Statistical Interference Course Project. Part 1

Daria Kravets

Overview

This project involves investigating the properties of the exponential distribution in R and comparing it with the Central Limit Theorem. The exponential distribution can be simulated in R using the rexp(n, lambda) function, where lambda is the rate parameter. For this project, lambda will be set at 0.2 for all simulations.

The aim is to investigate the distribution of averages of 40 exponentials by conducting a thousand simulations. Specifically, the project demonstrates a difference between samle characteristics, such as mean and variance, and theoretical values of it.

Finally, the normality of the distribution will be tested using appropriate methods such as a histogram or QQ plot to demonstrate that the distribution is approximately normal.

Sample simulation

First, for having a stable random sample we will set the seed:

```
set.seed(5327)
```

Then we need to set some values needed for our project

```
lambda <- 0.2
n <- 40
size <- 1000
```

And finally we can set our sample by producing random values:

```
simulated_sample = NULL
for (i in 1 : size) simulated_sample = c(simulated_sample, mean(rexp(n, lambda)))
```

Sample mean and Theoretical mean comparisson

Calculating mean from the simulations

```
sample_mean <- mean(simulated_sample)
sample_mean
```

```
## [1] 4.97759
```

Calculating theoretical mean using formula

```
theor_mean <- lambda^(-1)
theor_mean
```

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

Comparing of this two values by calculating their difference

```
mean_dif <- abs(sample_mean - theor_mean)
mean_dif
```

```
## [1] 0.02241024
```

From this calculations, it is clear that our result has only slight difference from a theoretical value.

Sample variance and Theoretical variance comparisson

Calculating variance from the simulations

```
sample_var <-var(simulated_sample)
sample_var
```

```
## [1] 0.6452081
```

Calculating theoretical variance using formula

```
theor_var <- 1 / (lambda^2 * n)
theor_var
```

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

Comparing of this two values by calculating their difference

```
var_dif <- abs(sample_var - theor_var)
var_dif
```

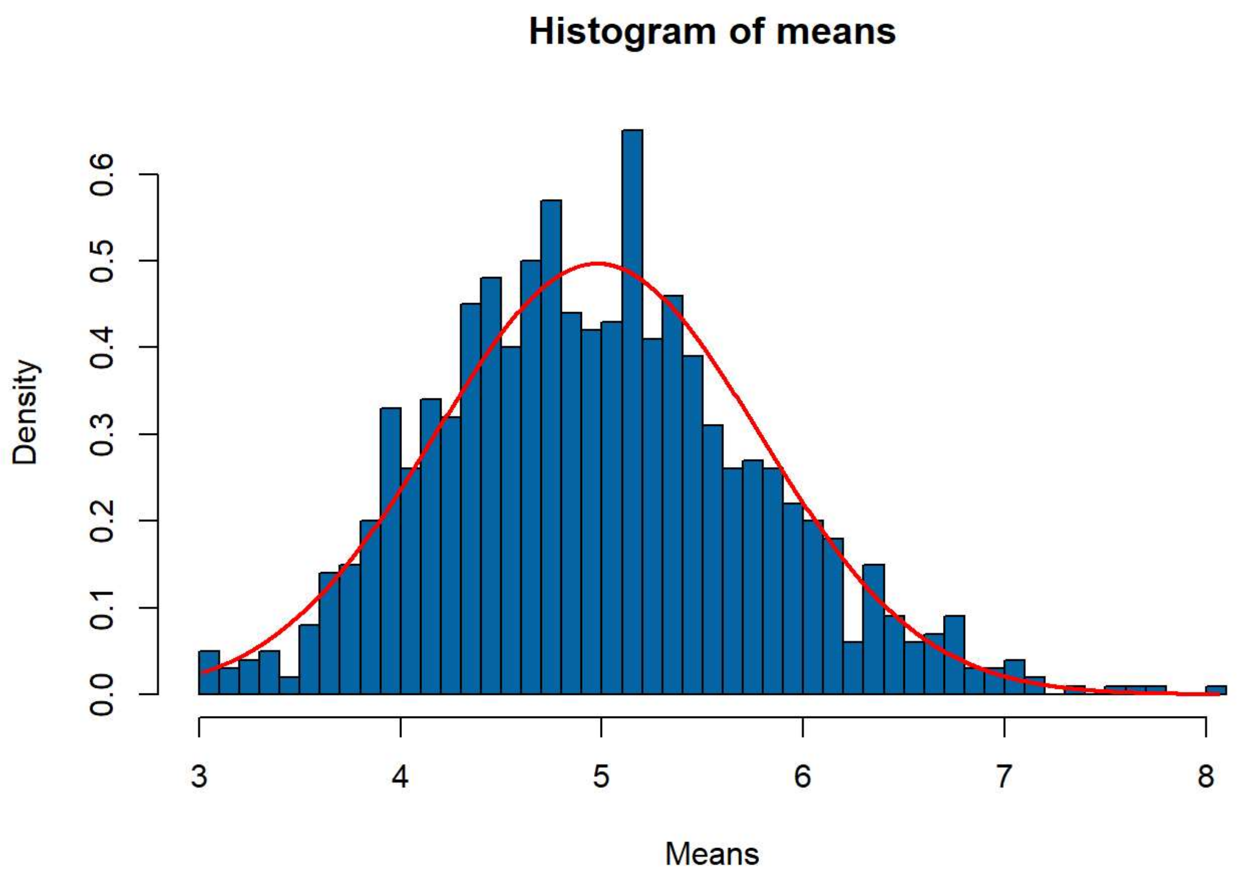
```
## [1] 0.02020813
```

From this calculations, it is clear again that our result has only slight difference from a theoretical value.

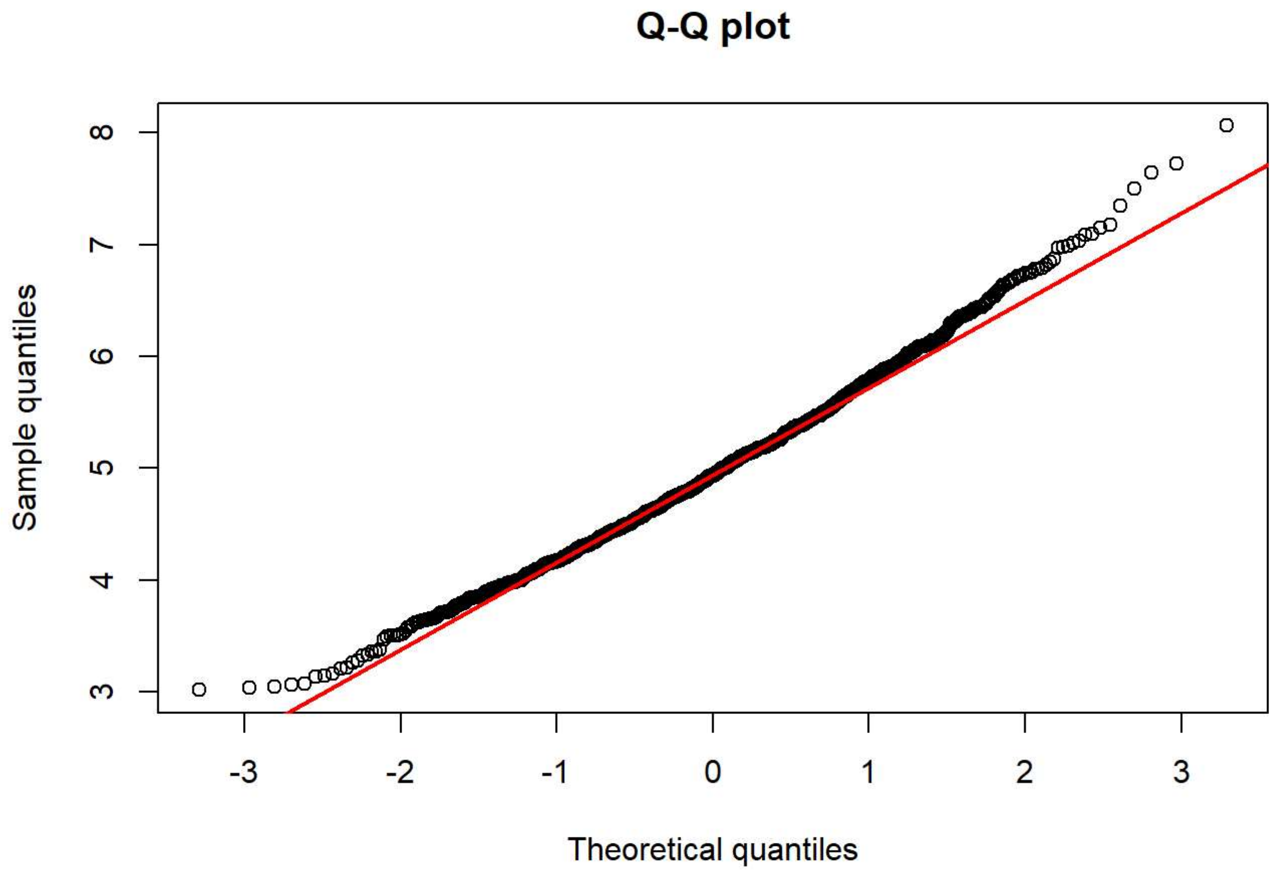
Distribution

By plotting a histogram and a normailty Q-Q plot, we can conclude whether the distribution is normal.

```
xx <- seq(min(simulated_sample), max(simulated_sample), length=1000)
hist(simulated_sample, breaks = n, freq=F, col='#0665a5', xlab='Means', ylab='Density', main
='Histogram of means')
lines(xx, dnorm(xx, mean=sample_mean, sd=sd(simulated_sample)), col='red', lwd=2)
```




```
qqnorm(simulated_sample, xlab='Theoretical quantiles', ylab='Sample quantiles', main = 'Q-Q p
lot')
qqline(simulated_sample, col='red', lwd=2)
```



Both plots are showing that the distribution is approximately normal.