Statistical Inference - Course Project

Part 1

This is the project for the statistical inference class. I will use simulation to explore inference and do some simple inferential data analysis.

Simulation

The exponential distribution can be simulated in R with `rexp(n, lambda)` where lambda (λ) is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also also $1/\lambda$. For the simulation, I set $\lambda = 0.2$ and I will investigate the distribution of 1000 averages of 40 exponential(0.2)s to answer the following questions:

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

First, let's prepare the data

```
set.seed(1337)
lambda <- 0.2
sample <- 40
nbrSimulation <- 1000

dist <- matrix(rexp(sample*nbrSimulation, lambda),nbrSimulation)
dist.rowMeans <- rowMeans(dist)
```

`dist` is a matrix of 1000 rows (number of simulations) by 40 columns (each exponential) that is converted to `dist.rowMeans`, a vector of size 1000 corresponding to the means of each number of simulation (each rows).

As stated previously, the theoretical center of the distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. However, because the distribution is a sample of averaged exponentials, the standard deviation (σ) is $\sigma = \frac{(1/\lambda)}{\sqrt{40}}$

```
theo_mean <- 1/lambda
theo_sd <- (1/lambda)/sqrt(sample)
theo_var <- theo_sd^2

dist_mean <- mean(dist.rowMeans)
dist_sd <- sd(dist.rowMeans)
dist_var <- dist_sd^2
c('Theorical Mean'=theo_mean,'Standard Deviation'=theo_sd,'Variance'=theo_var)
```

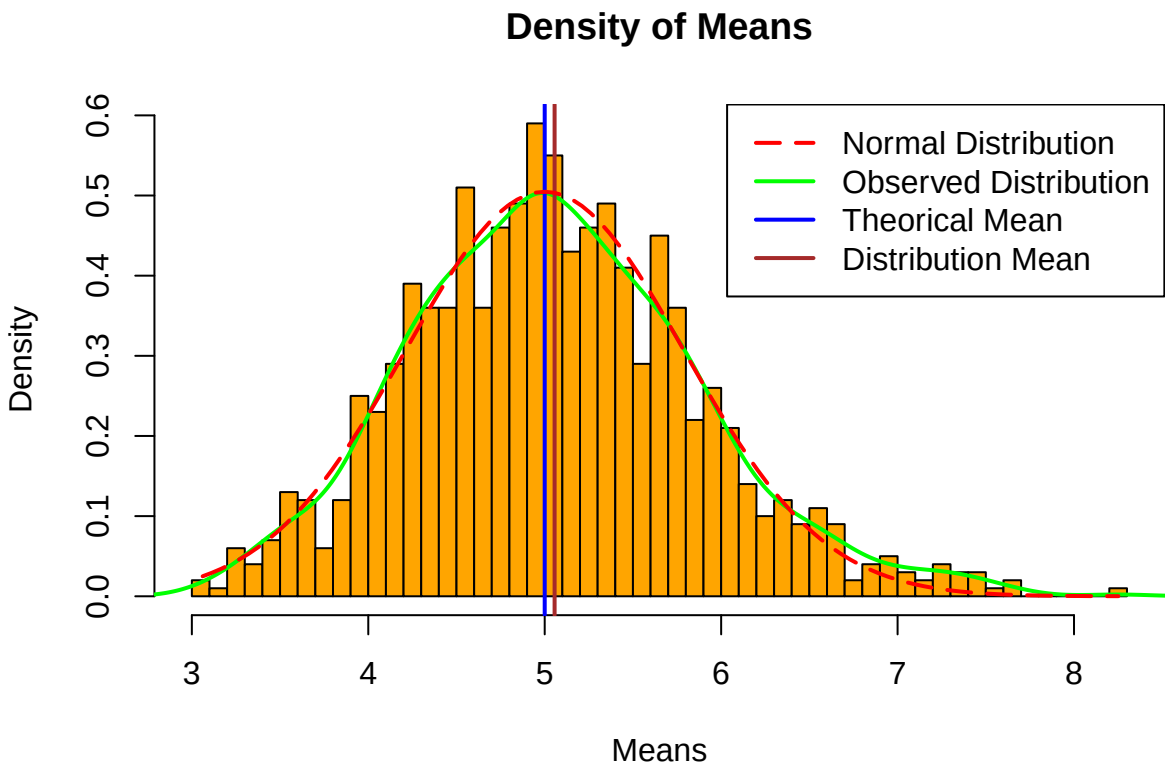
```
##     Theorical Mean Standard Deviation           Variance
##             5.0000             0.7906             0.6250
```

```
c('Distrivution Mean'=dist_mean,'Standard Deviation'=dist_sd,'Variance'=dist_var)
```

```
##  Distrivution Mean Standard Deviation           Variance
##             5.0560             0.8155             0.6651
```

The mean and the standart deviation of the distribution sample are μ=5.056 and σ=0.815 respectively which is close to the theorical values.

Here is a plot of the distribution.



Answers

I will refer to the previous plot to answer the questions.

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

The center of the distribution is the mean . They are represented by the blue and brown lines. The values are μ=5 for the theoretical mean and μ=5.056 for the distribution mean

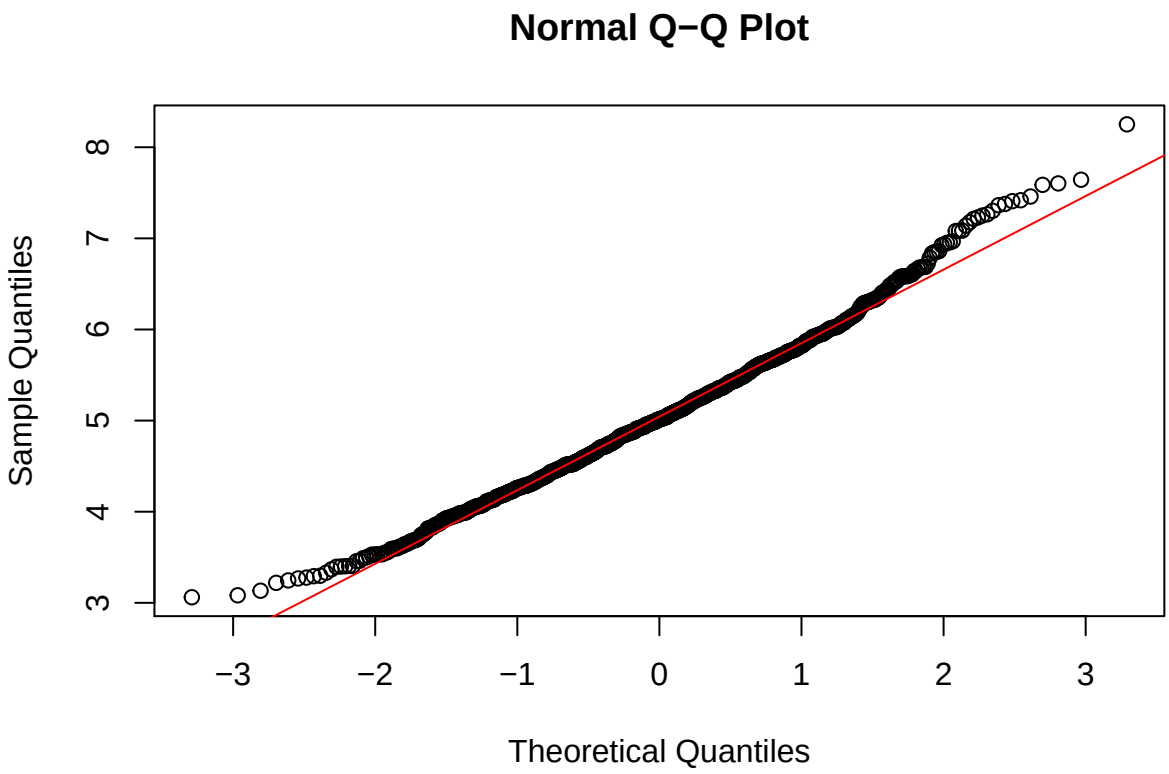
2. Show how variable it is and compare it to the theoretical variance of the distribution.

The variance ($Var(X)$=σ^2) is also close to the theoretical one: 0.6651 versus 0.6250.

3. Show that the distribution is approximately normal.

The plot shows the curve of a normal distribution in red. We can see that the green line (the actual density) is similar to the normal curve. Here is another way to see how close of a normal distribution the sample is:

```
qqnorm(dist.rowMeans)
qqline(dist.rowMeans, col = 'red')
```


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
3
4
5
6
7
8
−3
−2
−1
0
1
2
3