Statistical Inference Assignment

Carey Raychelle

2023-06-19

Assignment Description

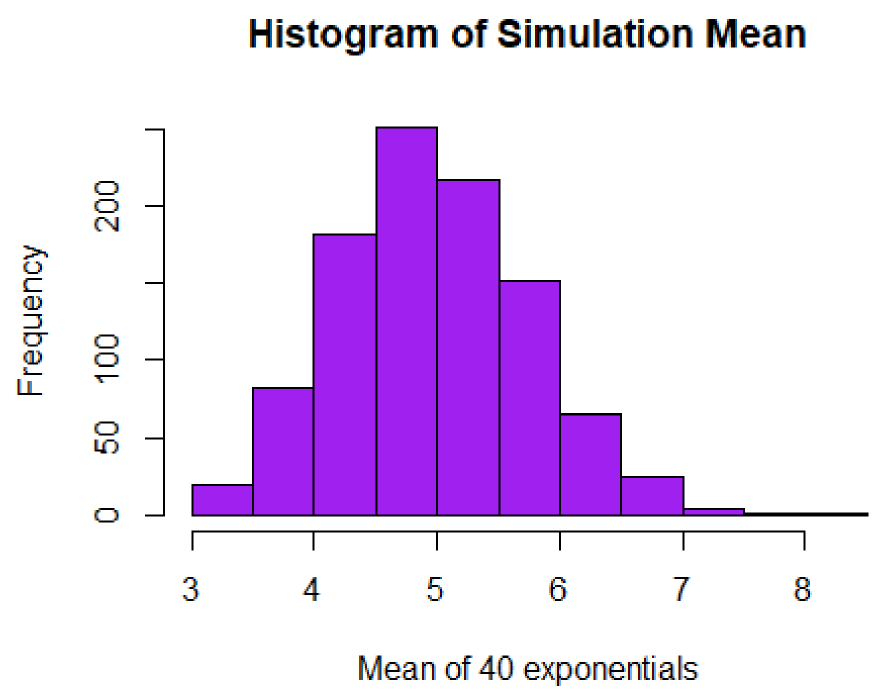
Investigating exponential distribution in R and comparing with the Central Limit Theorem. Exponential distribution in R can be simulated by rexp(n,lamda) where lamda is rate.Mean and standard deviation of distribution is 1/lamda Lamda is set to be 0.2

```
#install the packages needed
library(knitr)

## Warning: package 'knitr' was built under R version 4.2.3

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.2

#making the report reproducible
set.seed(12345)
```

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. 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. Note that you will need to do a thousand simulations.

```
#setting variations for the simulation
n <- 40
lamda <- 0.2
sim <- 1000
#Create a matrix with 1000 rows and columns 40 of random simulations
matsim <- matrix(rexp(sim * n,rate = lamda),sim,n)
#Calculate the means and plot
sim_mean <- rowMeans(matsim)
hist(sim_mean, xlab="Mean of 40 
exponentials",ylab="Frequency",main="Histogram of Simulation 
Mean",col="purple")
```



Sample Mean and Theoretical Mean

```
sample_mean <- mean(sim_mean)
sample_mean

## [1] 4.971972

theoritical_mean <- 1/lamda
theoritical_mean

## [1] 5
```

From the above the sample mean which is 4.97192 is very close to the theoretical mean which is 5

Sample Variance and Theoretical Variance

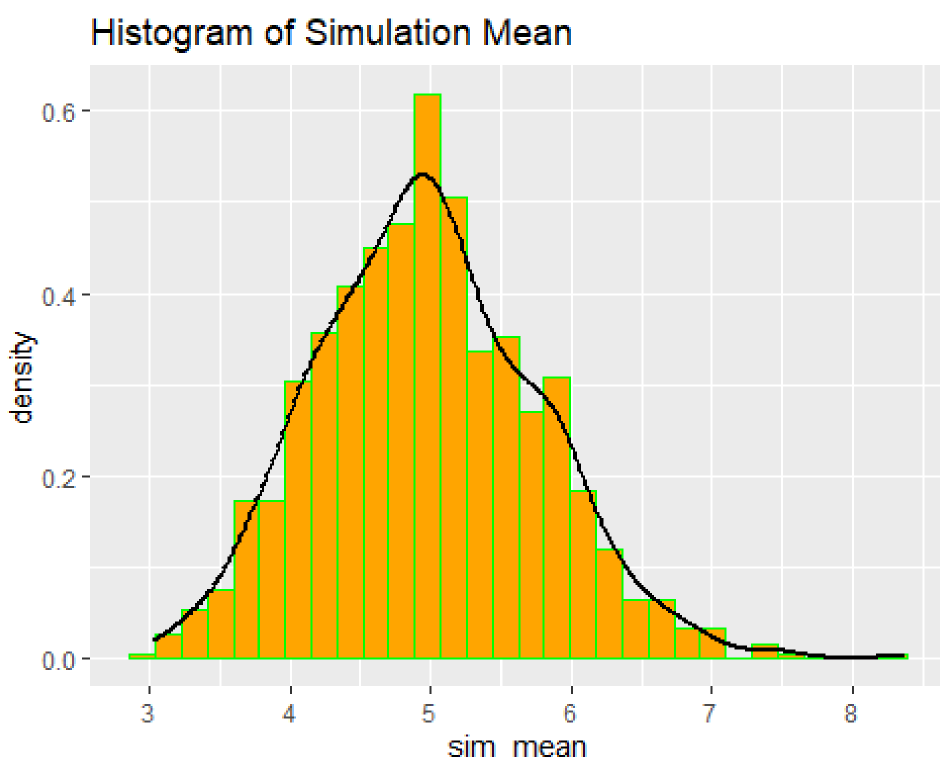
```
sample_var <- var(sim_mean)
sample_var

## [1] 0.6157926

theoretical_var <- (1/lamda)^2/n
theoretical_var

## [1] 0.625
```

The sample variance is also close to the theoretical variance

Approximate Normal Distribution

Creating an approximate normal to see how the sample differs from it

```
plotnorm <- data.frame(sim_mean)
g <- ggplot(plotnorm,aes(x=sim_mean))
g=g+geom_histogram(aes(y=after_stat(density)),colour="green",fill="orange")
g=g+geom_density(colour="black", size=1)

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

g=g+ggtitle("Histogram of Simulation Mean")
g

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



The plot indicates that the histogram can be approximated with the normal distribution # Comparing Confidence Intervals

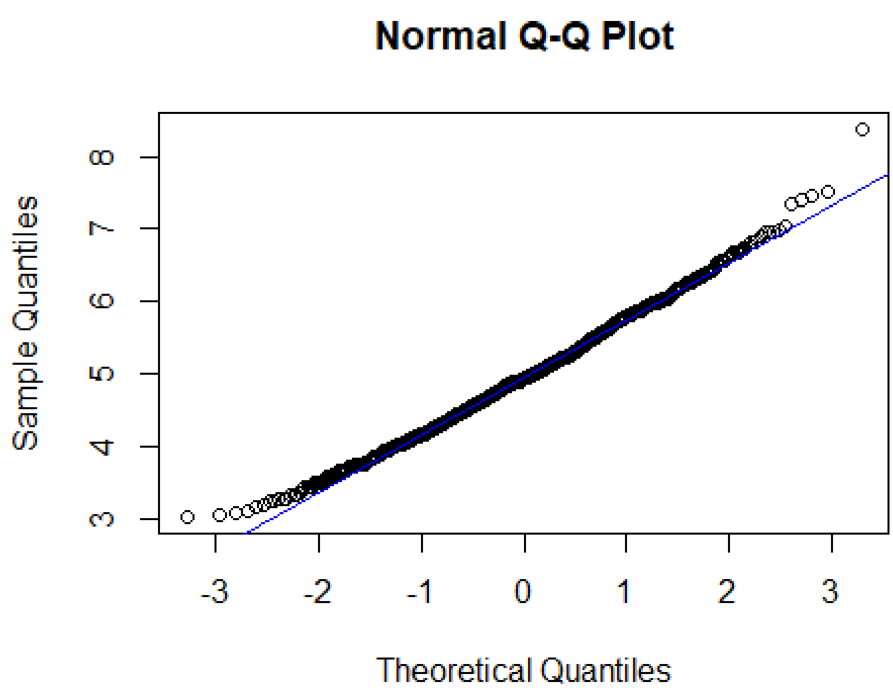
```
sample_conf <- round(mean(sim_mean)+c(-1,1)*1.96*sd(sim_mean)/sqrt(n),3)
sample_conf
```

```
## [1] 4.729 5.215

theoretical_conf <- round(mean(theoritical_mean)+c(-
1,1)*1.96*sqrt(theoretical_var)/sqrt(n),3)
theoretical_conf

## [1] 4.755 5.245
```

The sample 95% confidence interval [4,729,5.215] is close to the theoretical 95% confidence interval of [4.755,5.245]

Q-Q Plot for Quantiles

```
qqnorm(sim_mean,main = "Normal Q-Q Plot", xlab="Theoretical
Quantiles",ylab="Sample Quantiles")
qqline(sim_mean,col="blue")
```



Distribution is approximately normal