Statistical Inference: Final Project Part 1

Objectvies:

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.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Preparations

```
library(ggplot2)

#Settings
set.seed(123)
lambda<- 0.2
n <- 40
simulations <- 1000

#Simulate
simulated_exp <- replicate(simulations,rexp(n,lambda))

#Calculate the mean
mean_si <- apply(simulated_exp,2,mean)
```

Q1. Show the sample mean and compare it to the theoretical mean of the distribution.

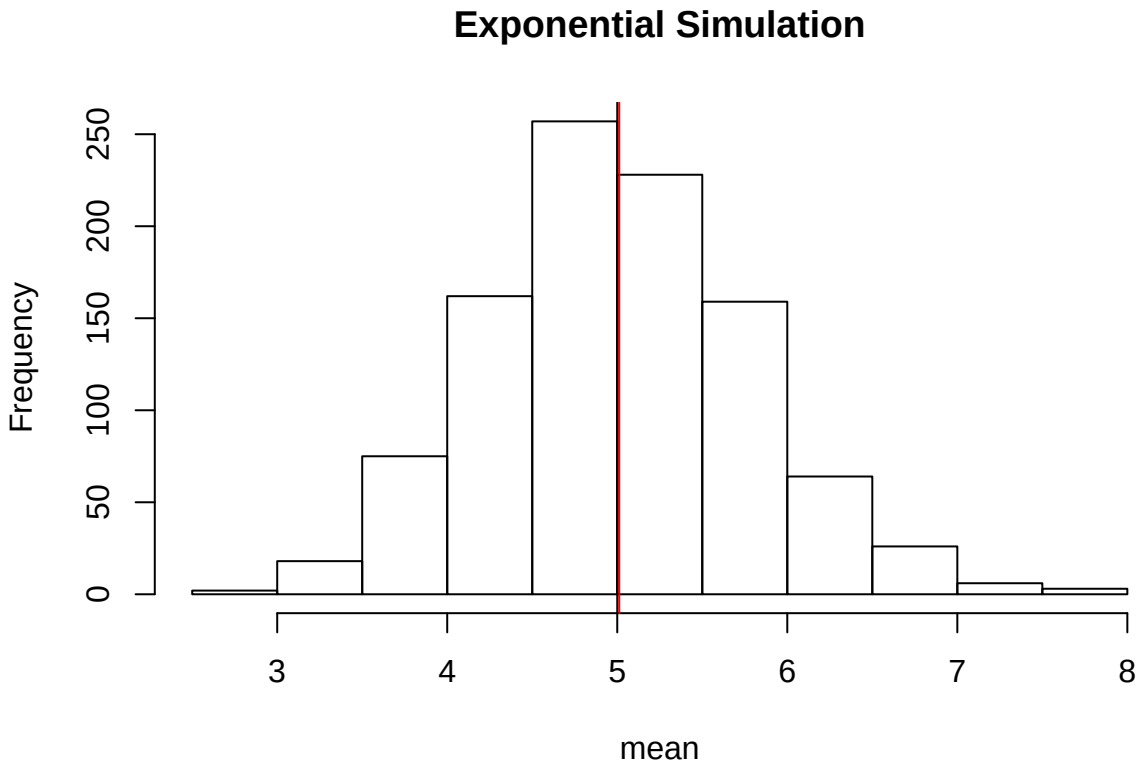
```
sample_mean <- mean(mean_si)
sample_mean
```

```
## [1] 5.011911
```

```
theory_mean <- 1/lambda
theory_mean
```

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

```
#Visual
hist(mean_si, xlab= "mean", main= "Exponential Simulation")
abline(v= sample_mean, col="red")
abline(v= theory_mean, col="black")
```



Ans the sample mean is close to the theoretical mean.

Q2. Show how variable it is and compare it to the theoretical variance of the distribution

```
#std of distribution
std_dist <- sd(mean_si)
std_dist
```

```
## [1] 0.7749147
```

```
#std from theory
std_theory <- (1/lambda)/(sqrt(n))
std_theory
```

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

```
#Variance of distribution
var_dist <- (std_dist)^2
var_dist
```

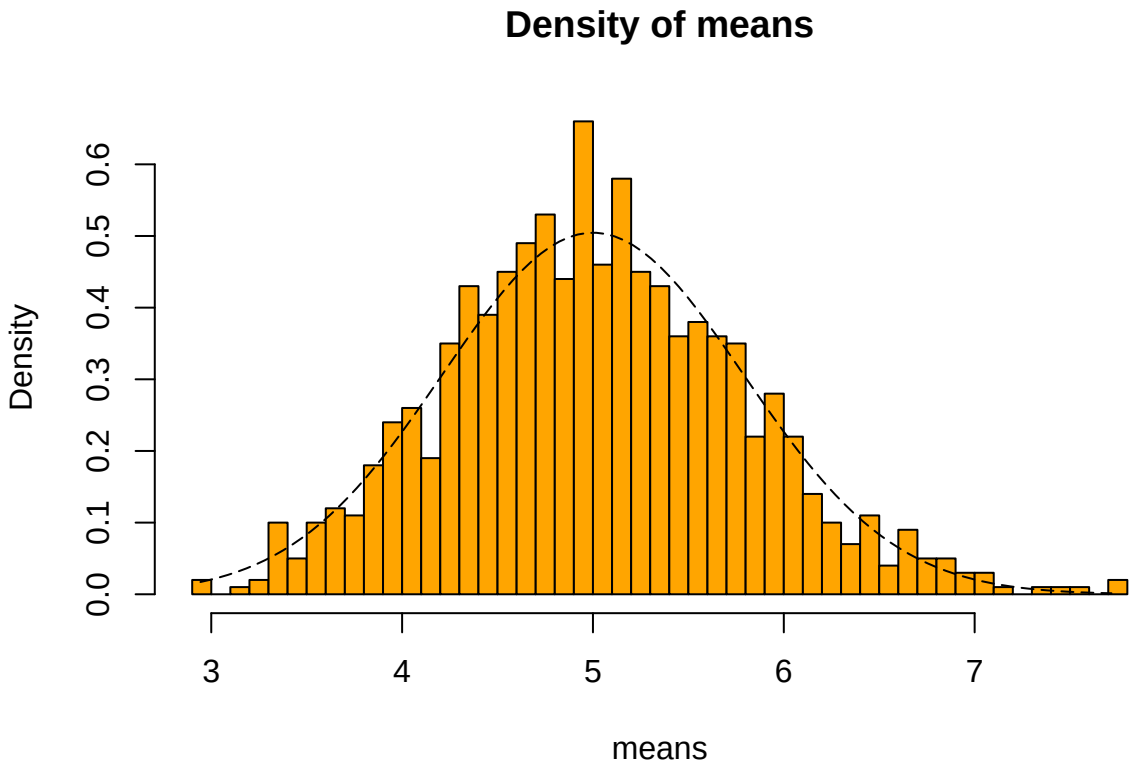
```
## [1] 0.6004928
```

```
#variance from theory
var_theory <- ((1/lambda)*(1/sqrt(n)))^2
var_theory
```

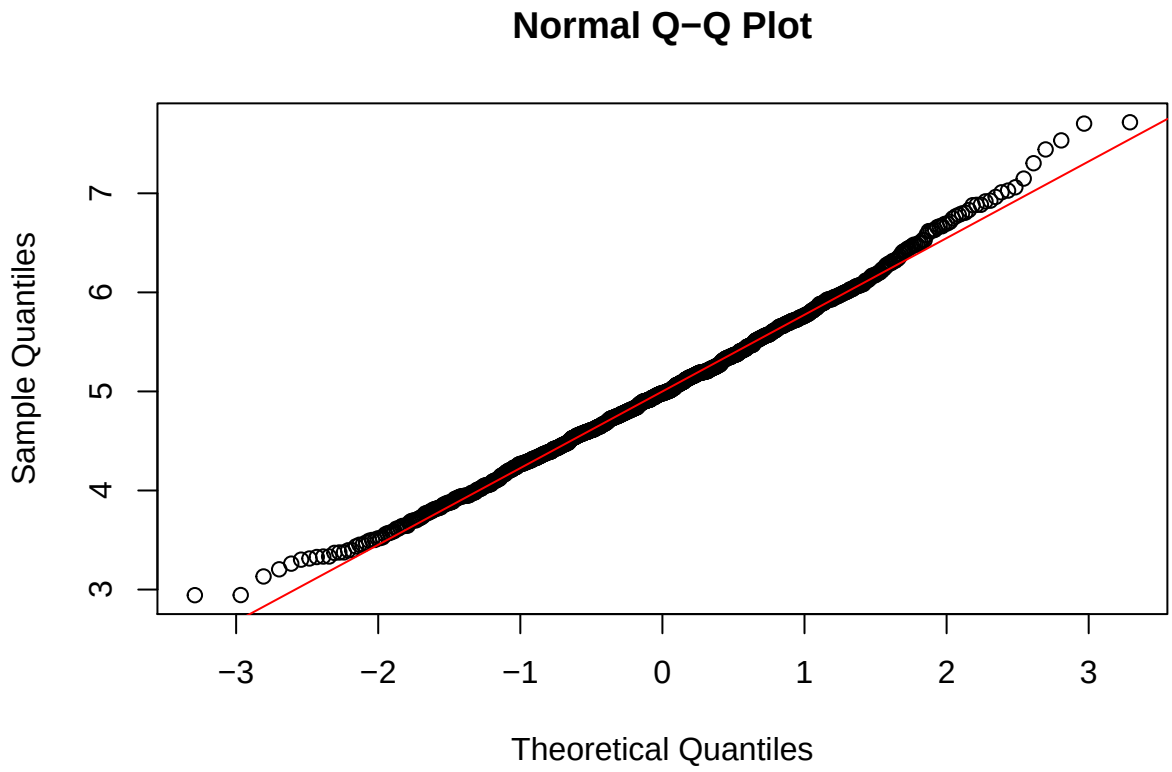
```
## [1] 0.625
```

Q3 Show that the distribution is approximately normal.

```
x <- seq(min(mean_si), max(mean_si), length=100)
y <- dnorm(x, mean=1/lambda, sd=(1/lambda/sqrt(n)))
hist(mean_si,breaks=n,prob=T,col="orange",xlab = "means",main="Density of means")
lines(x, y, pch=22, col="black", lty=5)
```




```
# compare the distribution of averages of 40 exponentials to a normal distribution
qqnorm(mean_si)
qqline(mean_si, col = 2)
```



According to CLT, the distribution of avg of 40 exponentials is close to normal distribution.