# Statistical Inference Project Part 1: Simulation Exercise Instructions

*Susmitha Saripalli*

## 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)
```

## Task

```
# set seed for reproducability
set.seed(0)

# lambda = 0.2
lambda <- 0.2

# 40 samples
n <- 40

# 1000 simulations
sims <- 1000

# simulating the data
dat <- replicate(sims, rexp(n, lambda))

# means of exponentials
means_exponentials <- apply(dat, 2, mean)
```

## Question 1

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

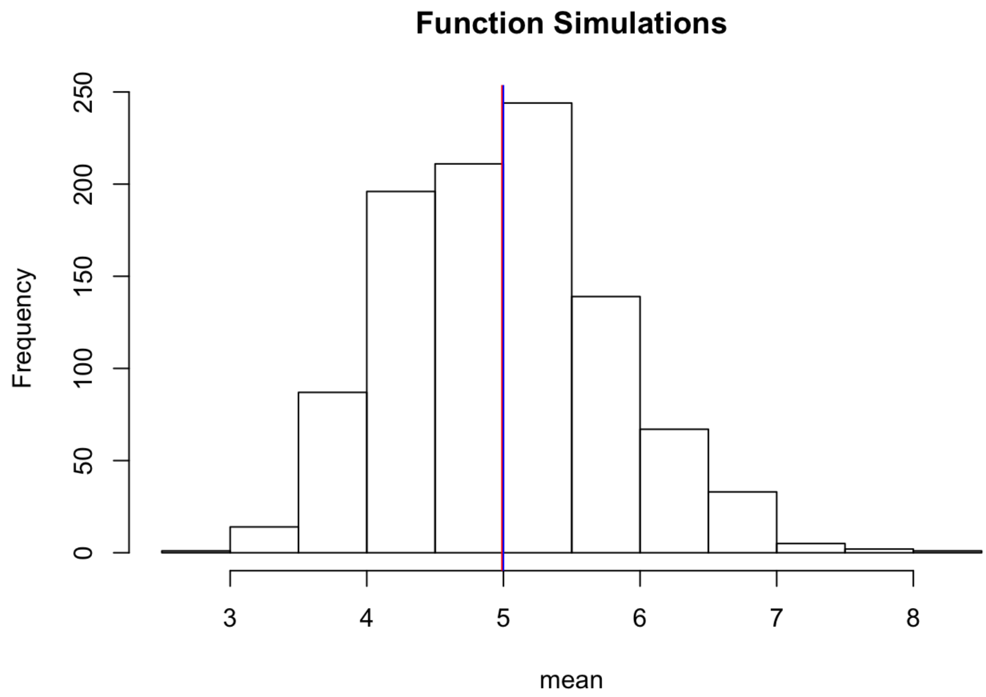
```
# sample mean
sample_mean <- mean(means_exponentials)
sample_mean
```

```
## [1] 4.989678
```

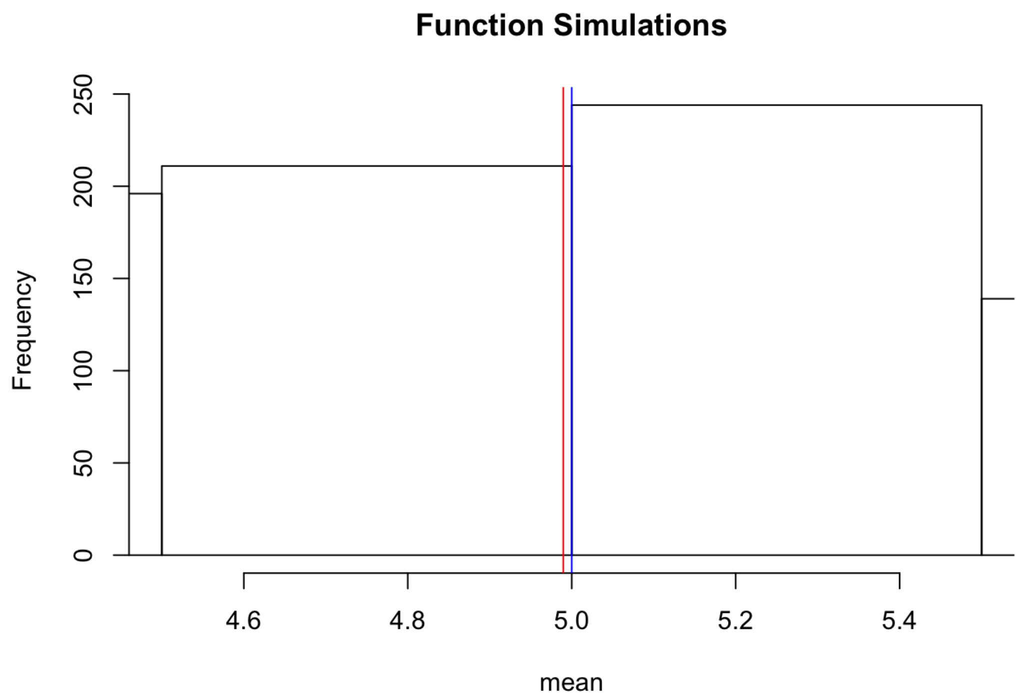
```
# theoretical mean
theoretical_mean <- 1/lambda
theoretical_mean
```

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

```
# illustration
hist(means_exponentials, xlab = "mean", main = "Function Simulations")
abline(v = sample_mean, col = 2)
abline(v = theoretical_mean, col = 4)
```




```
# zoomed in on histogram
hist(means_exponentials, xlab = "mean", main = "Function Simulations", xlim = c(4.5,5.5))
abline(v = sample_mean, col = 2)
abline(v = theoretical_mean, col = 4)
```




The sample mean is 4.989678 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 and the two means are also very close together.

# Question 2

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

```
# stdv of simulated distribution
stdv_dist <- sd(means_exponentials)
stdv_dist
```

```
## [1] 0.7862304
```

```r
# stdv of theoretical expression
stdv_theoretical <- (1/lambda)/sqrt(n)
stdv_theoretical
```

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

```r
# variance of simulated distribution
var_dist <- var(means_exponentials)
var(stdv_dist) #to check tomorrow
```

```
## [1] NA
```

```r
var_dist
```

```
## [1] 0.6181582
```

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

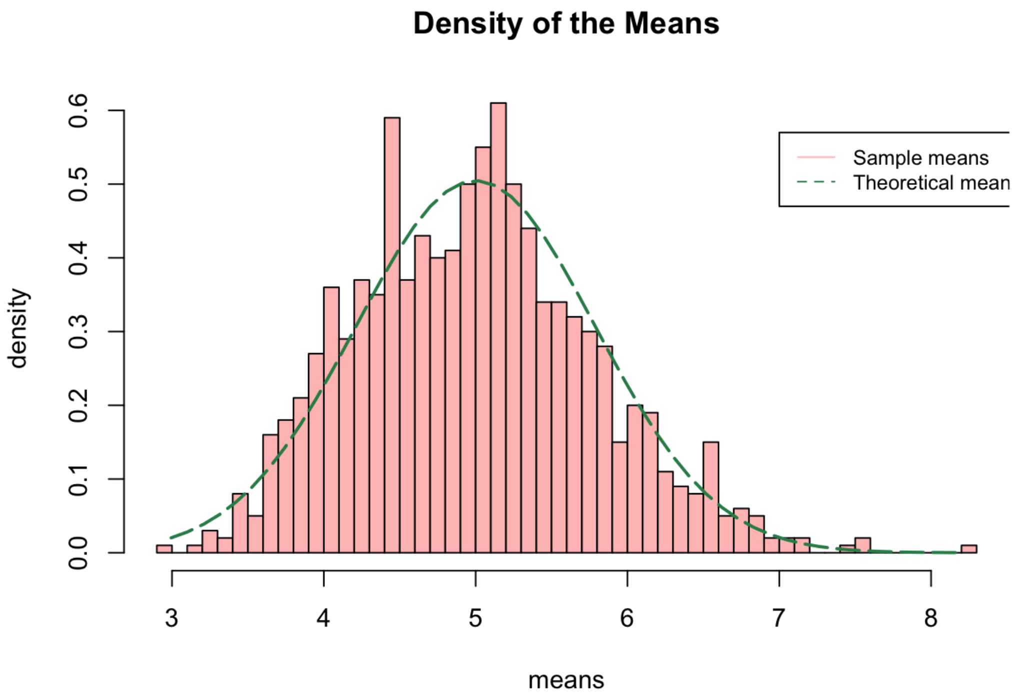
```
## [1] 0.625
```

Standard deviations of the simulated distribution and the theoretical distributions are 0.7862304 and 0.7905694, respectively. The variances are 0.7905694 for the simulated distribution and 0.625 for the theoretical distribution calculated by ((1 / lamda) * (1/n))^2.

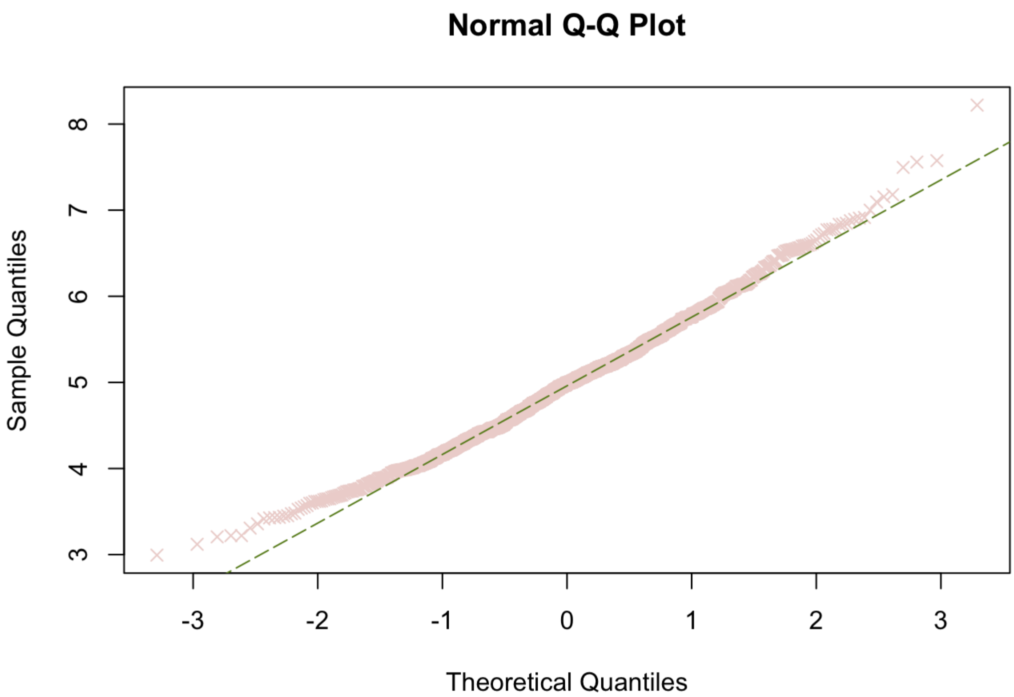
# Question 3

Show that the distribution is approximately normal.

```r
x_theo <- seq(min(means_exponentials), max(means_exponentials), length=50)
y_theo <- dnorm(x_theo, mean=theoretical_mean, sd=stdv_theoretical)
hist(means_exponentials,breaks=n,prob=T,col="rosybrown1",xlab = "means",main="Density of the Means",ylab="de
nsity")
lines(x_theo, y_theo, pch=22, col="seagreen", lty=5, lwd =2)
legend(7,0.57,legend=c("Sample means", "Theoretical mean"),col=c("rosybrown1", "seagreen"), lty=1:2, cex=0.8
)
```




```
# compare the distribution of averages of 40 exponentials to normal distribution
qqnorm(means_exponentials, col = "mistyrose2", pch=4)
qqline(means_exponentials, col = "olivedrab4", lty=5)
```




The distribution of averages of the 40 exponentials is very close to a normal distribution, as seen in the two graphs above.