# Statistical Inference Course Project - Part 1

*Chandrasekar Ganesan*

*September 1, 2017*

## Contents



### Overview

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

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.

### Data Simulation

```r
require(ggplot2)
```

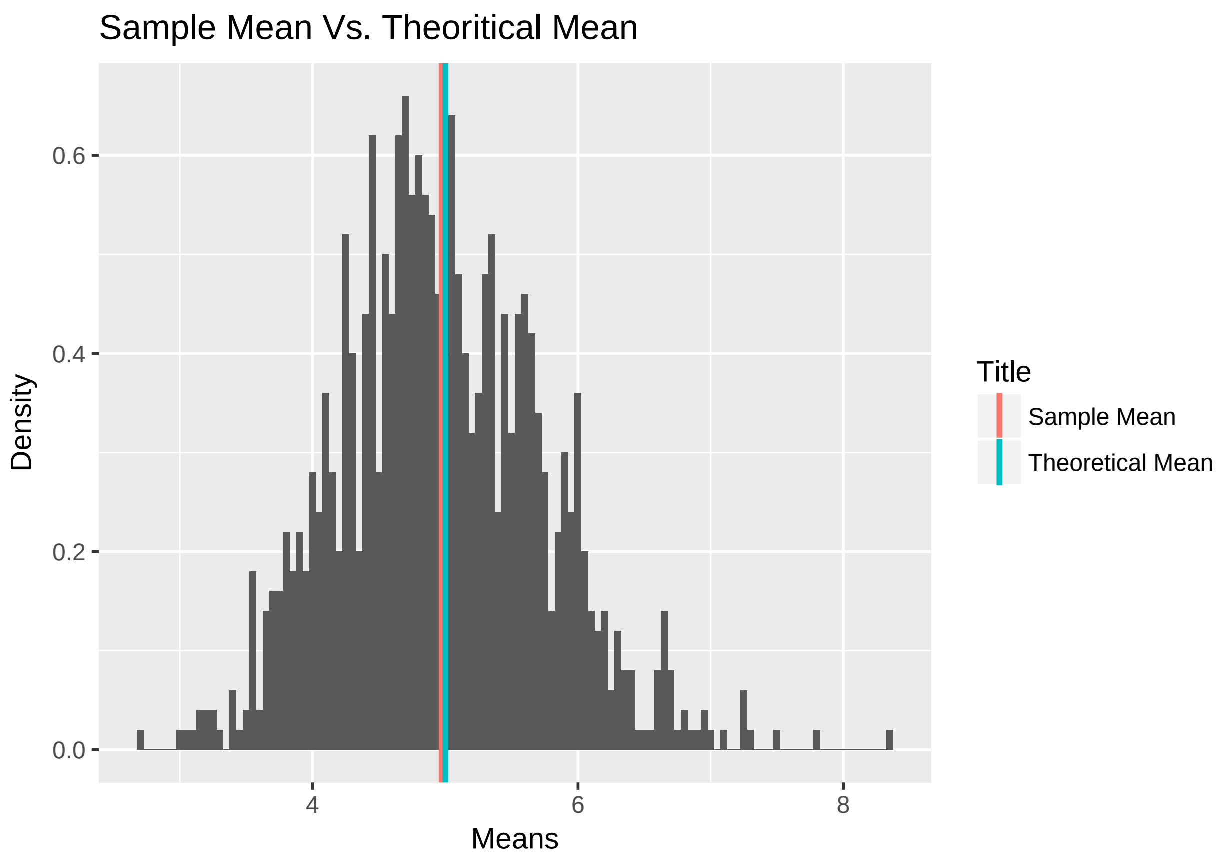
```
## Loading required package: ggplot2
```

```r
lambda = 0.2
n = 40
simulations = 1:1000
set.seed(12345)
simDf <-
data.frame(x = sapply(simulations, function(x) {
    mean(rexp(n, lambda))
}))
```

### Sample Mean Vs. Theoritical Mean

```r
mean_data <- data.frame(Title=c("Sample Mean", "Theoretical Mean"), vals=c(colMeans(simDf), 1/lambda))
g = ggplot(data = simDf, aes(x = x))
g = g + geom_histogram(binwidth = 0.05, aes(y = ..density..))
g = g + geom_vline(data = mean_data, mapping=aes(xintercept = vals, colour=Title), size=1, show.legend=
```

```r
g = g + labs(title = "Sample Mean Vs. Theoritical Mean")
g = g + labs(x="Means") + labs(y = "Density")
g
```




**Conclusion**

As shown below the theoritical mean of the distribution is **5.00** and sample mean is **4.97**

```
##                Title     vals
## x      Sample Mean 4.971972
##   Theoretical Mean 5.000000
```

## Sample Vs. Theoritical Std.Dev & Variance

```r
sampleSD <- sd(simDf$x)
sampleVariance <- sampleSD ^ 2

theoriticalSD <- (1/lambda) / sqrt(n)
theoriticalVariance <-theoriticalSD ^ 2

resultDF = data.frame(c("Sample", "Theoritical"),
                      c(sampleSD, theoriticalSD),
                      c(sampleVariance, theoriticalVariance))
```

```r
colnames(resultDF) <- c("","Std.Dev","Variance")
```

### Conclusion

From the table shown below the difference between theoritical and sample values for Std.dev and variance are very less.

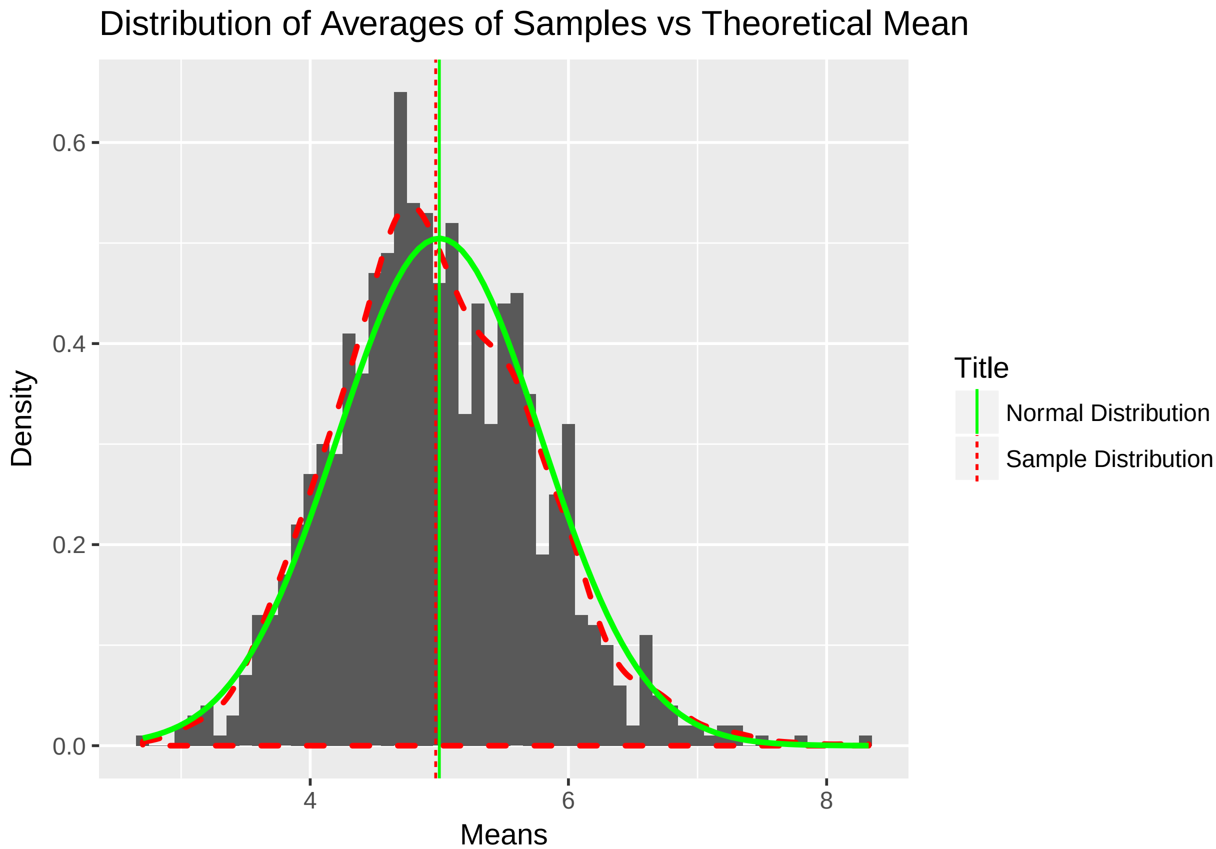
```
##                 Std.Dev  Variance
## 1      Sample 0.7716456 0.5954369
## 2 Theoritical 0.7905694 0.6250000
```

## Distribution

```r
df <- data.frame(Title=c("Sample Distribution", "Normal Distribution"), Vals=c(colMeans(simDf), 1/lambda

g = ggplot(data = simDf, aes(x = x))
g = g + geom_histogram(binwidth=0.1, aes(y=..density..))
g = g + geom_density(color="red", size=1, linetype=2)
g = g + geom_vline(data=df[df$Title == "Sample Distribution", ], mapping=aes(xintercept=df[df$Title ==
g = g + stat_function(fun=dnorm, args=list(mean=df$Vals[df$Title == "Normal Distribution"], sd=theoriti
g = g + geom_vline(data=df[df$Title == "Normal Distribution", ], mapping=aes(xintercept=df[df$Title ==
g = g + guides(linetype=guide_legend(override.aes=list(colour = c("green","red"))))
g = g + labs(title = "Distribution of Averages of Samples vs Theoretical Mean") + labs(x="Means") + lab

g
```




**Conclusion**

The normal distribution line refers to lambda = 0.2. The sample distribution refers to the averages of similated samples. The graph shows that the two distribution lines are well aligned thus the distribution of simulated data is approximately normal.