# Simulating Exponential Distributions

## Overview

* The following report shows that the sample mean and the sample standard deviation are good estimator of the population mean and standard deviation respectively.
* This report also verifies the Central Limit Theorem.

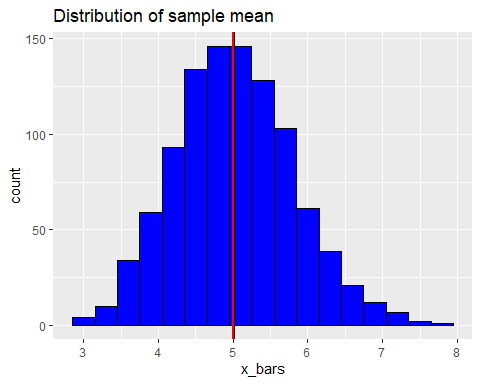
## Simulations

The following code takes sample size of 40 and calculates the sample mean. This is simulated 1000 times and stored in the vector x\_bars.  
This also calculated the standard deviation of the simulated data and stores it in the vector s.

x\_bars = sapply(1:1000,function(x) mean(rexp(40,0.2)))  
library(ggplot2)  
x\_bars = data.frame(x\_bars)  
s = sapply(1:1000,function(x) sd(rexp(40,0.2)))  
s = data.frame(s)

## Sample Mean versus Theoretical Mean

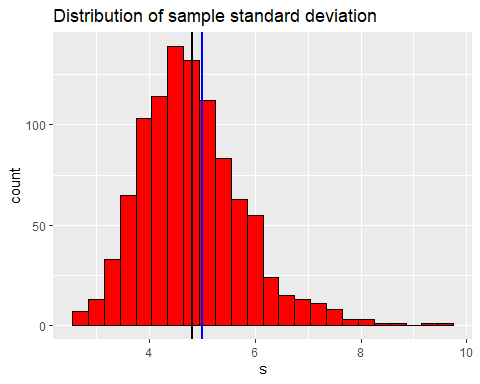
g = ggplot(x\_bars,aes(x\_bars))  
mu\_est = mean(x\_bars$x\_bars)  
diff = abs(mu\_est-5)  
g+geom\_histogram(binwidth = 0.3,fill = "blue",color = "black")+geom\_vline(xintercept = mu\_est,size = 1)+geom\_vline(xintercept = 5,size = 1,col = "red")+labs(title = "Distribution of sample mean")



Note that the histogram of the sample mean of the simulated data is plotted along with the theoretical mean.  
The red line indicates the theoretical mean and the black line is the mean of the simulated data.  
This figure clearly shows that the distribution of the simulated data is centered around the theoretical mean. The difference between the sample mean and the theoretical mean is **0.0223431**, which is fairly small.

## Sample Variance versus Theoretical Variance

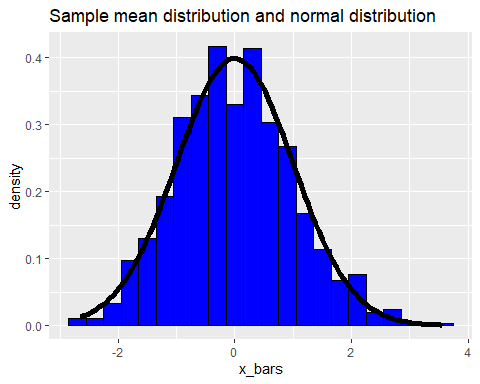
g = ggplot(s,aes(s))  
sd\_est = mean(s$s)  
diff2 = abs(sd\_est-5)  
g + geom\_histogram(binwidth = 0.3,color = "black",fill = "red") + geom\_vline(xintercept = 5,color = "blue",size = 1) + geom\_vline(xintercept = mean(s$s),size = 1)+labs(title = "Distribution of sample standard deviation")



Note that the histogram of the sample standard deviation of the simulated data is plotted along with the theoretical standard deviation.  
The blue line indicates the theoretical standard deviation and the black line is the standard deviation of the simulated data.  
This figure clearly shows that the distribution of the simulated data is centered around the theoretical standard deviation. The difference between the sample standard deviation and the theoretical standard deviation is *0.1865923* and thus the difference between their variances is the square of it which is **1.8311068**, which is fairly small.

##Distribution

x\_sn = (x\_bars-mu\_est)/sd(x\_bars$x\_bars)  
g = ggplot(x\_sn,aes(x = x\_bars))  
g+geom\_histogram(aes(y = ..density..),binwidth = 0.3,color = "black",fill = "blue")+stat\_function(fun = dnorm,args = list(mean = mean(x\_sn$x\_bars),sd = sd(x\_sn$x\_bars)),size = 2)+labs(title = "Sample mean distribution and normal distribution")



This plot shows that the sample mean distribution is approximately equal to the normal distribution.  
As the normal distribution almost overlays the sample mean plot.