Statistical Inference Project Part 1: Simulation Exercise

## Overview

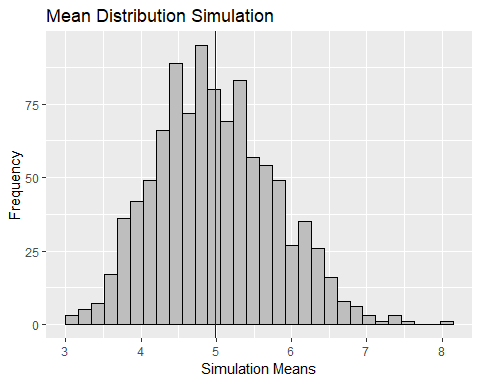
This project investigate the exponential distribution in R and compare it to the Central Limit Theorem. To do this we investigate the distribution of averages of 40 exponentials by doing a thousand simulations.

library(ggplot2)  
lambda <- 0.2  
n <- 40  
simulations <- 1000  
set.seed(1)

## Constructing and Plotting

SimulationMatrix <- matrix(rexp(simulations\*n,lambda),nrow = simulations, ncol = n)  
SimulationMean <- rowMeans(SimulationMatrix)  
SimulationData <- data.frame(cbind(SimulationMatrix,SimulationMean))  
  
ggplot(data = SimulationData, aes(SimulationData$SimulationMean)) + geom\_histogram(col = 'black', fill = 'gray') + labs(title = 'Mean Distribution Simulation',x = 'Simulation Means', y = 'Frequency') + geom\_vline(aes(xintercept = mean(SimulationData$SimulationMean)), color = 'blue')

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



We can see that the Histogram is very close toa normal bell curve and the mean is around 5.

## Sample Mean vs Theoretical Mean

ActualMean <- mean(SimulationMean)  
TheoMean <- 1/lambda  
TheoMean - ActualMean

## [1] 0.009974799

The difference of the simulated mean and the actual mean is small (0.04).

## Sample variance vs Theoretical variance

ActualVar <- var(SimulationMean)  
TheoVar <- ((1/lambda)^2)/n  
TheoVar - ActualVar

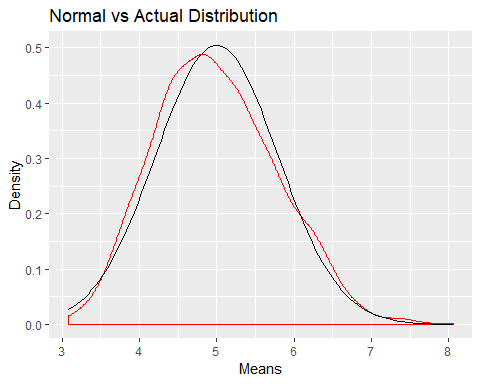
## [1] 0.007292825

Just like the mean above, the difference between the theoretical and simulated variance is very small (0.0082)

## Distribution

We compare the distribution of our simulated data vs the normal distribution and see if they differ to much.

ggplot(data = SimulationData,aes(SimulationData$SimulationMean)) + geom\_density(color = 'red') + stat\_function(fun = dnorm,args = list(mean=(1/lambda),sd = ((1/lambda))/sqrt(n))) + labs(title = 'Normal vs Actual Distribution', x = 'Means', y = 'Density')



We can see that it does not differ that much from the normal distribution.

## Comparing 95% confidence interval (theoretical vs simulated)

ActualConfInt <- ActualMean + c(-1,1) \* 1.96 \* sqrt(ActualVar) / sqrt(n)  
print(ActualConfInt)

## [1] 4.746459 5.233592

TheoConfInt <- TheoMean + c(-1,1)\*1.96\*sqrt(TheoVar)/sqrt(n)  
print(TheoConfInt)

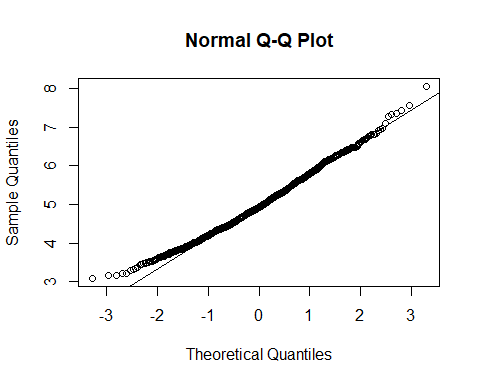
## [1] 4.755 5.245

Both of them are approximately the same

## QQ Plot

Finally we can create a QQ plot to see if our simulated data follows a straight line.

qqnorm(SimulationMean)  
qqline(SimulationMean)



Hence we can conclude that our simulated distribution is approximately normal