The Course Project for Statistical Inference_Part 1

Pengfei LI

7/5/2020

Overview

The requirement of this project is separated into two parts: Part 1 require a investigation on the exponential distribution in R and comparasion between the exponential distribution and the Central Limit Theorem

For the exponential distribution, it could be sumulated with rexp(n and lambda), where the lambda is the rate parameter. The mean of exponential distribution is 1/lambda, as well as the standard deviation.

In the requirement, the lambda is set to be 0.2. And the investigation will be on the distribution of averages of 40 exponentials among 1000 simulations.

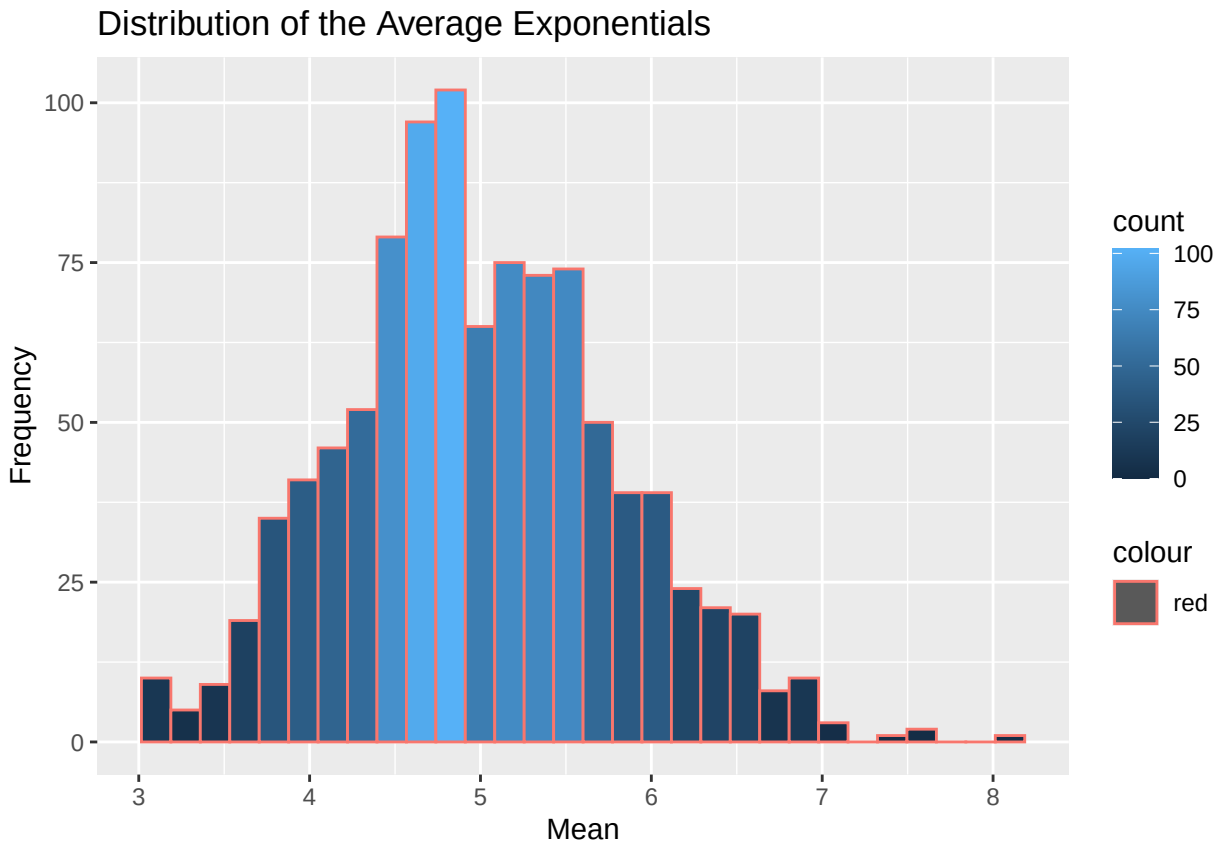
1. Simulations

In the beginning, the general data analysis is performed.

```
##Set the efficient of Lambda
set.seed(100)
lambda <- 0.2
n <- 40
simulations <- data.frame(x=replicate(1000,rexp(n,lambda)))
simulations_mean <- data.frame(x=apply(simulations,2,mean))

library(ggplot2)
h <- ggplot(simulations_mean,aes(x=x))+
        geom_histogram(aes(y=..count..,fill=..count..,color="red"))
h+labs(title = "Distribution of the Average Exponentials",y="Frequency",x="Mean")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



2. Sample Mean versus Theoretical Mean

This section presents the comparasion for sample mean and theoretical mean. Based on the calculation, we could find that sample mean is not equal to theoretical mean.

```
sample_mean <- mean(simulations_mean$x)
theory_mean <- 1/lambda
sample_mean
```

```
## [1] 4.999702
```

```
theory_mean
```

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

```
Difference_mean <- sample_mean-theory_mean
Difference_mean
```

```
## [1] -0.0002980731
```

3. Sample Variance versus Theoretical Variance

This section presents the comparasion for sample variance and theoretical variance Based on the calculation, we could find that sample variance is bigger than theoretical variance.

```
sample_variance <- var(simulations_mean$x)
theory_variance <- ((1/lambda)^2)/n
sample_variance
```

```
## [1] 0.6432442
```

```
theory_variance
```

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

```
Difference_variance <- sample_variance-theory_variance
Difference_variance
```

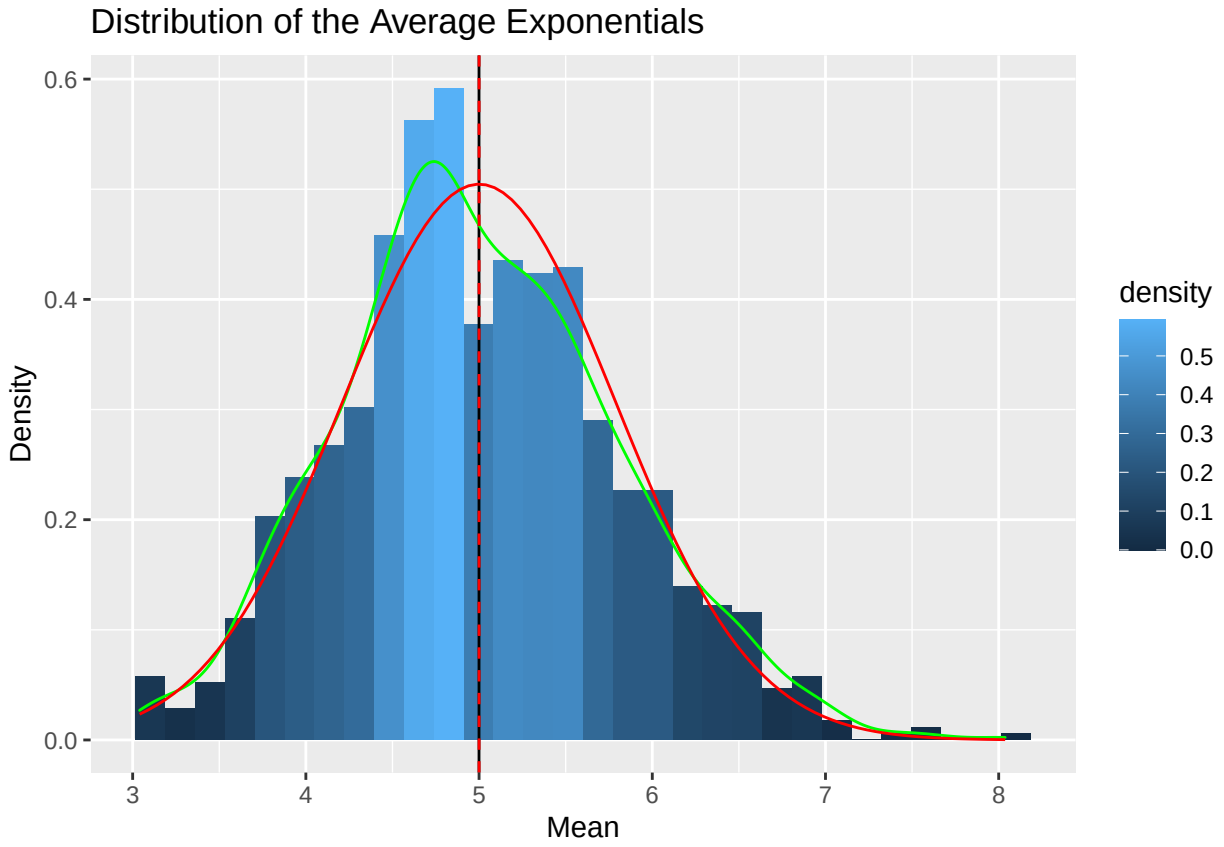
```
## [1] 0.01824422
```

4. Distribution

In this section, the distribution will be plotted into a histogram.

```
histogram_dist <- ggplot(simulations_mean,aes(x=x))+
        geom_histogram(aes(y=..density..,fill=..density..))+labs(title = "Distribution of the Average E
histogram_dist+stat_function(fun=dnorm,args=list( mean=1/lambda, sd=sqrt(theory_variance)),
              color = "red")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Conclusion

From the last histogram,the sample mean is close to the theoretical mean for a normal distribution.