Statistical inference project1

YOUNGSANG KIM

2020 12 13

A simulation exercise.

In this project you 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.

we will compare the Central Limit Theorem with the averages of 40 exponential distributions. The parameter lambda of the exponetial distribution is set to 0.2 and the number of iteration is 1000.

```
#Load packages
library(dplyr, warn.conflicts = F)
library(ggplot2)

#Exponential function parameters
lambda <- 0.2
n <- 40
n_sim <- 1000
```

Objectives

The specific objectives of this report are the following:

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Simulation

First, we should make the matrix

```
#set the seed
set.seed(3)

#Create a 1000x40 matrix containing the results of the simulation
sim_dist <- matrix(data=rexp(n * n_sim, lambda), nrow=n_sim)
means <- rowMeans(sim_dist)
```

1. Sample Mean versus Theoretical Mean

The sample mean, is obtained in what follows

```
sample_mean <- mean(means)
theoretical_mean <- 1 / lambda
sample_mean
```

```
## [1] 4.98662
```

```
theoretical_mean
```

```
## [1] 5
```

```
sample_mean - theoretical_mean
```

```
## [1] -0.01338033
```

there is a small difference between then, -0.01338003, to be exact

2. Sample variance versus Theoretical Variance

```
sample_variance <- var(means)
theoretical_variance <- (((1 / lambda)^2) / 40)
sample_variance
```

```
## [1] 0.6257575
```

```
theoretical_variance
```

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

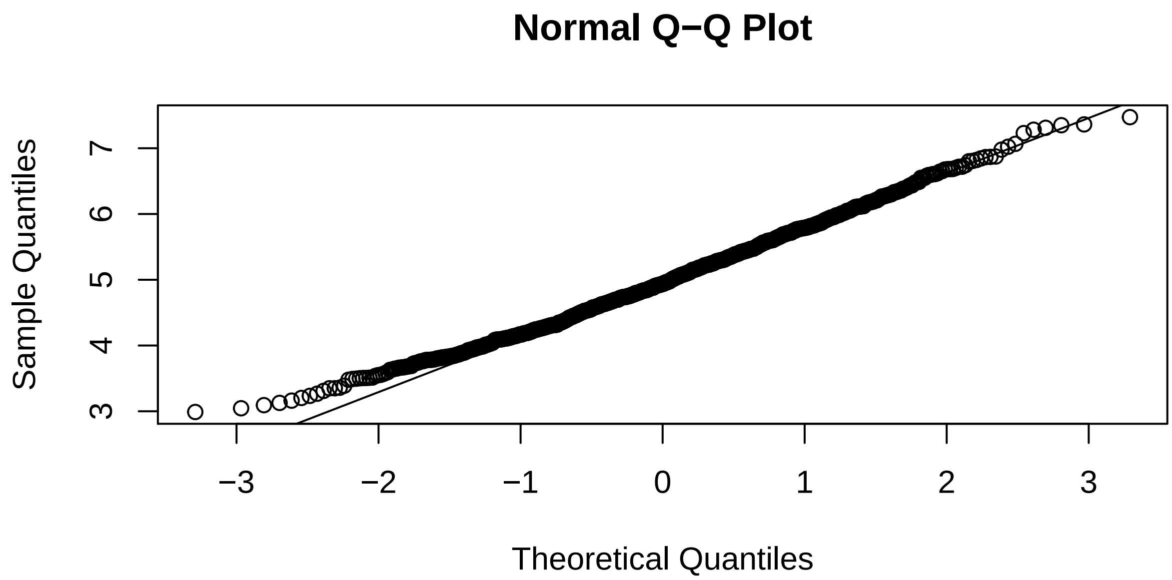
```
sample_variance - theoretical_variance
```

```
## [1] 0.0007575252
```

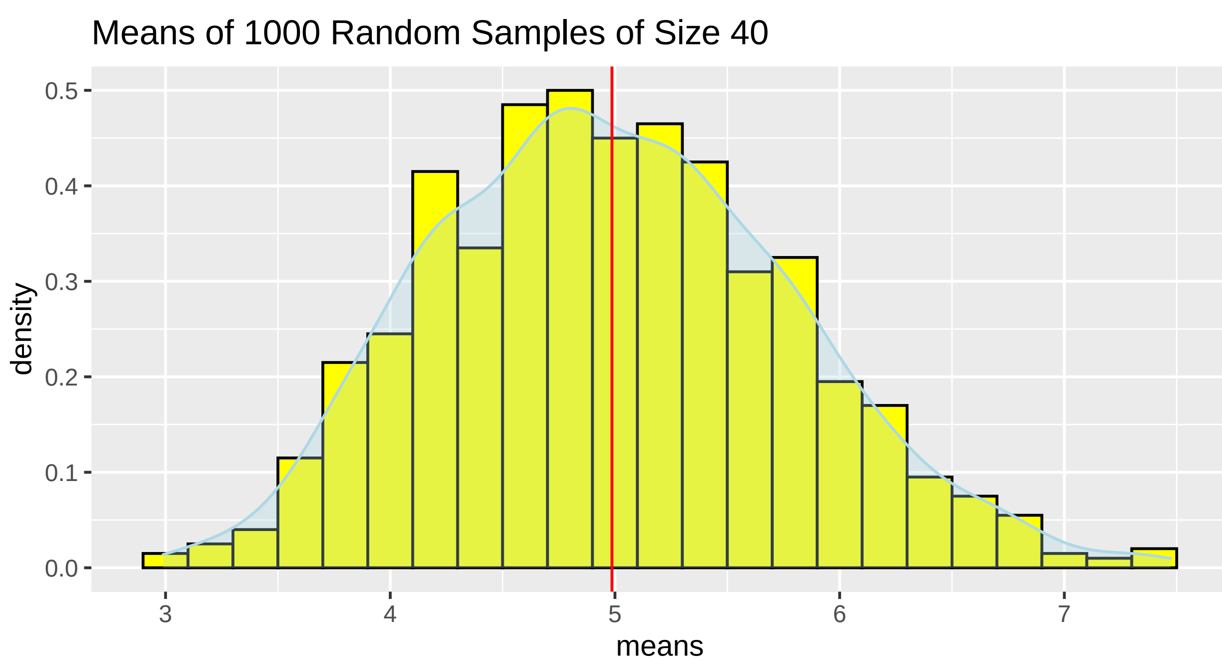
When comparing the estimated calue and the theoretical one, there is a small difference between then, 0.0007575252, to be exact

3. Distribution

```
qqnorm(means, main = "Normal Q-Q Plot")
qqline(means)
```




```r
g <- ggplot(data.frame(means), aes(x=means))
g <- g + geom_histogram(aes(y=..density..), binwidth = lambda, colour = "black", fill = "yellow")
g <- g + geom_density(colour="lightblue", fill="lightblue", alpha = .3)
g <- g + geom_vline(aes(xintercept = sample_mean), size = .5, colour = "red")
g <- g + ggtitle("Means of 1000 Random Samples of Size 40")
g
```



we can see that the distribution is approximately normal