# Statistical Inference Course Project

*Sohail Munir Khan*

*14 June 2015*

## Overview

Central Limit Theorem states that any large number of iterations of independent and identically distributed (i.i.d.) each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution. In this case we are exploring expontential distribution rexp(n, lambda) with fixed lambda = 0.2 which will have both mean and standard deviation as (1 / lambda) => (1 / 0.2) = 5

## Simulations

We shall be doing 1000 simulations each with a generation of 40 exponential values. Once we have Please find the simulation code below:

```
# Clean the environment
rm(list = ls())

# Set simulation parameters
#  n (number of exponentials generated per simulation run)
n <- 40
#  lambda (rate for exponential generation)
lambda <- 0.2
#  nosim (number of simulations)
nosim <- 1000

# Produces a massive dataMatrix with 1000 simulations of 40 exponentials each
dataMatrix <- matrix(data = rexp(n * nosim, rate = lambda),
                     nrow = nosim, ncol = n)
# Each simulation has now its own mean. 1000 * 1 means stored
dataMatrixMeans <- apply(dataMatrix, 1, mean)
# Mean for all means of the simulation
s_mean <- mean(dataMatrixMeans)
s_mean
```

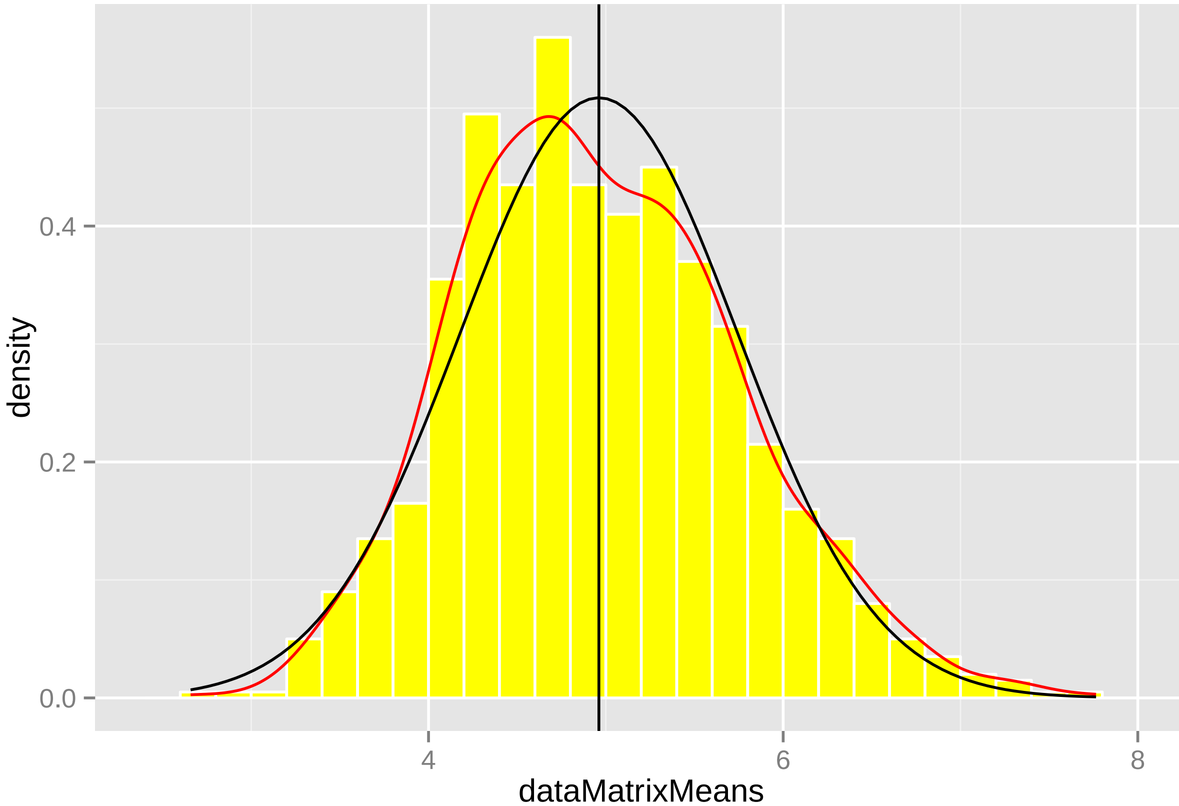
```
## [1] 4.960699
```

```
# Standard Deviation for all 1000 simulations calculated
s_sd <- sqrt(sum((dataMatrixMeans - s_mean)^2)/length(dataMatrixMeans))
s_sd
```

```
## [1] 0.784093
```

```
# To be able to create ggplot graphs, I have created a data.frame df that
#  will hold the data to be shown
df <- data.frame(dataMatrixMeans)
```

Histogram for the distribution of means & standard-deviation of simulated means:




---

## Sample Mean versus Theoretical Mean

We have calculated our simulated mean as 4.9606995 while the theoretical mean is 5. This shows that the deviation of simulated mean with theoretical mean is only 3.9300504%

---

## Sample Variance versus Theoretical Variance

We have calculated our simulated variance as 0.6148019 [s_sd^2] while the theoretical variance is 0.625 [(theoretical sd)^2 / n == (1/lambda)^2/n] to calculate simulated variance). This shows that the deviation of simulated variance with theoretical variance is only 1.0198117%

---

## Distribution

As you can see that the figure shows how closely related the red line (simulated distribution) is to the standard normal distribution (black line). The simulated and theoretical distributions are almost touching each other implying very accurate calculations of mean and standard deviation