Statistical Inference

JC

September 7, 2017

```
# set seed for reproducability
set.seed(1006)

# 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)
```

Question 1

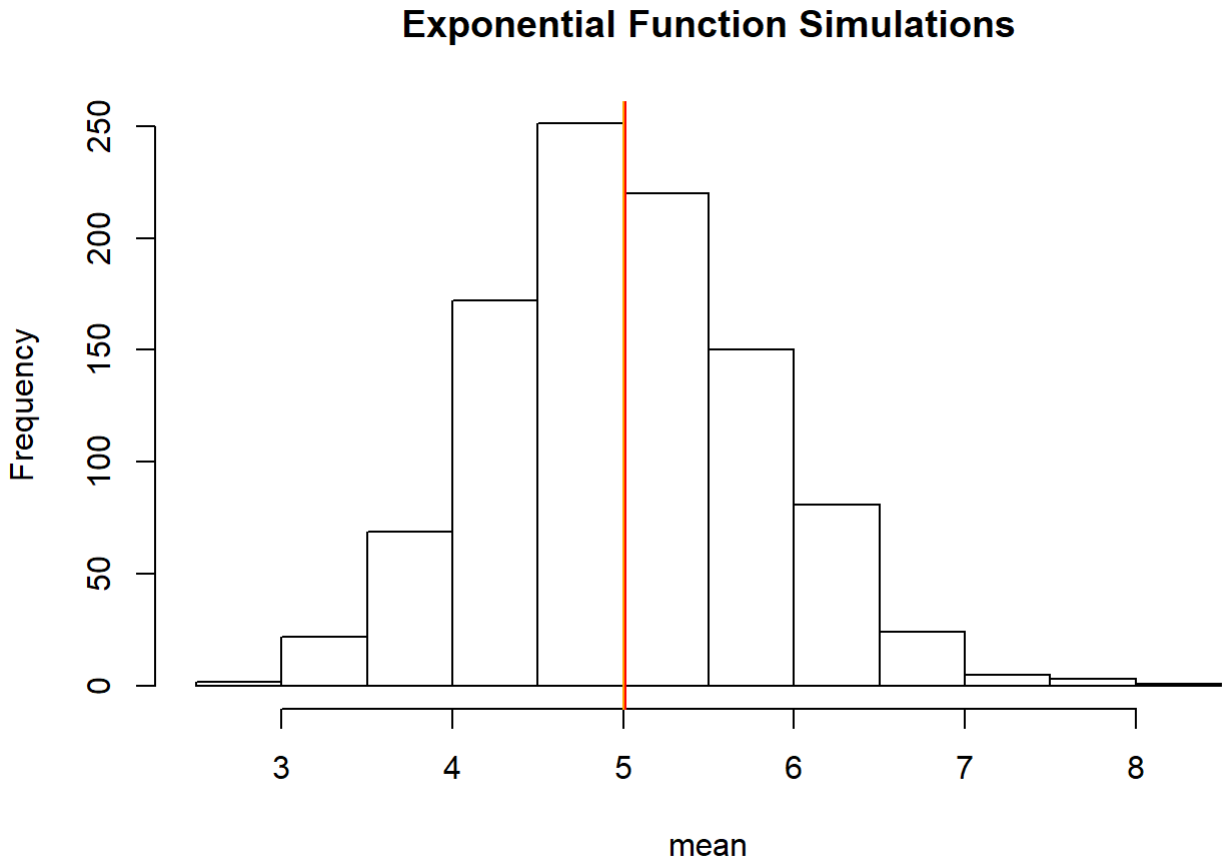
```
# distribution mean
simulation_mean <- mean(means_exponentials)
simulation_mean
```

```
## [1] 5.016005
```

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

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

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



Answer 1

The simulated mean is 5.02 , and the theoretical mean 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
standard_deviation_dist <- sd(means_exponentials)
standard_deviation_dist
```

```
## [1] 0.799028
```

```
# 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] 0.6384457
```

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

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

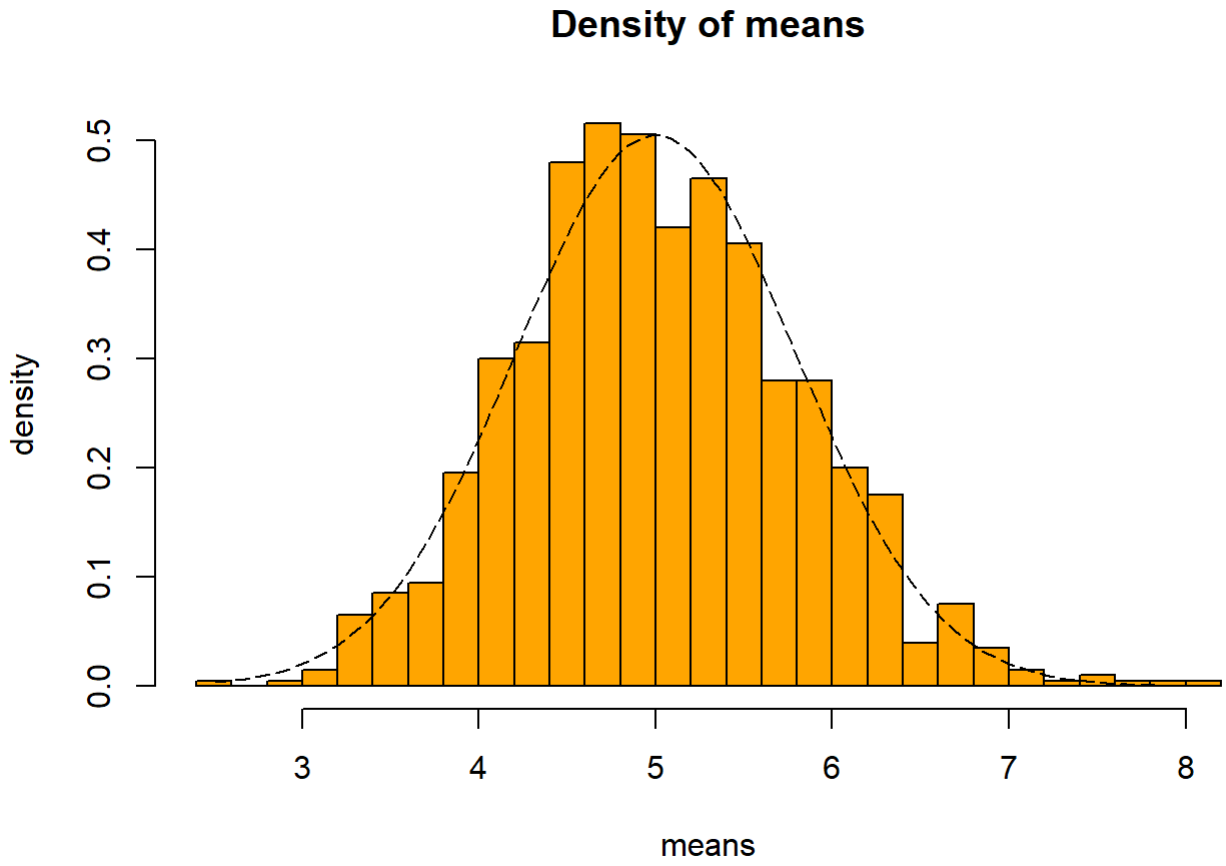
Answer 2

Standard Deviation of the distribution is 0.799 with the theoretical SD calculated as 0.791. The Theoretical variance is calculated as ((1/Lambda) * ((n)^.5)) ^ 2 = 0.625. The actual variance of the distribution is 0.638

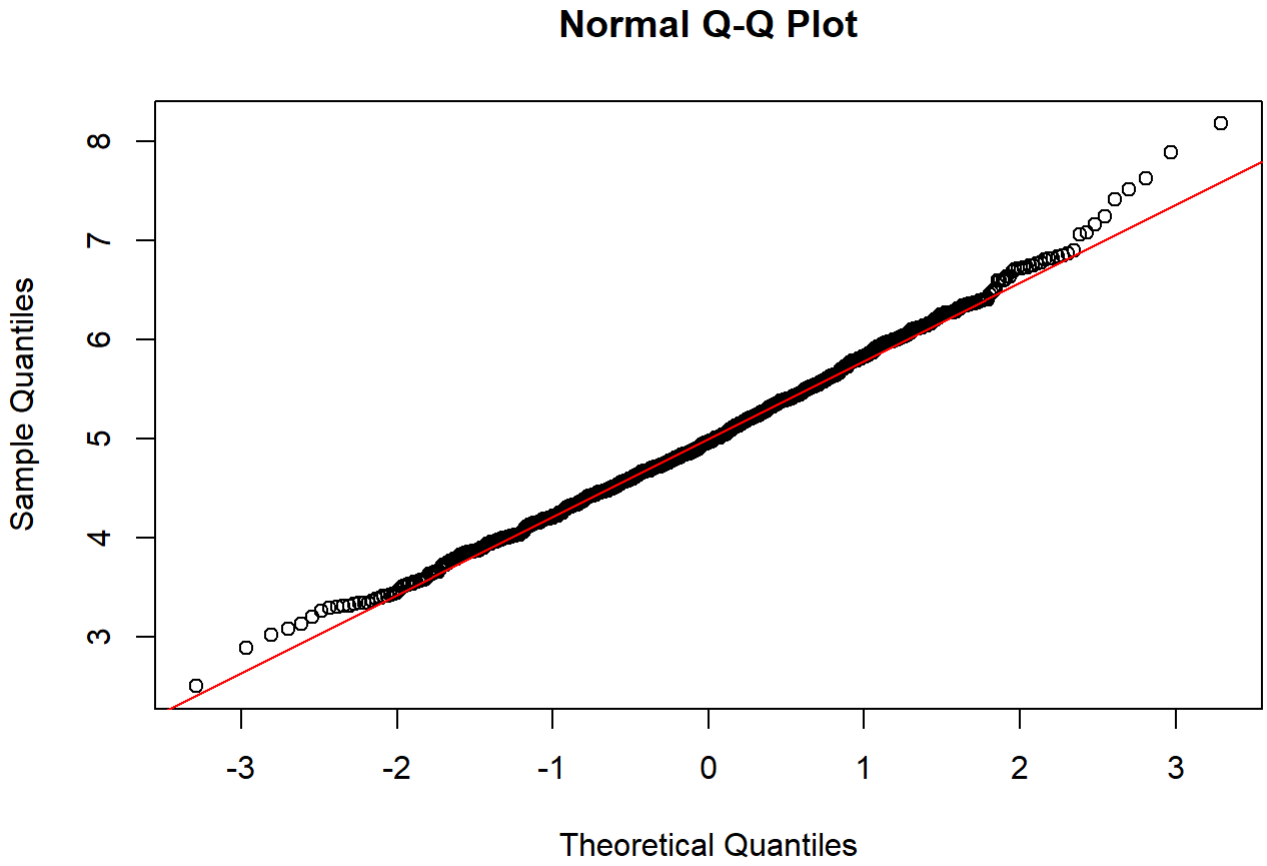
Question 3

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="density")
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)
```



Answer 3

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