Statistical Inference Project Part-1

Shubhendu Dash

14/07/2020

GitHub Repository for the Course: Statistical Inference

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 Libraries

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

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

Task

```
# set seed for reproducability

set.seed(31)

# set lambda to 0.2

lambda <- 0.2

# 40 samples

n <- 40

# 1000 simulations

Simulations <- 1000

# simulate
```

```r
SimulatedExponentials <- replicate(Simulations, rexp(n, lambda))

# calculate mean of exponentials

MeansExponentials <- apply(SimulatedExponentials, 2, mean)
```

Question 1

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

```r
AnaltyticalMean <- mean(MeansExponentials)
AnaltyticalMean
```

```
## [1] 4.993867
```

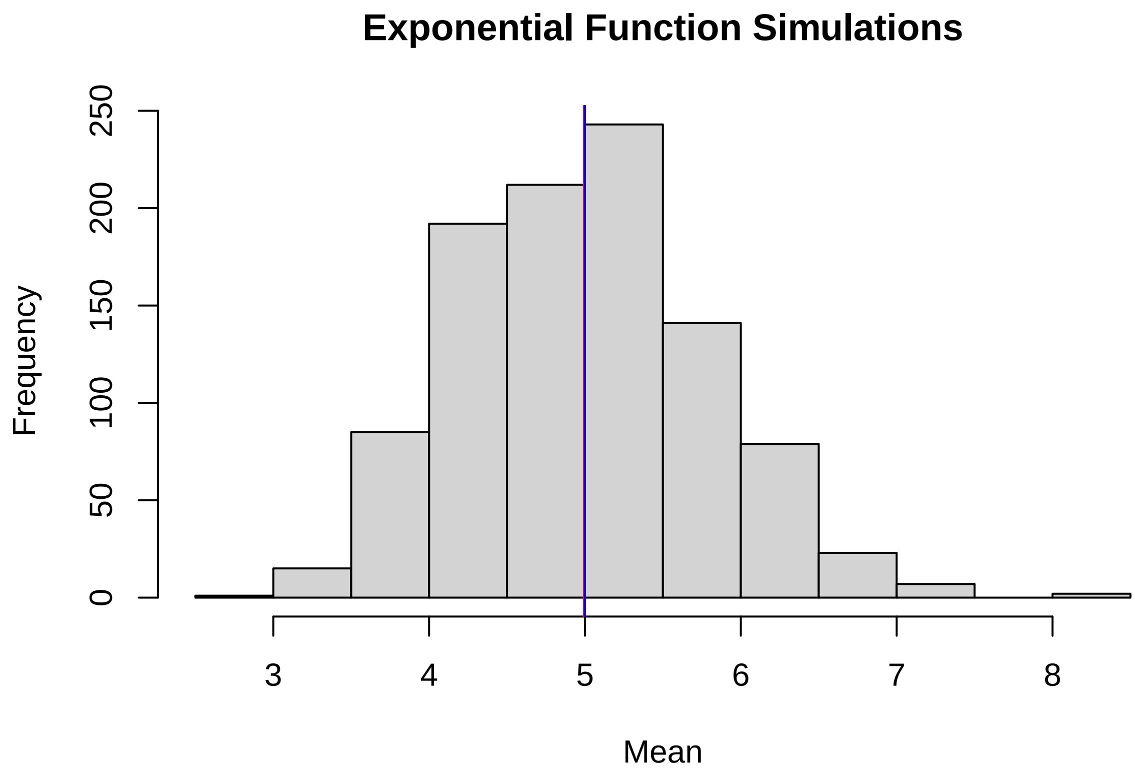
```r
# Analytical Mean

TheoreticalMean <- 1/lambda
TheoreticalMean
```

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

```r
# Visualization

hist(MeansExponentials, xlab = "Mean", main = "Exponential Function Simulations")
abline(v = AnaltyticalMean, col = "red")
abline(v = TheoreticalMean, col = "blue")
```



The analytics mean is 4.993867 and the 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 it is and compare it to the theoretical variance of the distribution.

```
# standard deviation of distribution

SDDistribution <- sd(MeansExponentials)
SDDistribution
```

```
## [1] 0.7931608
```

```
# standard deviation from analytical expression

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

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

```
# variance of distribution

VarDistribution <- SDDistribution^2
VarDistribution
```

```
## [1] 0.6291041
```

```r
# variance from analytical expression

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

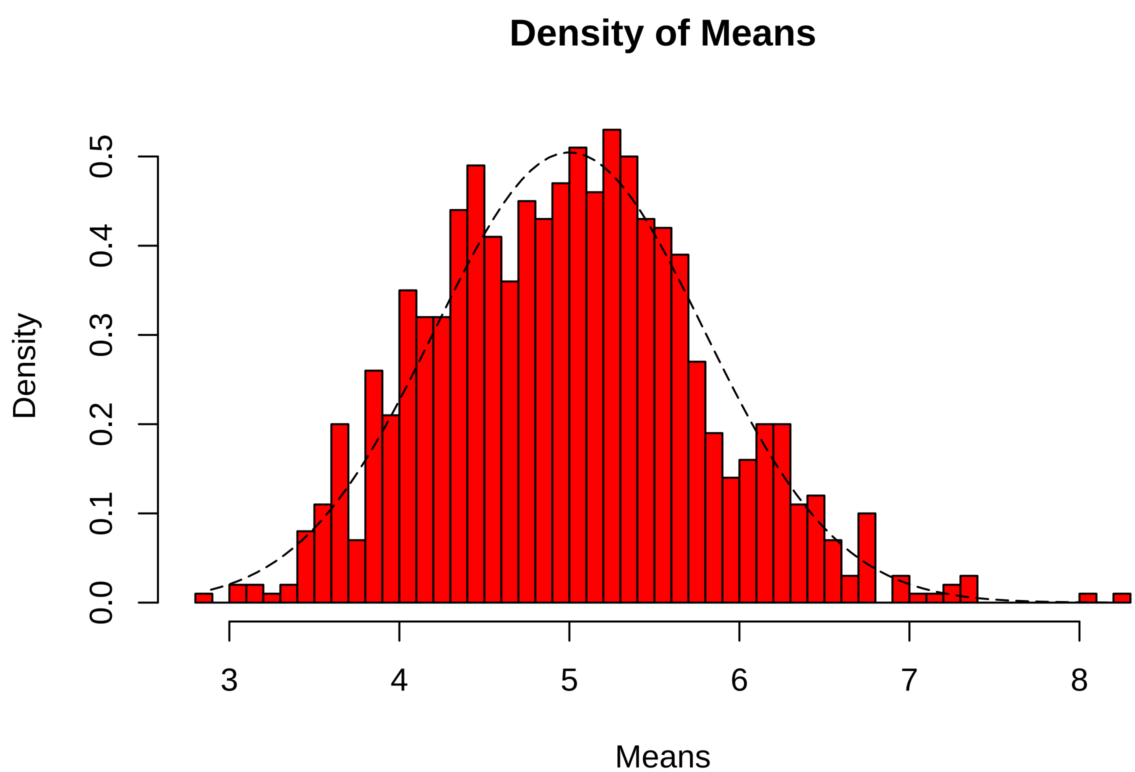
```
## [1] 0.625
```

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

Question 3

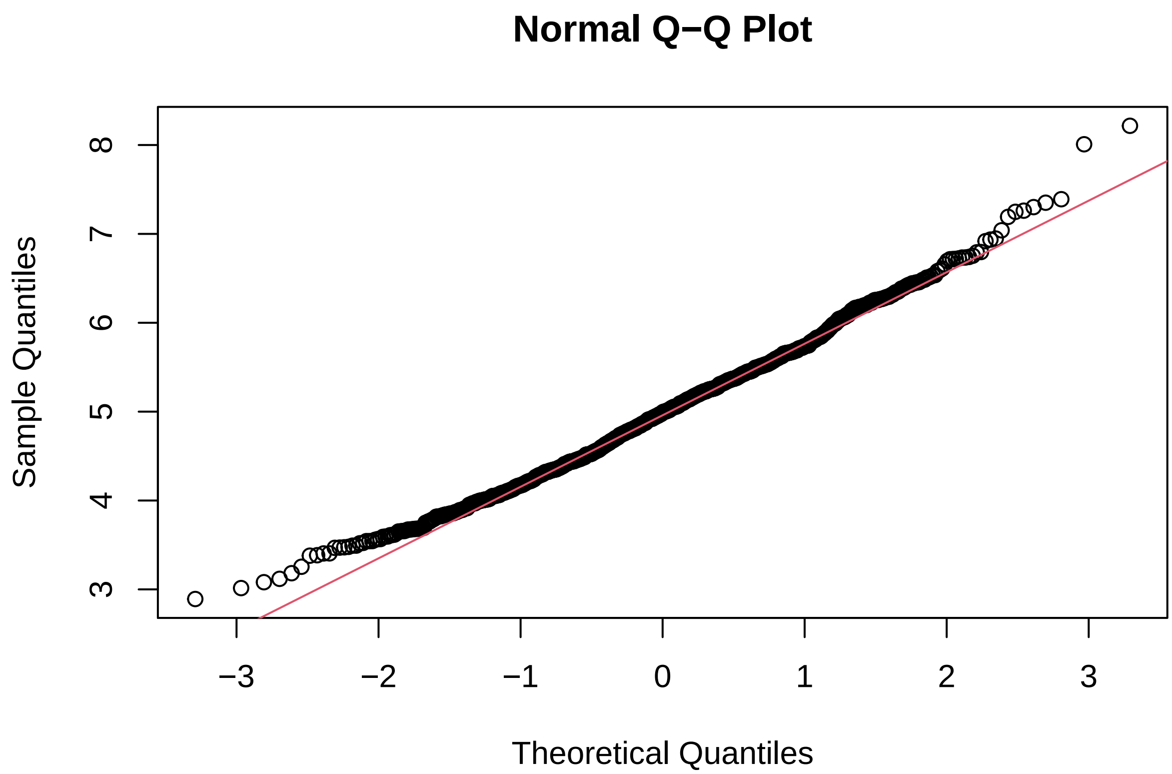
Show that the distribution is approximately normal.

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



```r
# compare the distribution of averages of 40 exponentials to a normal distribution

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



Conclusion

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