Statistical Inference Course Project Part 1

JB

2023-10-04

Overview

This report investigates the exponential distribution inR in comparison to Central Limit Theorem shows basic Inferential Data analysis on the ToothGrowth data set in part 2.

```
# Loading necessary packages
library(datasets)
library(ggplot2)
library(knitr)
library(dplyr)
```

```
##
## 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
```

```
library(tidyr)
```

Part 1: Simulation exercise.

This exercise starts with generating 1000 simulations of 40-size samples of exponential distribution. The simulation is done with a rexp() function in R. The mean of each simulation is calculated using the apply() function.

```
lambda<- 0.2
n<- 40
simData<- replicate(1000, rexp(n,lambda))

simMeans<- apply(simData, 2, mean)
```

1. Show the sample mean and compare it to the theoretical mean of the distribution.

```r
sampleMean<- mean(simMeans)
sampleMean
```

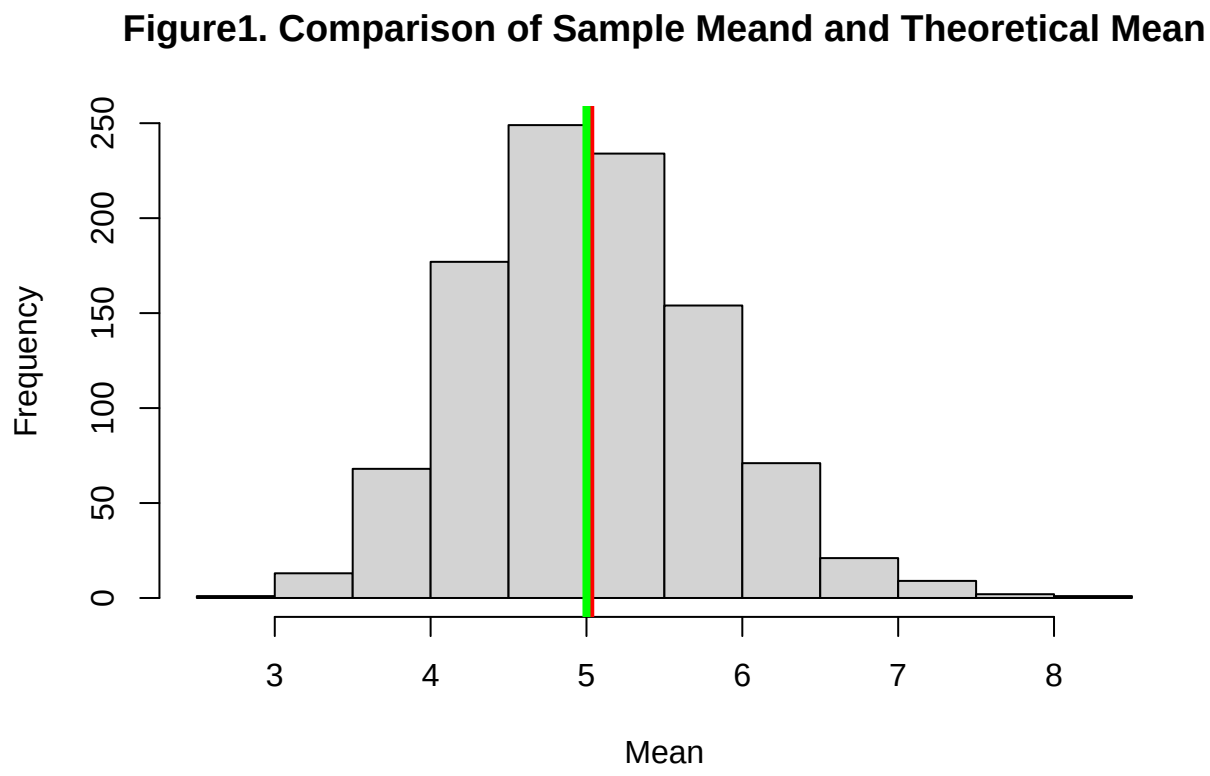
```
## [1] 5.024385
```

```r
theoreticalMean<- 1/lambda
theoreticalMean
```

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

2. Plot Histogram showing the distribution of 1000 mean values calculated for each simulation.

```r
hist(simMeans, main = "Figure1. Comparison of Sample Meand and Theoretical Mean", xlab="Mean")
abline(v= sampleMean, lw = 4, col= "red")
abline(v= theoreticalMean, lw = 4, col = "green")
```



The red line indicates the simulated data mean, whereas the green the theoretical one.

3. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```r
sampleSD<- sd(simMeans)
sampleSD
```

```
## [1] 0.7701222
```

```
theoreticalSD<- (1/lambda)/sqrt(n)
theoreticalSD
```

```
## [1] 0.7905694
```

```
sampleVar<- sampleSD^2
sampleVar
```

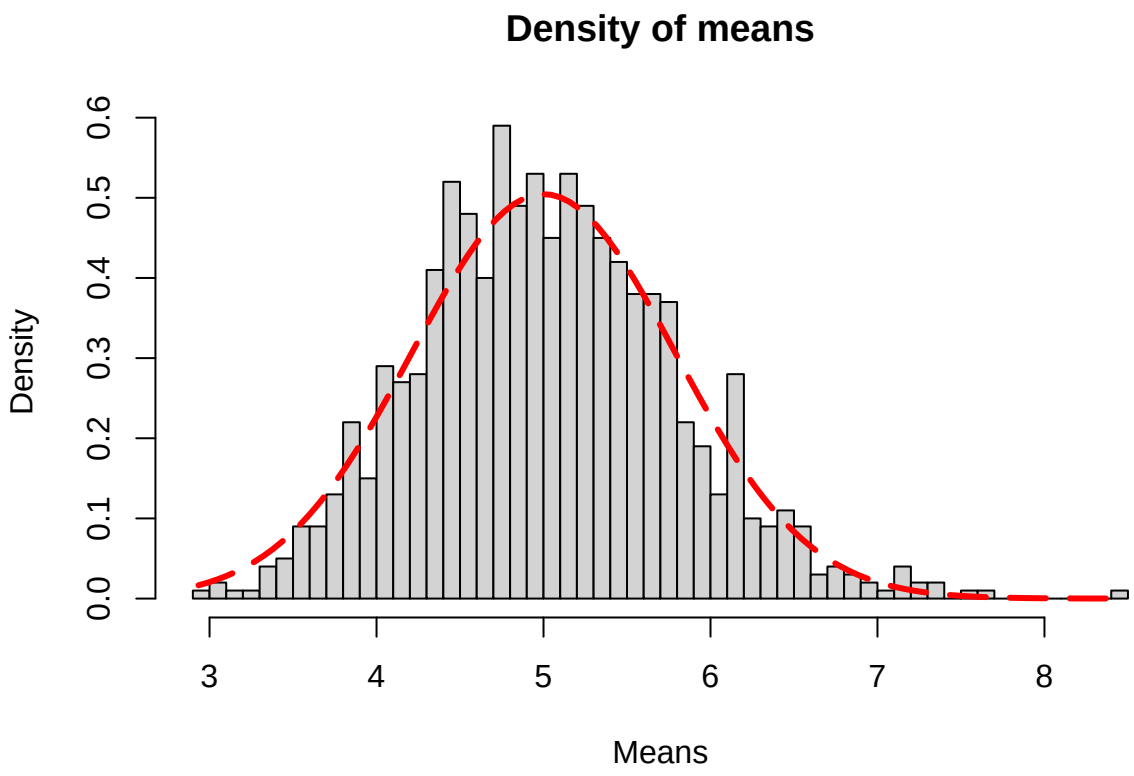
```
## [1] 0.5930882
```

```
theoreticalVAr<- theoreticalSD^2
theoreticalVAr
```

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

4. Show that the distribution is approximately normal.

```
xfit<- seq(min(simMeans), max(simMeans), length=100)
yfit<- dnorm(xfit, mean = 1/lambda, sd= (1/lambda)/sqrt(n))
hist(simMeans, breaks=n, prob = T, main = "Density of means", xlab = "Means", ylab = "Density" )
lines(xfit, yfit, pch = 22, lw=3, lty = 5, col="red")
```



The red curve indicates the normal distribution

Summary between the two distribution can be seen below

```
##             Variable Simulated Theoretical
## 1               Mean 5.0243848   5.0000000
## 2 Standard Deviation 0.7701222   0.7905694
## 3           Variance 0.5930882   0.6250000
```

Conclucion As we can see from the table all of mean, standard deviation and variance are closely similar. Due to the CLT the distribution of averages of 40 exponential is closely similar to a normal distribution.