Statistical Inference-Exponential Distribution vs CLT

Devara Izaz Fathan

8/22/2020

1. Overview

This assignment evaluates the exponential distribution versus Central Limit Theorem. We will use rexp function to do the simulation. The values used in this asignment are lambda= 0.2 and numberofsimulation=1000. We are going to do simulation, comparing sample mean vs theoritical mean, comparing sample variance versus theoritical variance and its distribution.

2. Simulation

The exponential distribution can be simulated in R 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. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. We will do 1000 simulation.

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## -- Attaching packages ------------------------------------------------------------------------------

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v stringr 1.4.0
## v tidyr   1.1.1     v forcats 0.5.0
## v readr   1.3.1

## -- Conflicts ---------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
lambda <- 0.2
n <- 40 #number of sample
B <- 1000 #number of simulation
set.seed(8)
## do the simulation
data_sample <- data.frame(matrix(rexp(n*B, lambda), nrow = B, ncol = n))
data_sample <- data_sample %>% mutate(mean_sample = rowMeans(data_sample))
```

3. Sample Mean VS Theoritical Mean

Based on theory, the mean for the theoritical data is 1/lambda. Here is the code:

```
theoritical_mean <- 1/lambda
print(theoritical_mean)
```

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

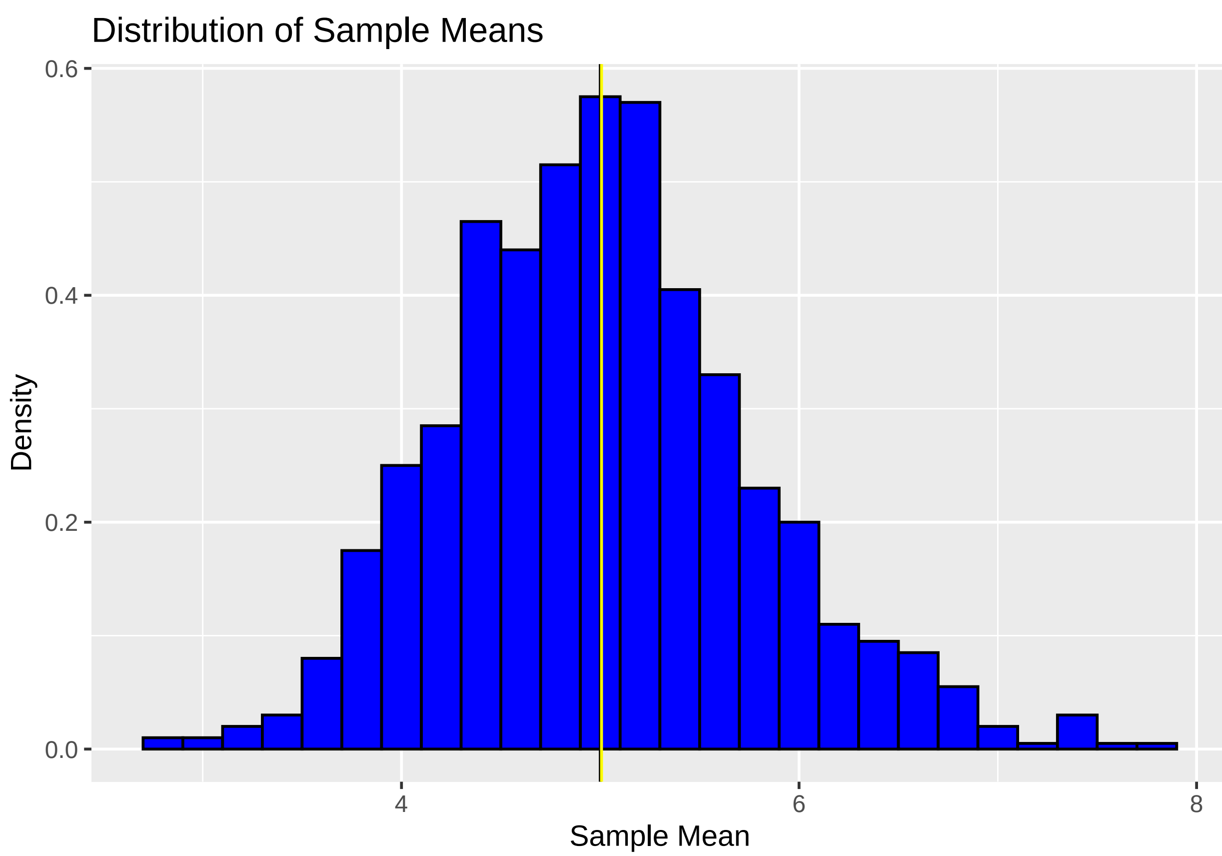
Then we compute the sample mean with data_sample dataset.

```
sample_mean <- mean(data_sample$mean_sample)
print(sample_mean)
```

```
## [1] 5.006442
```

Based on those two codes above, we have got the the theoretical mean = 5, and the sample mean = 5.018281. Those values are very close enough to each other.

```
data_sample %>% ggplot(aes(mean_sample, ..density..)) +
      geom_histogram(binwidth = 0.2, position = 'identity', col= 'black', fill='blue') +
      xlab('Sample Mean') +
      ylab('Density') +
      ggtitle('Distribution of Sample Means') +
      geom_vline(xintercept = theoritical_mean, color='black') +
      geom_vline(xintercept = sample_mean, color= 'yellow')
```



4. Sample Variance versus Theoretical Variance

Based on theory, the variance for the theoretical data is sdˆ2/n. Here is the code:

```
theoritical_variance <- (1/lambda)^2/n
print(theoritical_variance)
```

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

Then we compute the sample variance using data_sample dataset that we have been created.

```
sample_variance <- var(data_sample$mean_sample)
print(sample_variance)
```

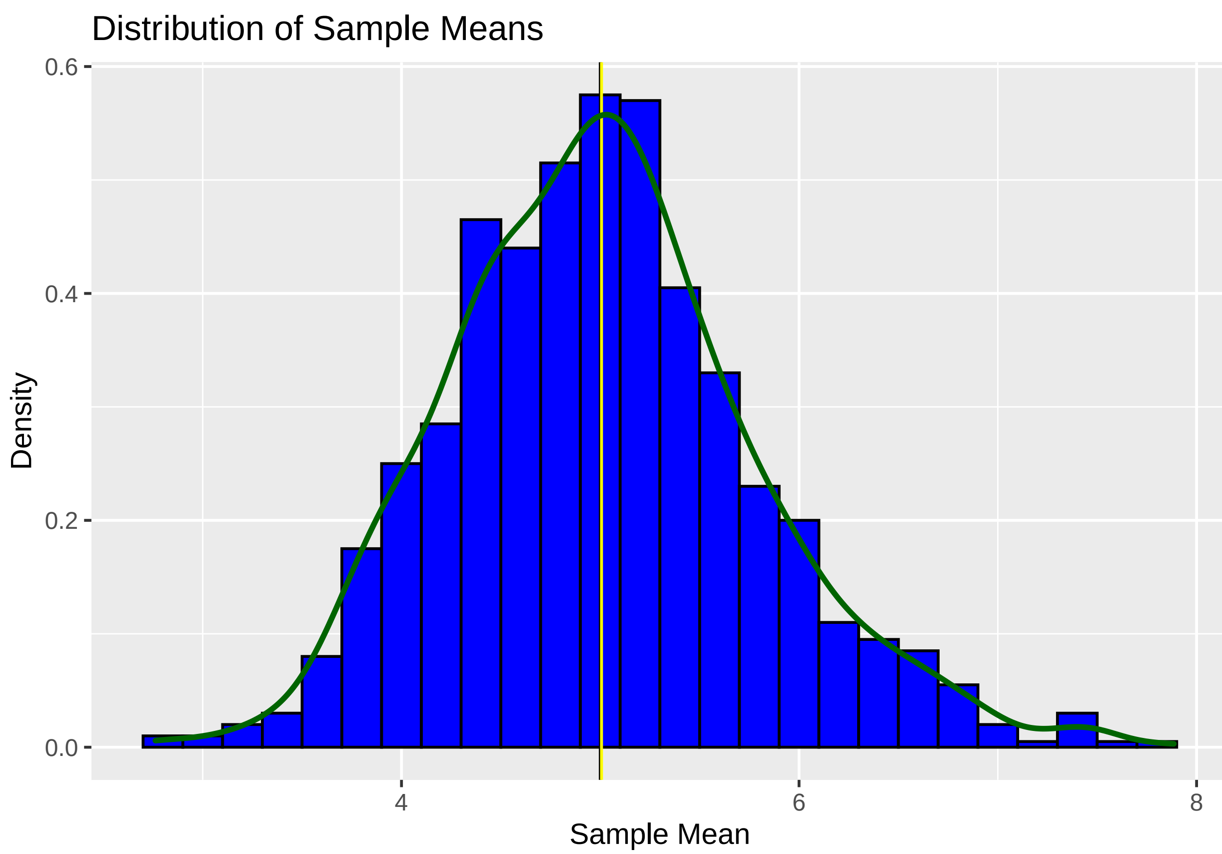
```
## [1] 0.5908158
```

Based on those two codes above, we have got the theoretical variance = 0.625, and the sample variance = 0.636219

5. Distribution

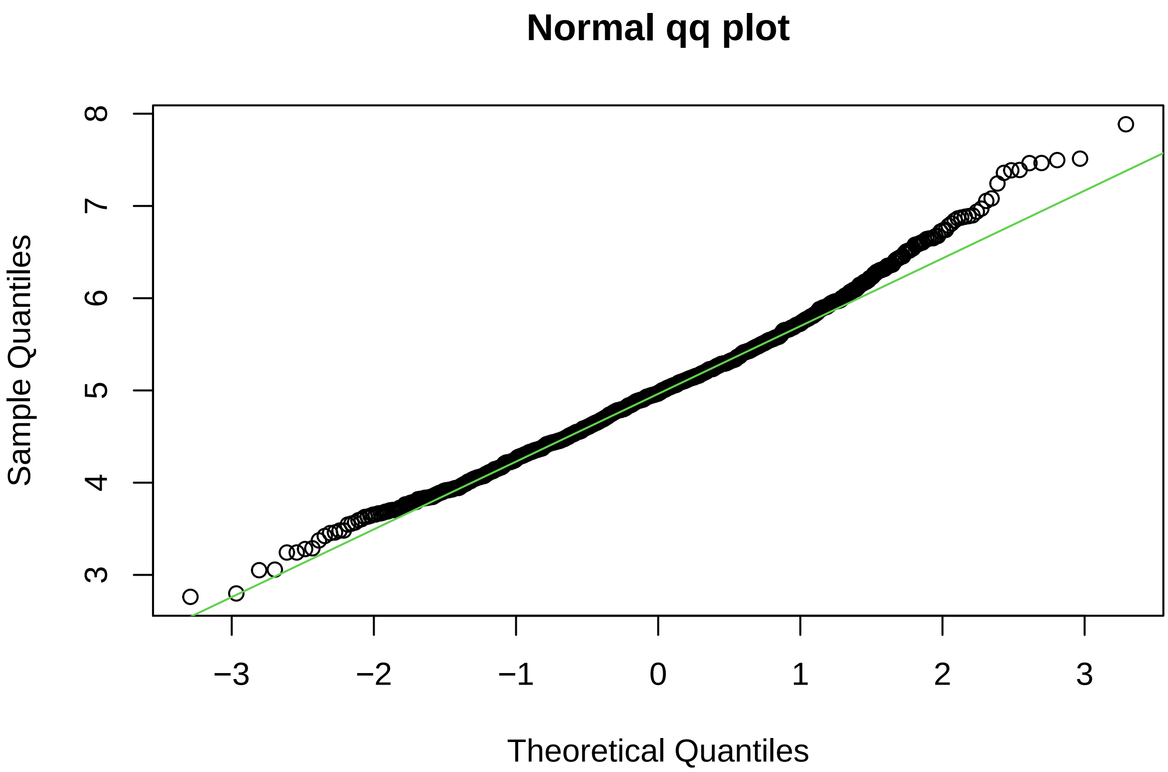
Based on Central Limit Theorem, the distribution of the sample mean should follow this theorem. We will see the density plot here

```
data_sample %>% ggplot(aes(mean_sample, ..density..)) +
      geom_histogram(binwidth = 0.2, position = 'identity', col= 'black', fill='blue') +
      xlab('Sample Mean') +
      ylab('Density') +
      ggtitle('Distribution of Sample Means') +
      geom_vline(xintercept = theoritical_mean, color='black') +
      geom_vline(xintercept = sample_mean, color= 'yellow') +
      geom_density(colour='darkgreen', size=1)
```



To give more evidence that the distribution sample mean occurs in Central Limit Theorem, we use qq plot function.

```
mean_sample <- data_sample$mean_sample
qqnorm(mean_sample, main='Normal qq plot');qqline(mean_sample, col='3')
```



This QQ plot show a close relationship between the data sample mean and the theoritical mean. Thats why we can say that Our simulation is approximately a normal distribution.