# Course Project Part 1: Simulation Exercise

Aashay Binaykia

11 June 2017

## Overview

This project is intended to investigate the exponential distribution in R and compare it with the Central Limit Theorem. The central limit theorem (CLT) states that the means of random samples drawn from any distribution with mean m and variance s^2 will have an approximately normal distribution with a mean equal to m and a variance equal to s^2 / n.

## Simulation

Investigating the distribution of averages of 40 exponentials. A total of 1000 Simulations will be performed and stored in the variable 'means'

```
n <- 40
lambda <- 0.2
means <- numeric()
sim <- 1000 ## There would be a thousand simulations
for(i in 1:sim){
  means <- c(means,mean(rexp(n,rate = lambda)))
}
```

## Analyzation

Analyzing Sample values vs Theoretical Values

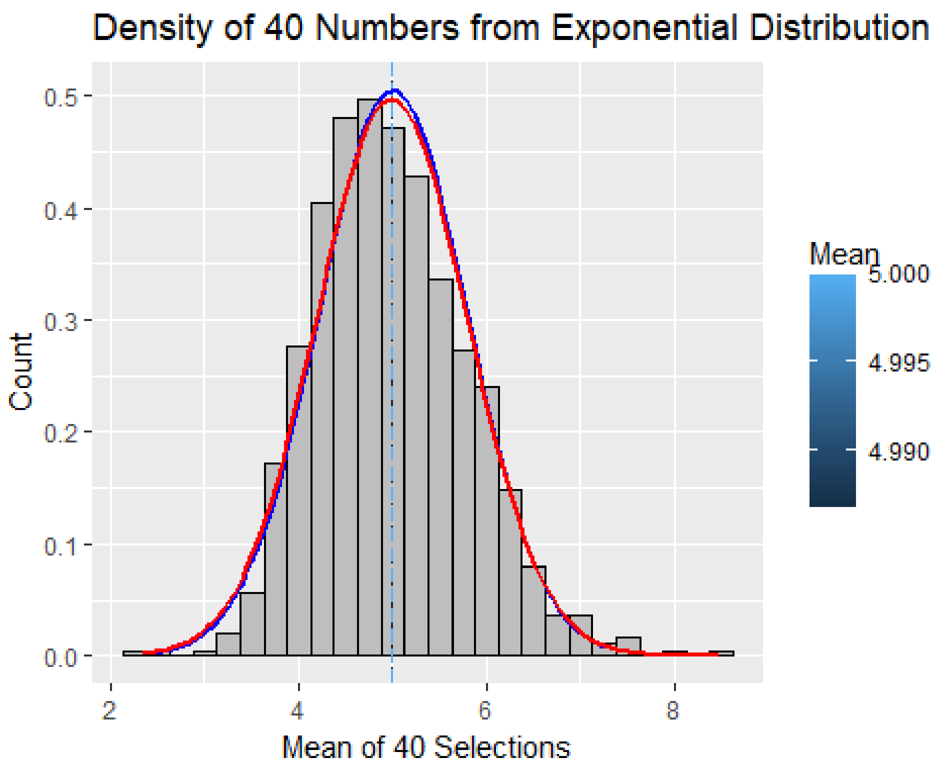
```
sampleMean<-mean(means)
thMean <- 1/lambda
sampleVar <- var(means)
thVar <- 1/(lambda^2*n)
mat<-data.frame(Title=c("Sample Values", "Theoretical Values"),
Mean=c(sampleMean, thMean),Variance=c(sampleVar,thVar))
```

Printing the Results Below:

```
mat

##                Title     Mean  Variance
## 1      Sample Values 4.986537 0.6455762
## 2 Theoretical Values 5.000000 0.6250000
```

Thus we see that the sample values are almost equal to the theoretical values
Plotting these values:

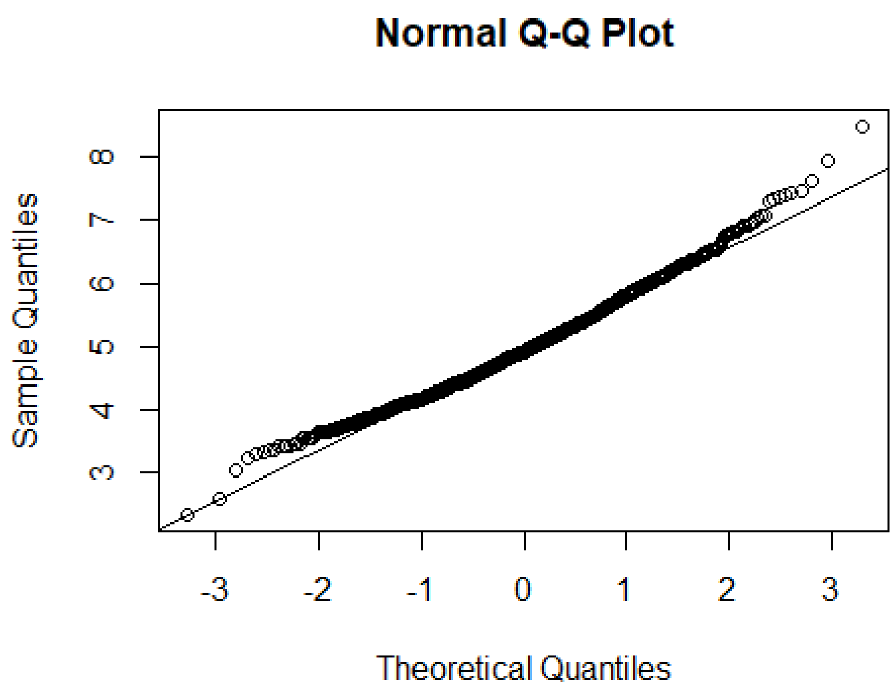
```r
require(ggplot2)

## Loading required package: ggplot2

ggplot(mapping = aes(x = means))+
  geom_histogram(aes(y = ..density..),color="black",fill="grey",binwidth =
.25)+
  labs(title="Density of 40 Numbers from Exponential Distribution", x="Mean
of 40 Selections", y="Count")+
  geom_vline(data =
mat,aes(xintercept=mat$Mean,color=Mean),linetype=mat$Mean,show.legend =
TRUE)+ # add a line for the actual mean
  stat_function(fun=dnorm,args=list(mean=thMean, sd=sqrt(thVar)),color =
"blue", size = 1.0)+
  stat_function(fun=dnorm,args=list(mean=sampleMean,
sd=sqrt(sampleVar)),color = "red", size = 1.0)
```




*The histogram shows that the simulated data is almost a normal distribution*. The graph above shows the sample mean and the theoritical mean of the simulated data. Both of which are almost indistinguishable. The red line shows the normal distribution formed with sapmle values, while the blue line shows the actual normal distribution.

To further establish that 'Means' forms a distribution which is almost normal, we use the qqnorm and qqline function. qqnorm is a generic function the default method of which produces a normal QQ plot of the values in y. qqline adds a line to a "theoretical", by default normal, quantile-quantile plot which passes through the probs quantiles, by default the first and third quartiles.

```r
qqnorm(means)
qqline(means)
```




In the plot, we see that the sample quantiles and theoritical quantiles are almost similar, thus establishing the hypothesis.