Statistical Interference Course Project

Harshad B.

18/07/2020

Tasks to accomplish

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")
```

Initialise

```
# set seed for reproducability
set.seed(31, sample.kind = "Rounding")
```

```
## Warning in set.seed(31, sample.kind = "Rounding"): non-uniform 'Rounding'
## sampler used
```

```
# set lambda to 0.2
lambda <- 0.2
# 40 samples
n <- 40
# 1000 simulations
simulations <- 1000
# simulate
simulated_exponentials <- replicate(simulations, rexp(n, lambda))
# calculate mean of exponentials
means_exponentials <- apply(simulated_exponentials, 2, mean)
```

Q1

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

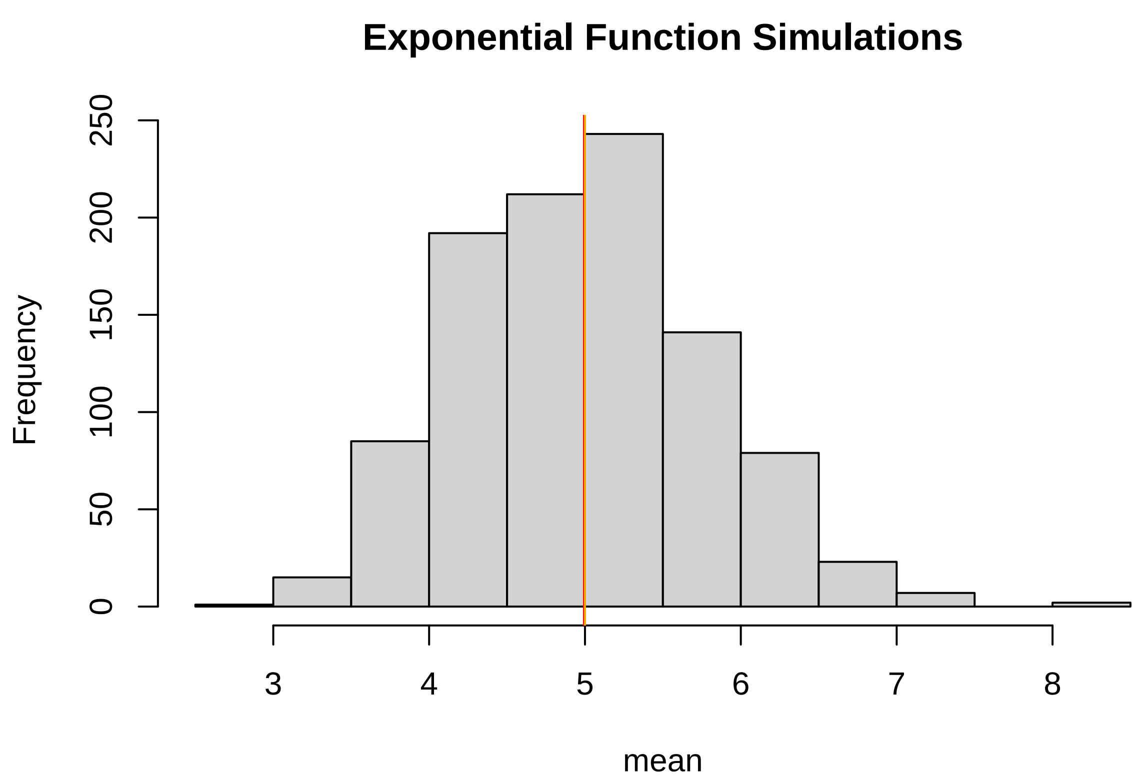
```r
analytical_mean <- mean(means_exponentials)
analytical_mean
```

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

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

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

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



- Analytics mean = 4.993867
- Theoretical mean = 5
- The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

Q2

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_exponentials)
standard_deviation_dist
```

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

```
# 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] 0.6291041
```

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

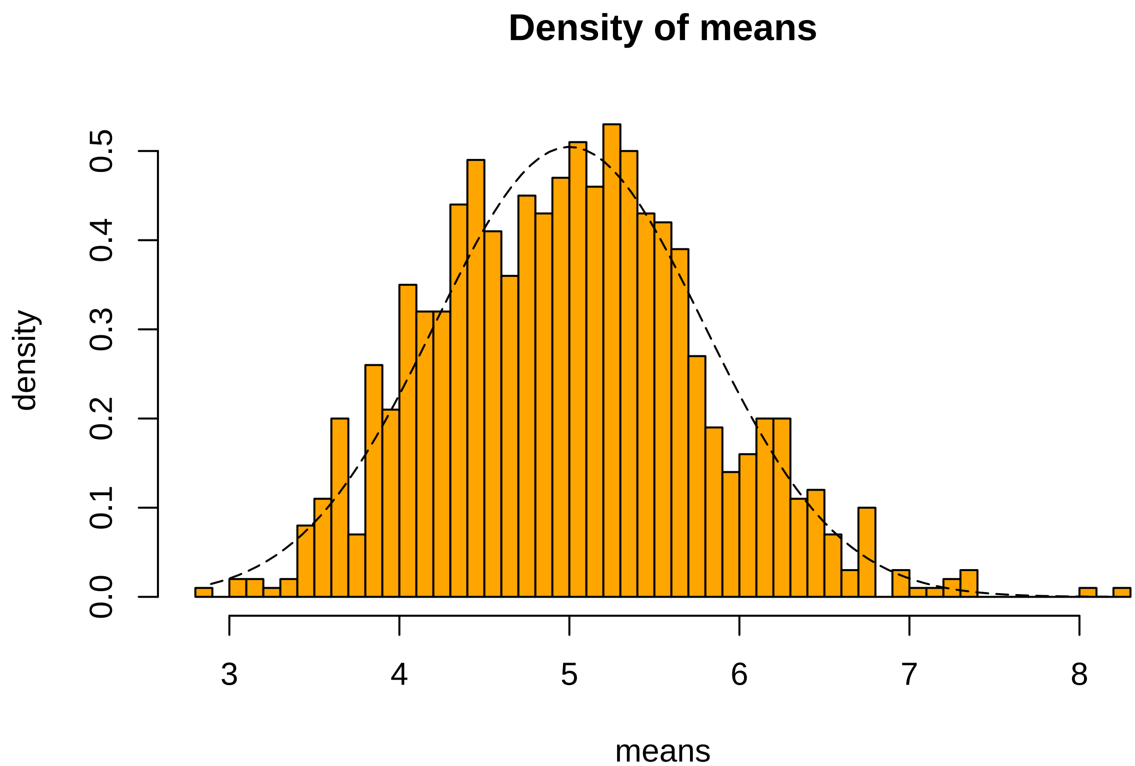
```
## [1] 0.625
```

- Standard Deviation of the distribution = 0.7931608
- Theoretical SD = 0.7905694
- Theoretical variance = 0.625.
- Actual variance of the distribution = 0.6291041

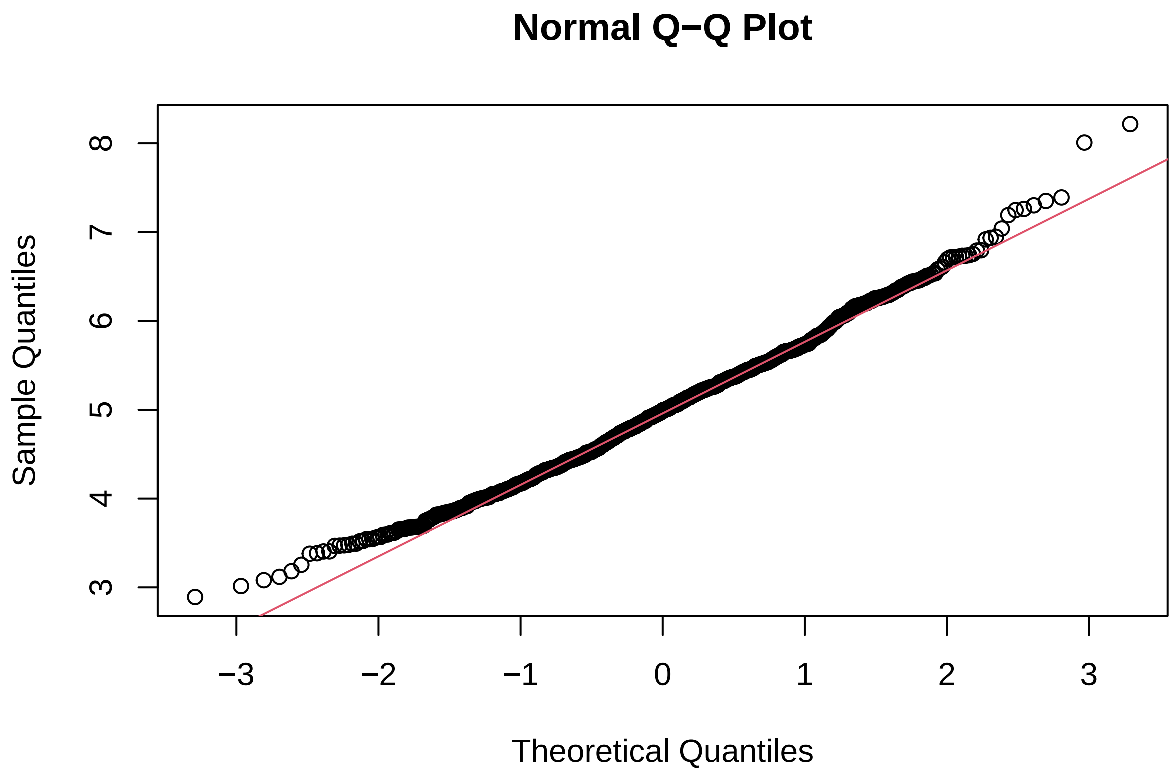
Q3

Show that the distribution is approximately normal.

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




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



- Due to the Central Limit Theorem, the distribution of averages of 40 exponentials is very close to a normal distribution.