# Statistical Inference - Part 1: Simulation Exercise

*Saeed Nusri*

*July 29, 2016*

## Synopsis

This project is focused on simulation to explore inference and conducting some simple inferential data analysis. The project consists of two parts:

1. A simulation exercise
2. Basic inferential data analysis

# 1. Simulation

### Setting up the simulation

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

```
#Setting seed for reproduci
set.seed(10)

#Stating parameters for running the simulations
lambda <- 0.2

#40 exponentials
n<-40


# running 1000 simulations
simulated_exponentials <- replicate(1000, rexp(n, lambda))

# calculate mean of exponentials from 1000 simulations
means_exponentials <- apply(simulated_exponentials, 2, mean)
```

### Distributions of Sample and Theoritical Means

To show where the distribution is centered at and to compare it to the theoretical center of the distribution the following is carried out.

```
#Sample mean from 1000 simulations
sample_mean <- mean(means_exponentials)
sample_mean
```

```
## [1] 5.04506
```

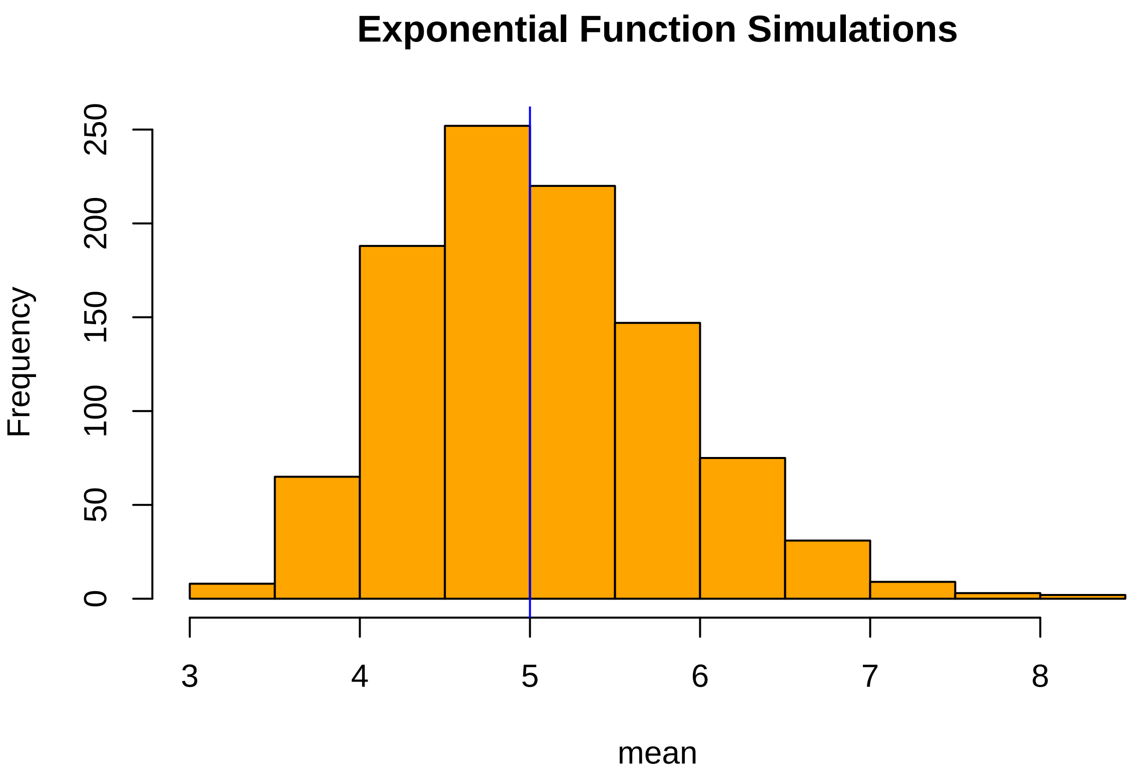
```r
#Calculating the theoritical mean
theo_mean <- 1/lambda
theo_mean
```

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

The results can also be demonstrated through the following histogram where the blue line indicates the theoretical mean

```r
hist(means_exponentials, xlab = "mean", main = "Exponential Function Simulations", col = "orange")

abline(v = theo_mean, col = "blue", b= 4)
```




Therefore, the simulated mean, which is 5.04506 is not away from the theoritical mean of 5.

### Distributions of Sample and Theoretical Standard Deviations

The stanadard deviations were compared. Note that standard deviations of the theoretical distribution is given by

$$\frac{1/\lambda}{\sqrt(n)}$$

and variance is given by

$$(\frac{1}{\lambda} * \frac{1}{\sqrt(n)})^2$$

```
#Sample stabdard deviation
sample_sd <- sd(means_exponentials)
sample_sd
```

```
## [1] 0.7982821
```

```
#Theoritical stabdard deviation
theo_sd <- (1/lambda)/sqrt(n)
theo_sd
```

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

```
#Sample variance
sample_variance <- sample_sd^2
sample_variance
```

```
## [1] 0.6372544
```

```
#Theoritical variance
theo_variance <- ((1/lambda)*(1/sqrt(n)))^2
theo_variance
```

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

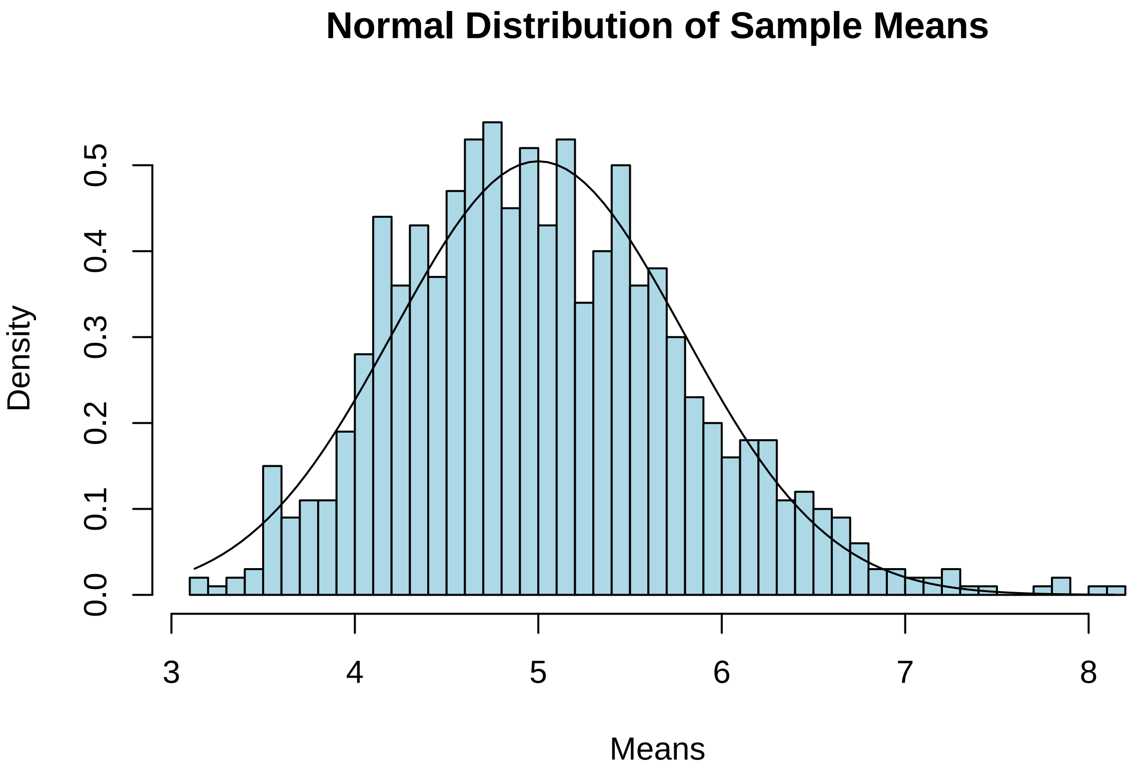
Hence, as seen, the sample variance and standard deviations are quite close to the theoretical values.

### Showing Normal Distribution in Simulated Sample

The normal distribution of the 1000 simulations of 40 exponentials is seen by the follwoing histogram.

```
#Plotting histogram of the sample distribution
hist(means_exponentials,breaks=n, prob = TRUE,
     col="light blue",xlab = "Means", ylab="Density", main="Normal Distribution of Sample Means")

# Adding theoretical distribution lines
x <- seq(min(means_exponentials), max(means_exponentials), length = 100)
y <- dnorm(x, mean=1/lambda, sd= 1/lambda/sqrt(n))
lines(x, y, pch=22, col="black", lty=1)
```




As seen, the sample distribution follows a normal distribution trend due to Central Limit Theorem.