

# Statistical Inference Project Part 1

Andres Mauricio Castro

## Synopsis

In this project, we are going to investigate the exponential distribution in R and compare it with the Central Limit Theorem, and will do some simple inferential data analysis.

## Assignment Description

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should 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.

## Start The Simulation

- set the seed to reproduce the same results
- set lambda to 0.2
- Set 40 samples
- Number of simulations is 1000

```
set.seed(31)
lambda <- 0.2
samples <- 40
simulations <- 1000
```

- Run the simulation
- Calculate mean of exponentials

```
simulatedExponentials <- replicate(simulations, rexp(samples, lambda))
meansExponentials <- apply(simulatedExponentials, 2, mean)
```

## Question 1

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

Calculate of the sample mean and the theoretical mean:

```
sampleMean <- mean(meansExponentials)
sampleMean

## [1] 4.993867
```

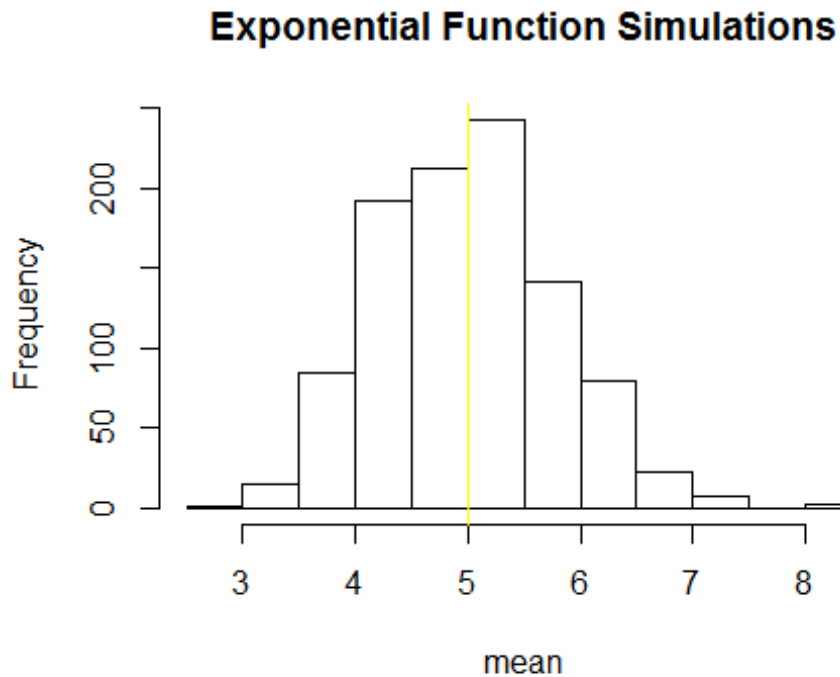
```

theoreticalMean <- 1/lambda
theoreticalMean

## [1] 5

hist(meansExponentials, xlab = "mean", main = "Exponential Function
Simulations")
abline(v = sampleMean, col = "blue")
abline(v = theoreticalMean, col = "yellow")

```



From the histogram, we can see the theoretical mean is very close to the sample mean. The difference is almost imperceptible.

## Question 2

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

```

sdDistribution <- sd(meansExponentials)
sdDistribution

## [1] 0.7931608

theoreticalSD <- (1/lambda)/sqrt(samples)
theoreticalSD

## [1] 0.7905694

```

```

varianceDistribution <- sdDistribution^2
varianceDistribution

## [1] 0.6291041

theoreticalVariance <- ((1/lambda)*(1/sqrt(samples)))^2
theoreticalVariance

## [1] 0.625

```

We can see the Standard Deviation of the distribution is close to the theoretical Standard Deviation; same for the variance.

### Question 3

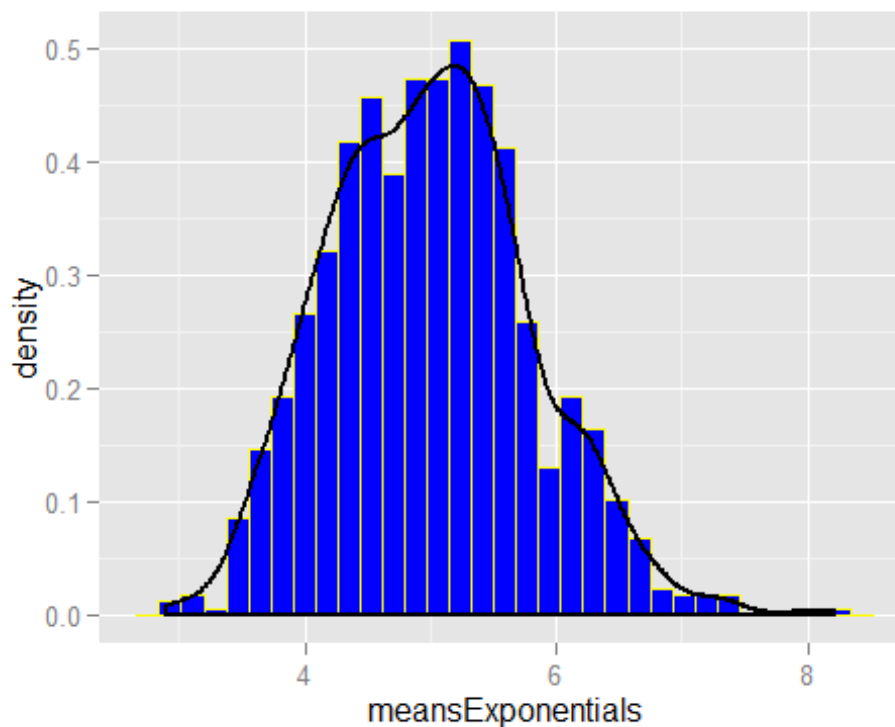
Show that the distribution is approximately normal.

Here, we create an approximate normal distribution and see how the sample data fits to it.

```

library(ggplot2)
meansDataFrame <- data.frame(meansExponentials);
m <- ggplot(meansDataFrame, aes(x =meansExponentials))
m <- m + geom_histogram(aes(y=..density..), colour="yellow",
fill = "blue")
m + geom_density(colour="black", size=1);

```



We can see above, the distribution is close to fit a bell curve.

By comparing the confidence intervals, we can see the theoretical ones are close to the distribution ones.

```
actualConfidenceInterval <- round (mean(meansExponentials) + c(-  
1,1)*1.96*sd(meansExponentials)/sqrt(samples),3)  
actualConfidenceInterval
```

```
## [1] 4.748 5.240
```

```
theoreticalConfidenceInterval <- theorethicalMean + c(-  
1,1)*1.96*sqrt(theoreticalVariance)/sqrt(samples);  
theoreticalConfidenceInterval
```

```
## [1] 4.755 5.245
```

Compare the distribution of averages of 40 exponentials to a normal distribution

```
qqnorm(meansExponentials); qqline(meansExponentials)
```

