Statistical Inference: Project report

## Part 1 -Introduction

This project investigates the exponential distribution in R and compares it with the Central Limit Theorem. 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 1/lambda. Lambda is se to 0.2 for all of the simulations. Simulations are repeated thousand times with samples of 40 numbers generated with exponential distribution.

It will ilustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. It should

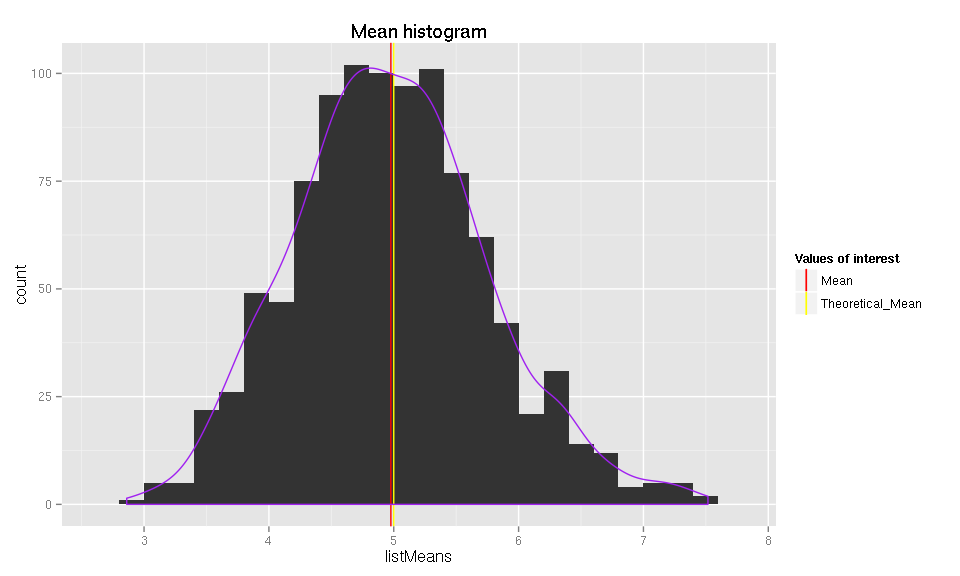
1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

## Calculating sample's mean and variation

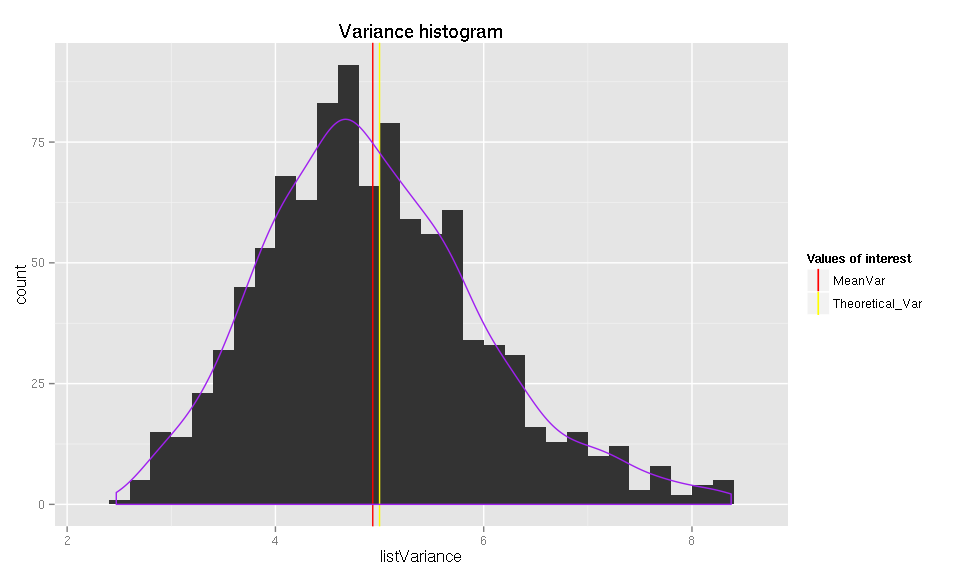
The following r code takes 40 samples with exponential distribution and calculates its variance and mean. Then repeats that simulation a thousand times and plots variance and mean histograms of the given data.

library(data.table)  
library(ggplot2)  
library(gridExtra)  
  
listMeans <- data.table(unlist(lapply(1:1000, function (x) { mean(rexp(40, 0.2)) })))  
listVariance <- data.table(unlist(lapply(1:1000, function (x) { sd(rexp(40, 0.2)) })))  
setnames(listMeans, "listMeans")  
setnames(listVariance, "listVariance")  
Mean <- mean(listMeans[,listMeans])  
MeanVariance <- mean(listVariance[,listVariance])

ggplot(listMeans, aes(x=listMeans)) +  
 geom\_histogram(bin=0.2) +   
 ggtitle("Mean histogram") +  
 geom\_vline(aes(xintercept = Mean,colour = "Mean"),show\_guide = TRUE) +  
 geom\_vline(aes(xintercept = 1/0.2,colour = "Theoretical\_Mean"),show\_guide = TRUE) +  
 geom\_density(aes(y = 0.2 \* ..count..), colour = "purple") +  
 scale\_colour\_manual(name="Values of interest", values=c(Theoretical\_Mean="yellow",Mean="red"))



ggplot(listVariance, aes(x=listVariance)) +  
 geom\_histogram(bin=0.2) +   
 ggtitle("Variance histogram") +  
 geom\_vline(aes(xintercept = MeanVariance,colour = "MeanVar"),show\_guide = TRUE) +  
 geom\_vline(aes(xintercept = 1/0.2, colour = "Theoretical\_Var"),show\_guide = TRUE) +  
 geom\_density(aes(y = 0.2 \* ..count..), colour = "purple") +  
 scale\_colour\_manual(name="Values of interest", values=c(Theoretical\_Var="yellow",MeanVar="red"))



We can see that distribution of sample means and variations closely resembles normal distribution and that sample's expected mean of 4.9769975 and Variance of 4.9347721 are very close to theoretical values of 5.