# Statistcal Infererence - Course Project

*Luke Moss*

*Sunday, April 26, 2015*

## Overview

This project will investigate the exponential distribution in R and compare it with the central Limit Theorem by conducting 1000 simulations with the distribution of 40 exponentials. The goal is to illustrate the properties of the distribution of the mean of the 40 exponentials. The following will be shown and discussed

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.

## Simulations

First, set the seed to ensure everything is reproducible. Then, run the simulations to generate the data.

```
set.seed(5)
lambda <- 0.2;
expon <- 40;
sim <- 1000
data <- data.frame(
    x  = apply(matrix(rexp(expon * sim, lambda), sim), 1, mean),
    sz = factor(rep(expon, sim)))
```

## Sample Means vs Theoretical Mean

The first objective is to compare the sample mean to the theoretical mean. The sample mean is calculated against the simulated data. The theoretical mean is calculated against the exponential distribution which is 1/lambda (`1/lambda`).

```
sim_mean <- mean(data$x);
theo_mean <- 1/lambda;
c("Simulation Mean"=sim_mean, "Theoretical Mean"=theo_mean)
```

```
##  Simulation Mean Theoretical Mean
##         5.043053         5.000000
```

### Result

The sample mean (5.043053) and the theoretical mean (5) are very close, but not identical.

## Sample Variance vs Theoretical Variance

The second objective is to compare the sample variance to the theoretical variance. The sample variance is calculated in R using the standard deviation function. The theoretical variance is calculated using the Central Limit Theorem ($\sigma/\sqrt{n}$) where $\sigma$ = equals `1/lambda` and n is the number of exponentials (40).

```
sim_variance <- sd(data$x);
theo_variance <- (1/lambda) / sqrt(expon);
c("Simulated Variance"=sim_variance, "Theoretical Variance"=theo_variance)
```

```
##   Simulated Variance Theoretical Variance
##            0.8238373            0.7905694
```

### Result

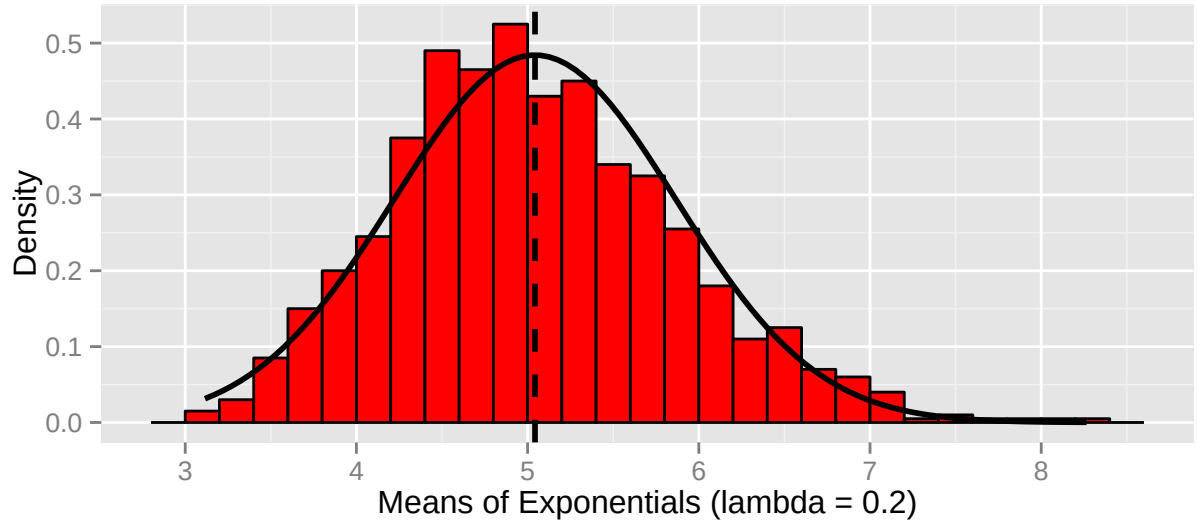
Similar to the means, the sample variance (0.8238373) and the theoretical variance (0.7905694) are close, but not the same.

## Distribution

The next objective is to show that the distribution of the simulated data is approximately normal. To illustrate this I will make two plots: A comparison to a normal probability density distribution curve and a q-q plot.

### Comparison to a Normal Probability Density Distrubution Curve

The plot below shows the simulated data against a normal probability density distribution curve. The parameters for this plot are the `sim_mean` and the `sim_variance`. The profile of the density curve very closely matches the profile of the simulation data.



**Quantile by Quantile Comparison**

The plot below compares, by quantile, the simulated data with a normal standard distribution. This is known as a q-q plot. To build this comparison, the distribution of the simulation mean is calculated using $\bar{Z} = \frac{\bar{X} - \mu_X}{\sigma_X}$. The plot below illustrates that much of the simulated data sits very close to the line with exceptions at the far left and far right. This indicates a very close match with the theoretical quantiles.