Statistical Inference Project 2

29 December 2017

Instructions

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.

Loading required Library

```
library("ggplot2")
library("data.table")
```

```
## Warning: package 'data.table' was built under R version 3.4.3
```

In this project we 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. We will investigate the distribution of averages of 40 exponentials.

```
# set seed for reproducability
set.seed(35)

# set lambda to 0.2
lambda <- 0.2

# 40 samples
n <- 40

# 1000 simulations
simulations <- 1000

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

# calculate mean of exponentials
means_exp <- apply(simulated_exp, 2, mean)
```

#Question 1

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

defining analytical mean

```
analytical_mean <- mean(means_exp)

analytical_mean
```

```
## [1] 4.988198
```

analytical mean

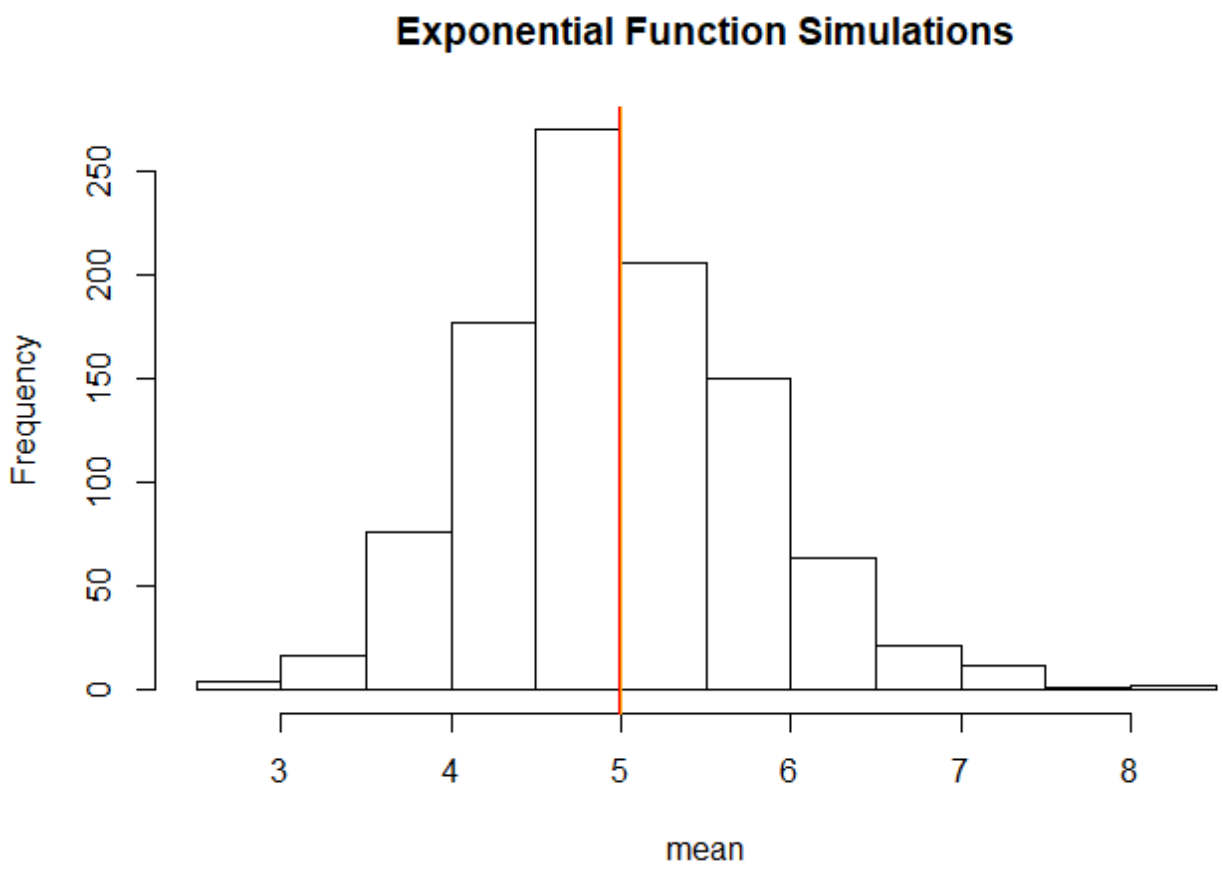
```
theory_mean <- 1/lambda
theory_mean
```

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

The analytics mean is 4.993867 the theoretical mean 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

Plotting Histogram of Exponential simulations

```
hist(means_exp, xlab = "mean", main = "Exponential Function Simulations")
abline(v = analytical_mean, col = "red")
abline(v = theory_mean, col = "orange")
```



Again from Histogram its confirm that Analytical mean and Theoretical mean are very close to centre of

distribution

#Question 2

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

standard deviation of distribution

```
standard_deviation_dist <- sd(means_exp)

standard_deviation_dist
```

```
## [1] 0.7988187
```

standard deviation from analytical expression

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

standard_deviation_theory
```

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

variance of distribution

```
variance_dist <- standard_deviation_dist^2

variance_dist
```

```
## [1] 0.6381113
```

variance from analytical expression

```
variance_theory <- ((1/lambda)*(1/sqrt(n)))^2

variance_theory
```

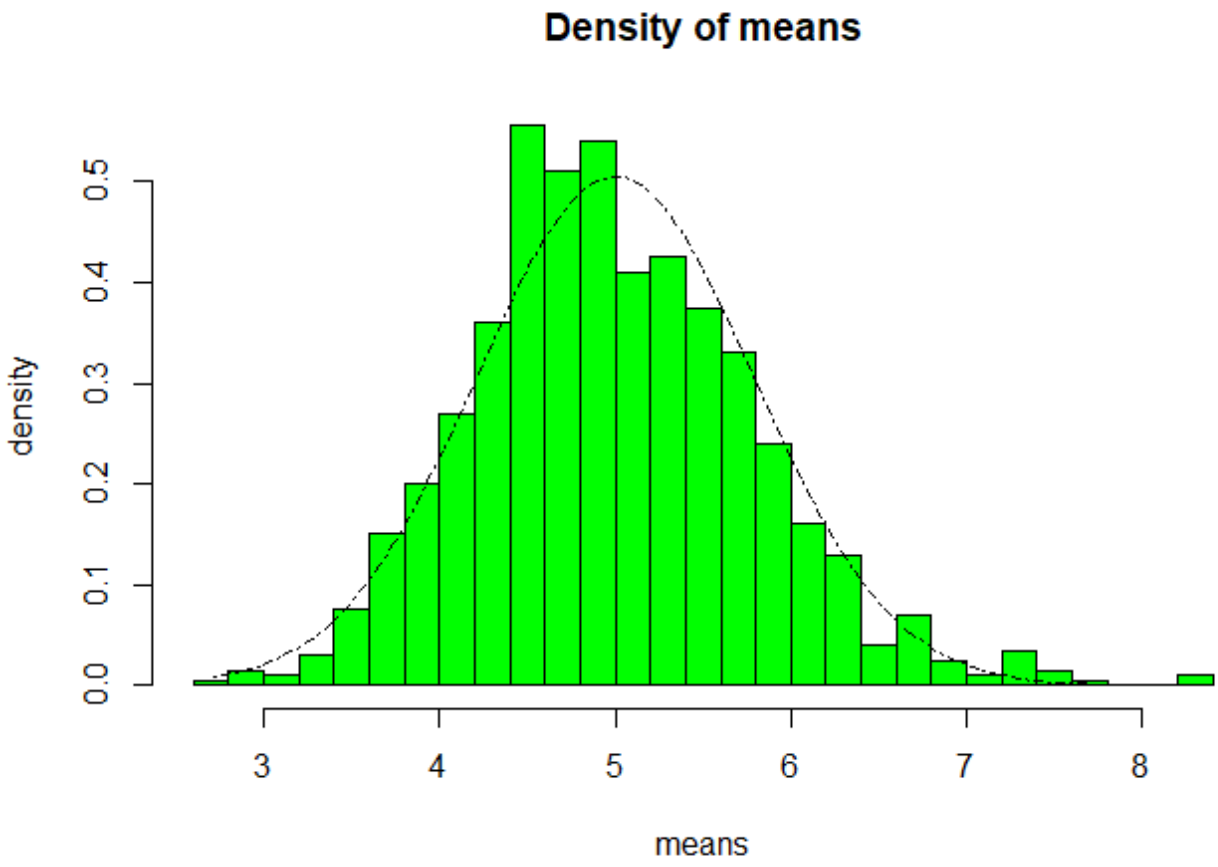
```
## [1] 0.625
```

Standard Deviation of the distribution is 0.7931608 with the theoretical SD calculated as 0.7905694. The Theoretical variance is calculated as ((1 /lambda) * (1/Sqrt(n))^2 = 0.625. The actual variance of the distribution is 0.6291041

Question 3

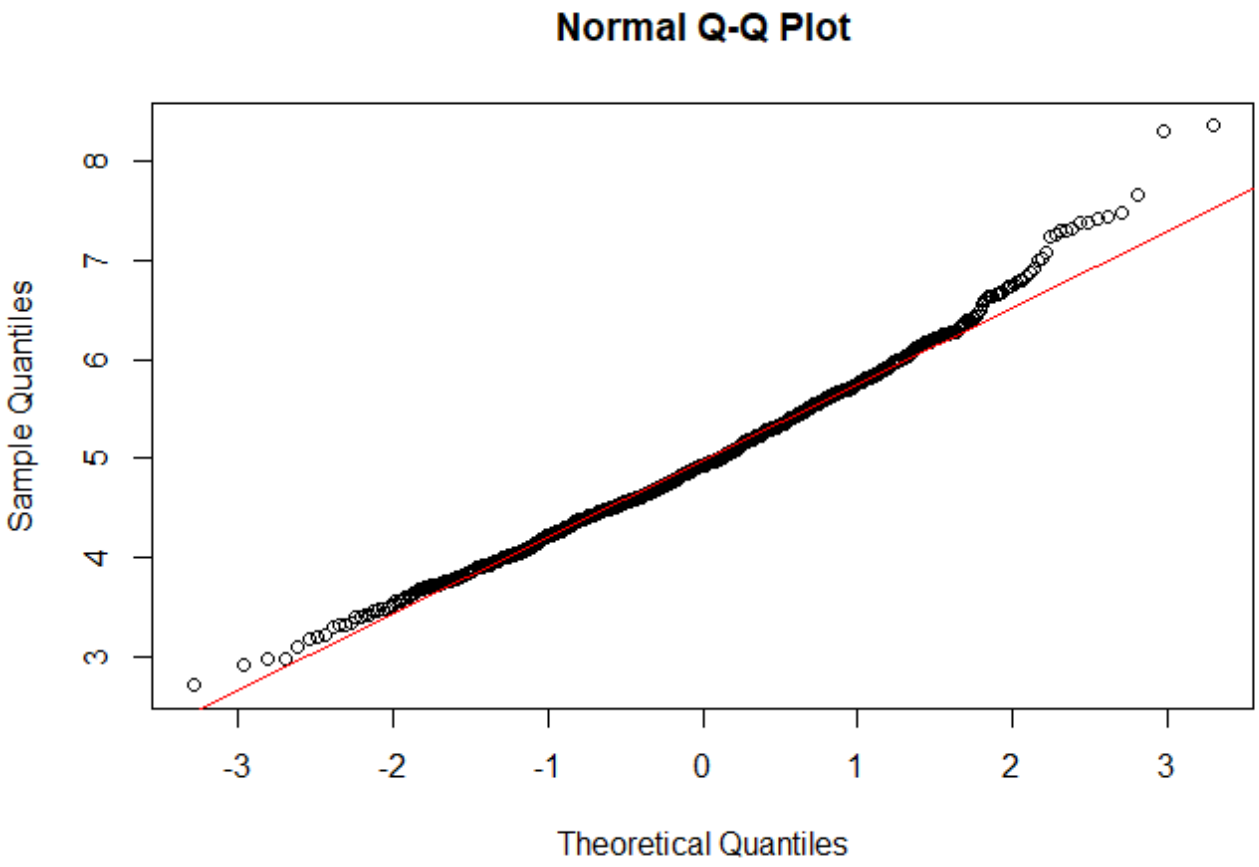
Show that the distribution is approximately normal.

```
xfit <- seq(min(means_exp), max(means_exp), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
hist(means_exp,breaks=n,prob=T,col="green",xlab = "means",main="Density of means",ylab
="density")
lines(xfit, yfit, pch=21, col="black", lty=4)
```



compare the distribution of averages of 40 exponentials to a normal distribution

```
qqnorm(means_exp)
qqline(means_exp, col = 2)
```



Due to Due to the central limit theorem (CLT), the distribution of averages of 40 exponentials is very close to a normal distribution.