# Statistical Inference Course Project Part 1

Sheida Vanouei

July 23, 2018

## Overview

This is the first part of the Statistical Inference Course Project from Coursera. It is demonstrated that the distribution proves the Central Limit Theorem. This assignment will make calculation and plots and compare confidence intervals, and eventually proves that the distribution is approximately normal.

## Tasks

In this project you will investigate the exponential distribution in R and compare 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. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should - Show the sample mean and compare it to the theoretical mean of the distribution. - Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. - Show that the distribution is approximately normal.

In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

## Analysis

## Setting

The mean of exponential distribution is 1/lambda and the standard deviation is also also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s.

```
set.seed(539)
lambda <- 0.2
n <- 40
size <- 1000
simulate <- replicate(size, rexp(n, lambda))
means_expo <- apply(simulate, 2, mean)
```

## Meam Comparison

The mean of the sample means is 5.000589 and the theoritical mean is 5! The sample mean and the theoretical mean (expected mean) are very close.

## Variance Comparison

```
sample_var <- var(means_expo)
THEOvar  <- (1 / lambda)^2 / (n)
sample_sd <- sd(means_expo)
THEOsd   <- 1/(lambda * sqrt(n))
```

The variance of the sample means is 0.6095467 and the thoeretical variance of the distribution is 0.625. Both variance values are very close to each other.Morever, the standard deviation of the sample means is 0.7807347 and the thoeretical standard deviation of the distribution is 0.7905694. Both standard deviation values are very close to each other.

## Distribution

```
library(ggplot2)
plot1 <- data.frame(means_expo)
p <- ggplot(plot1, aes(x =means_expo))
p <- p + geom_histogram(aes(y=..density..), colour="black",fill =
"lightblue")
p <- p + labs(title = "Distribution of means of 40 Samples", x = "Mean of 40
Samples", y = "Density")
p <- p + geom_vline(aes(xintercept = sample_mean, colour = "sample"))
p <- p + geom_vline(aes(xintercept = THEOmean, colour = "black" ))
p <- p + stat_function(fun = dnorm, args = list(mean = sample_mean, sd =
sample_sd), color = "green", size = 2.0)
p <- p + stat_function(fun = dnorm, args = list(mean = THEOmean, sd =
THEOsd), colour = "red", size = 2.0)
p

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

## Distribution of means of 40 Samples
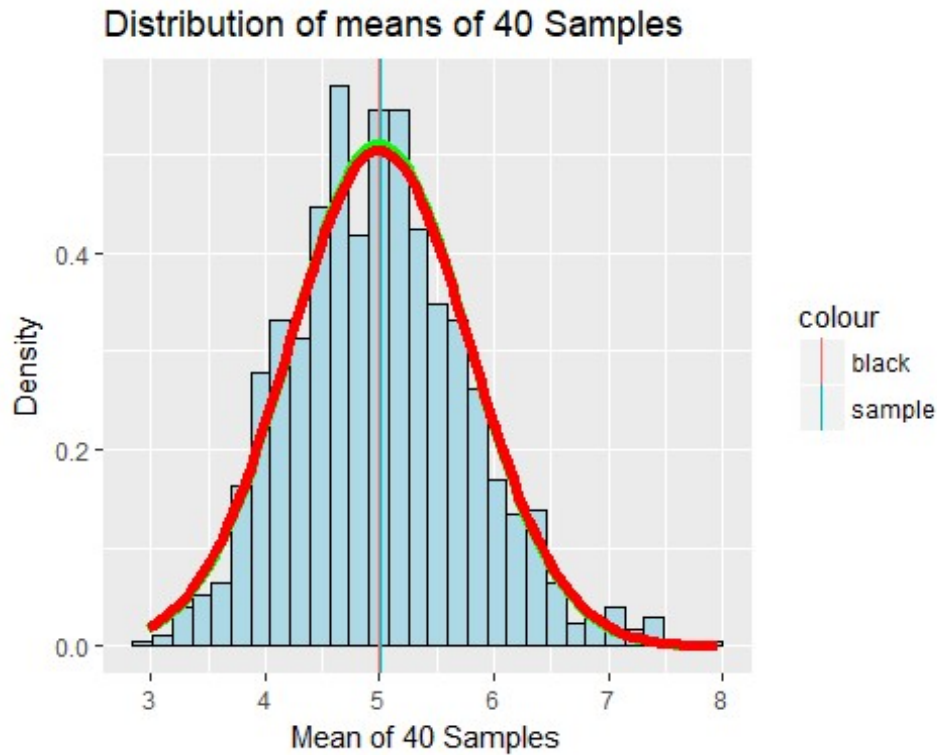


The density of the actual data is shown by the light blue bars. The theoretical mean and the sample mean are so close that they nearly overlap. The "red" line shows the normal curve formed by the the theoretical mean and standard deviation. The "green" line shows the curve formed by the sample mean and standard deviation. As you can see from the graph, the distribution of means of 40 exponential distributions is close to the normal distribution with the expected theoretical values based on the given lambda