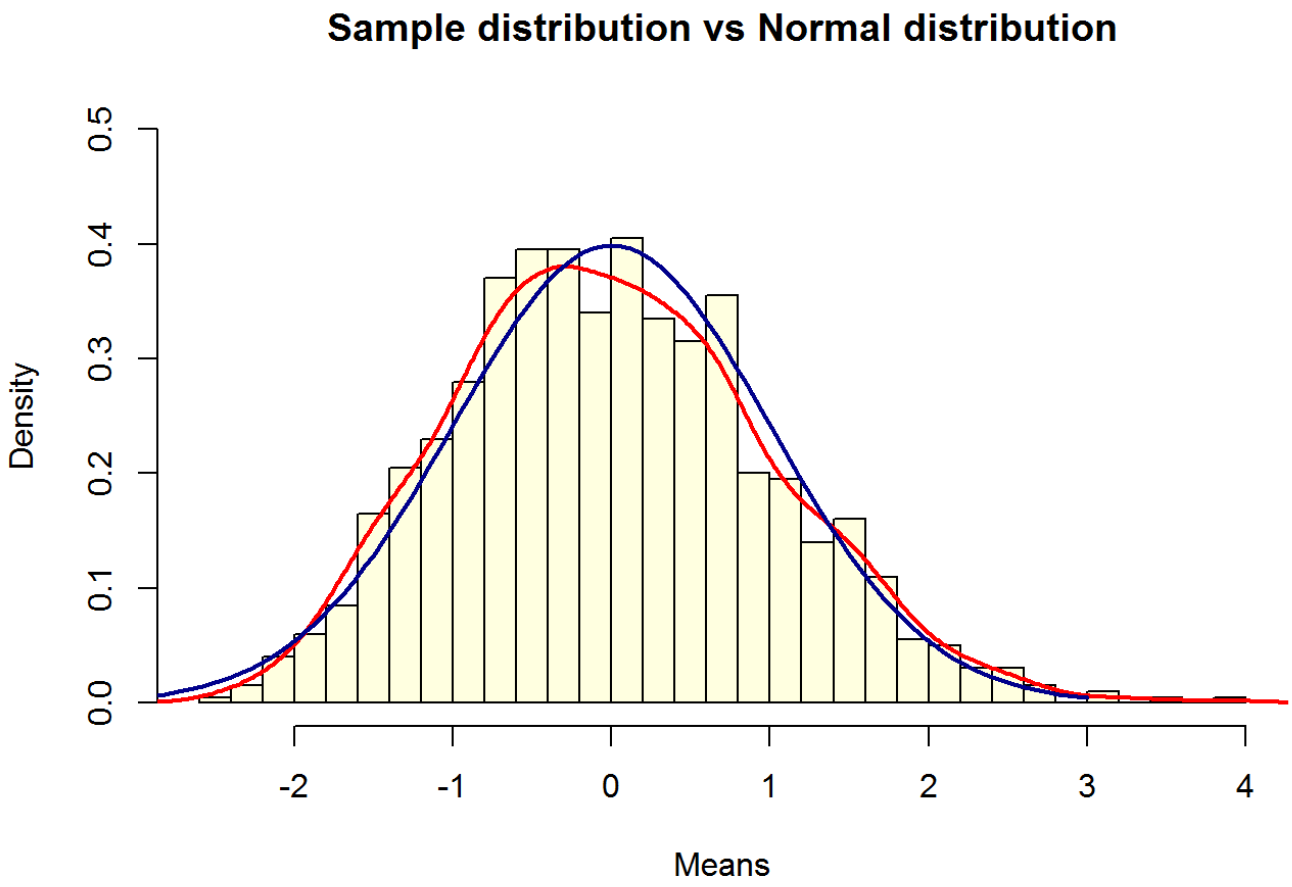
The theoretical variance is `(1/lambda)^2/n = 0.625`

The theoretical variance and the sample variance are still quite close to each other.

Normal distribution

1. Comparing to the sample data to a normal distribution on a plot:

```
scale.sim.avg <- scale(sim.avg)
hist(scale.sim.avg, prob=TRUE, main="Sample distribution vs Normal distribution", xlab = "Means", ylim=c(0, 0
.5), breaks = 30, col = "lightyellow")
lines(density(scale.sim.avg), lwd = 2, col = "red")
curve(dnorm(x,0,1), -3, 3, col="darkblue", lwd=2, add=TRUE)
```



From the plot we can see, that the distribution of the sample averages is approximately normal, as its curve lies very close the the normal distribution curve.