Statistical Inference Project Part 1

Akshay Amrit

11th November, 2019

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.

Load Libraries

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

```
## Warning: package 'data.table' was built under R version 3.6.1
```

```r
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 3.6.1
```

Assignment

```r
# set seed for reproducability
set.seed(3)
# Set values specified in the project:
lambda <- 0.2
n <- 40
simulations <- 1000
# simulate
simulated_exponentials <- replicate(simulations, rexp(n, lambda))
# calculate mean of exponentials
simulated_exponentials.mean <- apply(simulated_exponentials, 2, mean)
```

Question 1

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

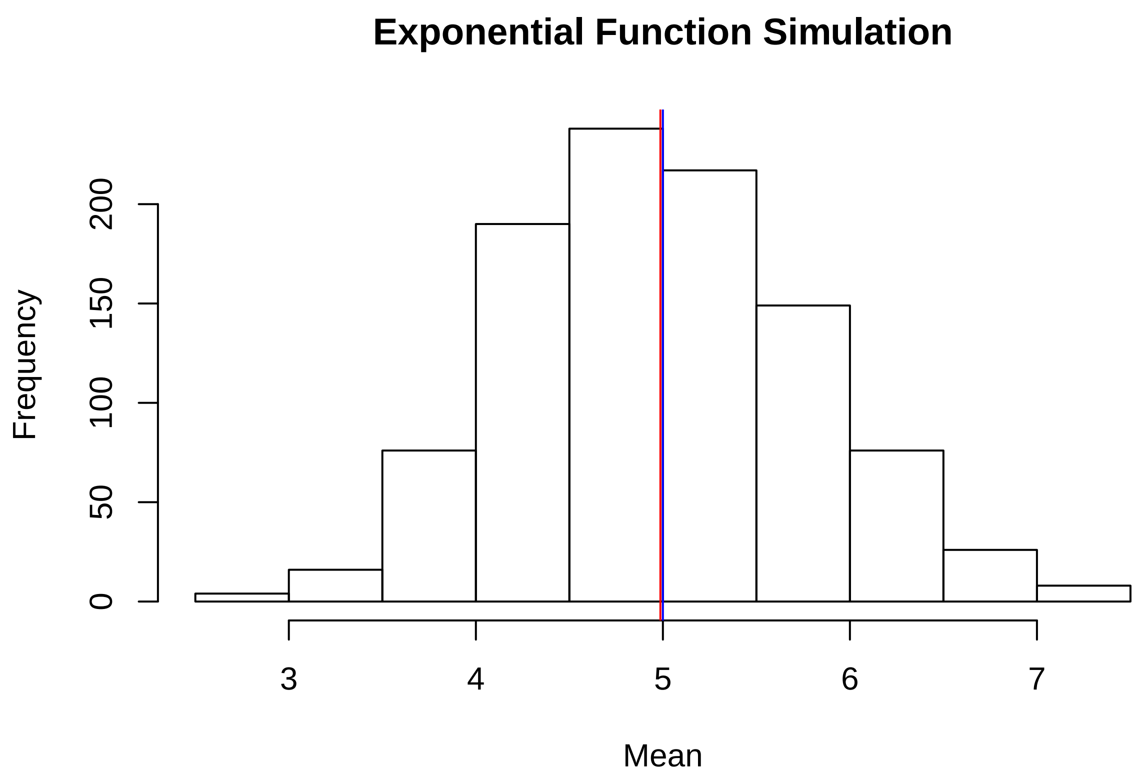
```r
analytical_mean <- mean(simulated_exponentials.mean)
analytical_mean
```

```
## [1] 4.98662
```

```r
# analytical mean
theory_mean <- 1/lambda
theory_mean
```

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

```r
# visualization
hist(simulated_exponentials.mean, xlab = "Mean", main = "Exponential Function Simulation")
abline(v = analytical_mean, col = "red")
abline(v = theory_mean, col = "blue")
```



The analytical mean is 4.98662 whereas theoretical mean is 5. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

Question 2

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

```r
# standard deviation of distribution
standard_deviation_dist <- sd(simulated_exponentials.mean)
standard_deviation_dist
```

```
## [1] 0.7947823
```

```r
# standard deviation using analytical expression
standard_deviation_theory <- (1/lambda)/sqrt(n)
standard_deviation_theory
```

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

```r
# variance of distribution
variance_dist <- standard_deviation_dist^2
variance_dist
```

```
## [1] 0.6316789
```

```r
# variance from analytical expression
variance_theory <- ((1/lambda)*(1/sqrt(n)))^2
variance_theory
```

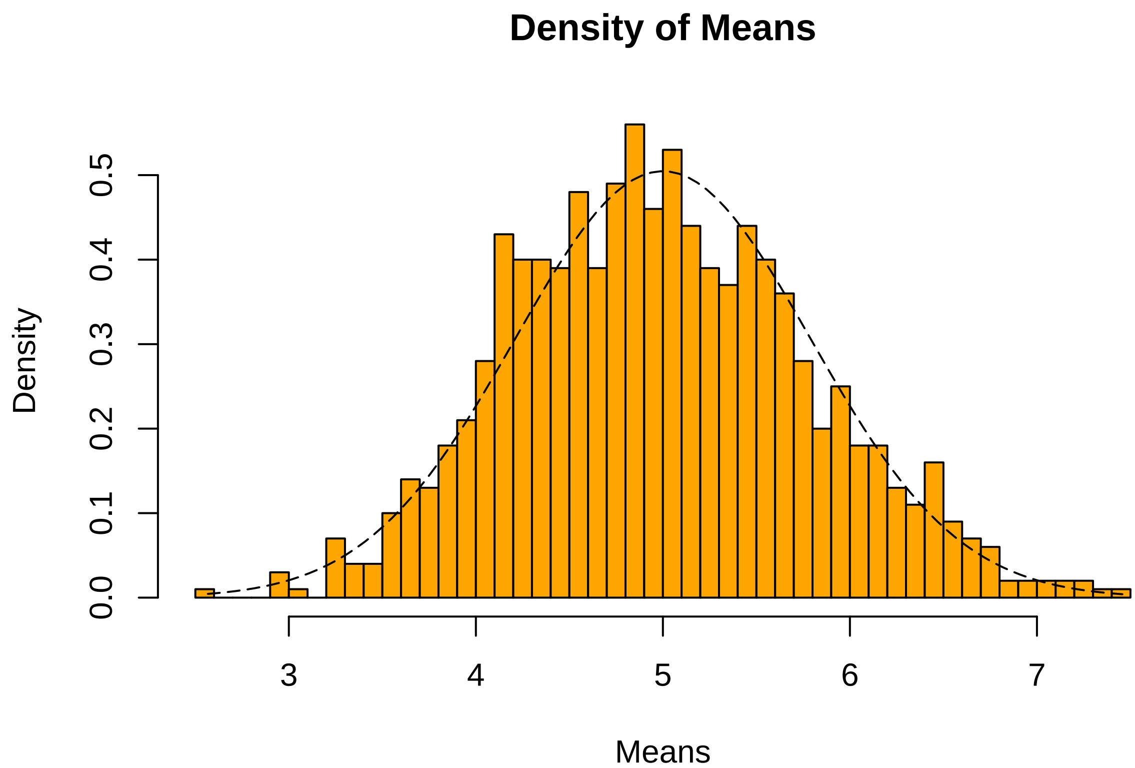
```
## [1] 0.625
```

Standard Deviation of the distribution is 0.7947823 with the theoretical standard deviation calculated as 0.7905694. The Theoretical variance calculated is 0.625. The actual variance of the distribution is 0.6316789.

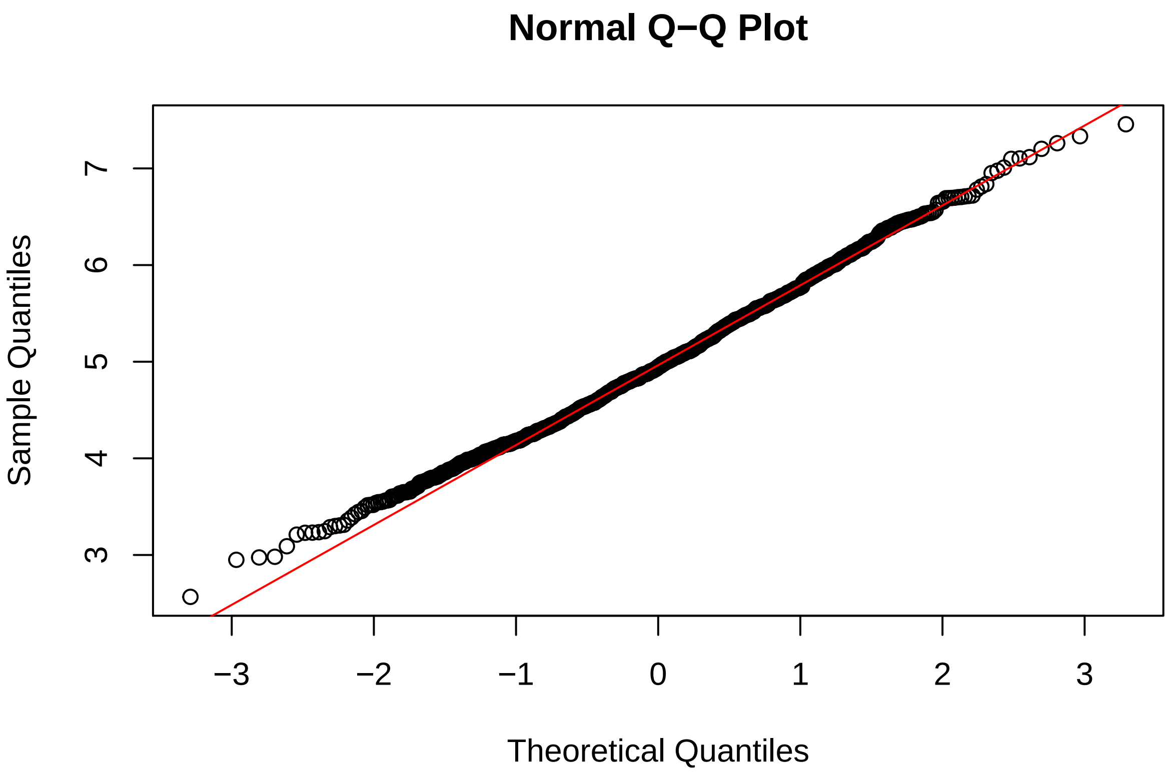
Question 3

Show that the distribution is approximately normal.

```r
xfit <- seq(min(simulated_exponentials.mean), max(simulated_exponentials.mean), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))
hist(simulated_exponentials.mean,breaks=n,prob=T,col="orange"
     ,xlab = "Means",main="Density of Means",ylab="Density")
lines(xfit, yfit, pch=22, col="black", lty=5)
```




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



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