Statistical Inference Course Project

Masha Maziuk

Overview

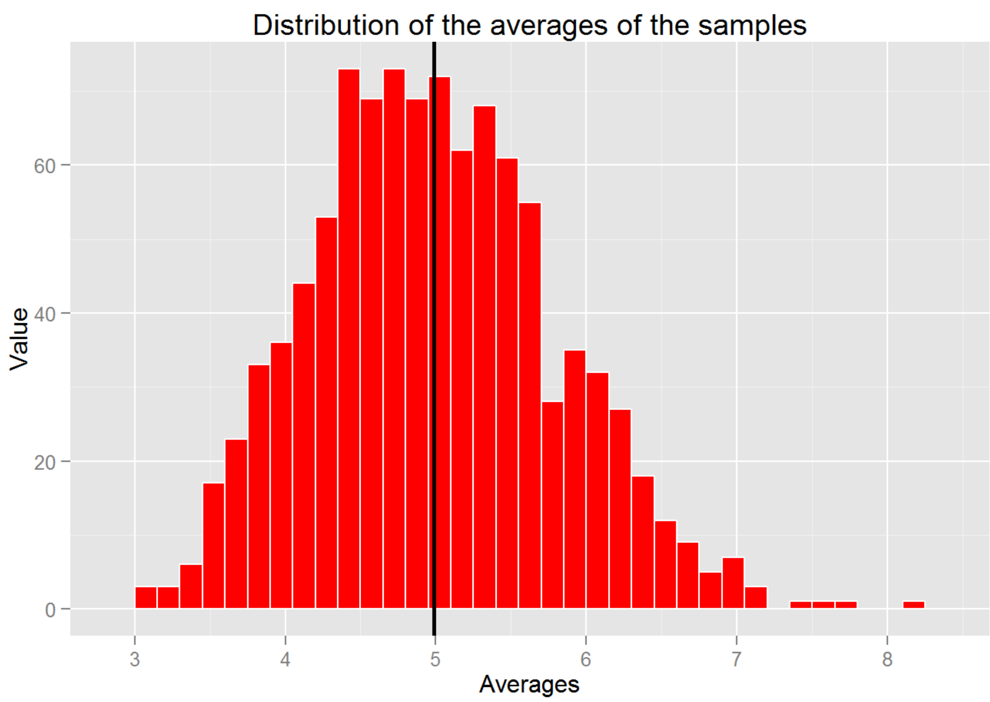
In this project I am investigating the exponential distribution in R and comparing it with the Central Limit Theorem. The exponential distribution is simulated with `rexp(n, lambda)` where `lambda` is the rate parameter, which is set to `0.2` for all the simulations. The mean of exponential distribution is `1/lambda` and the standard deviation is also `1/lambda`. I am investigating the distribution of averages of `40` exponentials, and I'm going to do `1000` simulations.

Simulations

```
# 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 using rexp function. Each row is a samp
le of size 40. The number of rows (or samples) is 1000.
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:

```
library(ggplot2)
#Creating a histogram for the distribution of the averages of the samples with a vertical line of the mean of
the averages.
ggplot(data = data.frame(Averages = sim.avg), aes(Averages)) +
  geom_histogram(col = "white", fill = "red", binwidth = 0.15) +
  labs(title = "Distribution of the averages of the samples", x = "Averages", y = "Value") +
  geom_vline(xintercept = mean(sim.avg), size = 1)
```



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

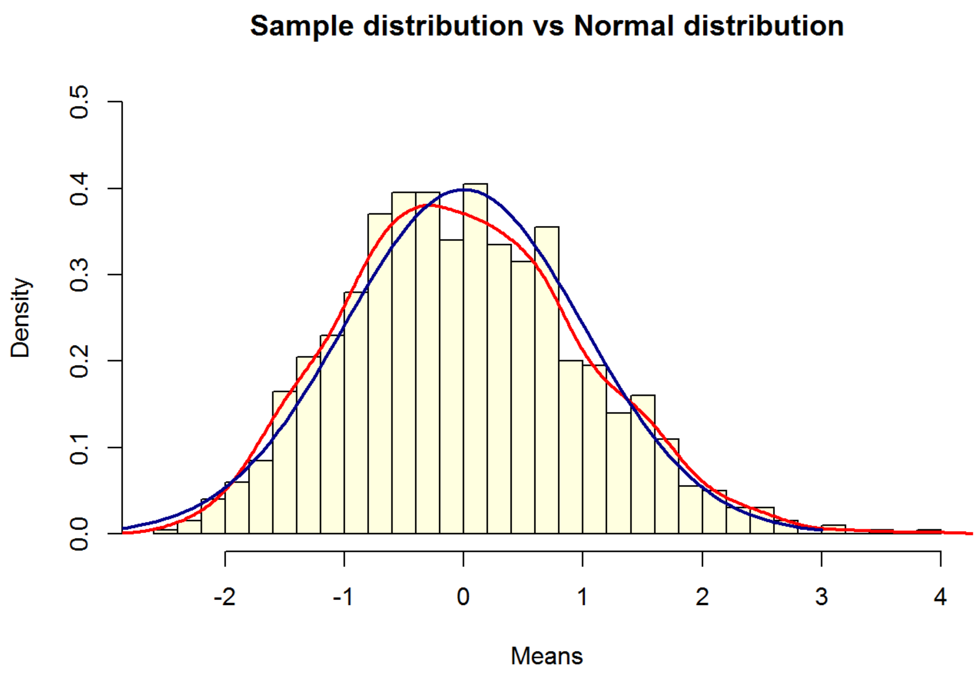
The theoretical variance is `(1/lambda)^2/n = 0.625`

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

Approximately normal distribution

The Central Limit Theorem states that if you have a population with mean mu and standard deviation sigma and take sufficiently large random samples from the population, then the distribution of the sample means will be approximately normally distributed. Our samples meet the conditions of the CLT: 1) sampled observations are independent (as they are random and make less the 10% of the entire popupaltion); 2) the sample size is large (more than 30).

The distibution of the mean of sample averages is bell-shaped and very close to the curve of the normal distribution. We can see it from the plot below:



Thus, we can say that the distribution of the sample averages is approximately normal.