# Statistical Inference Course Project Part 1

*Koji*

*2018/7/20*

## 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.

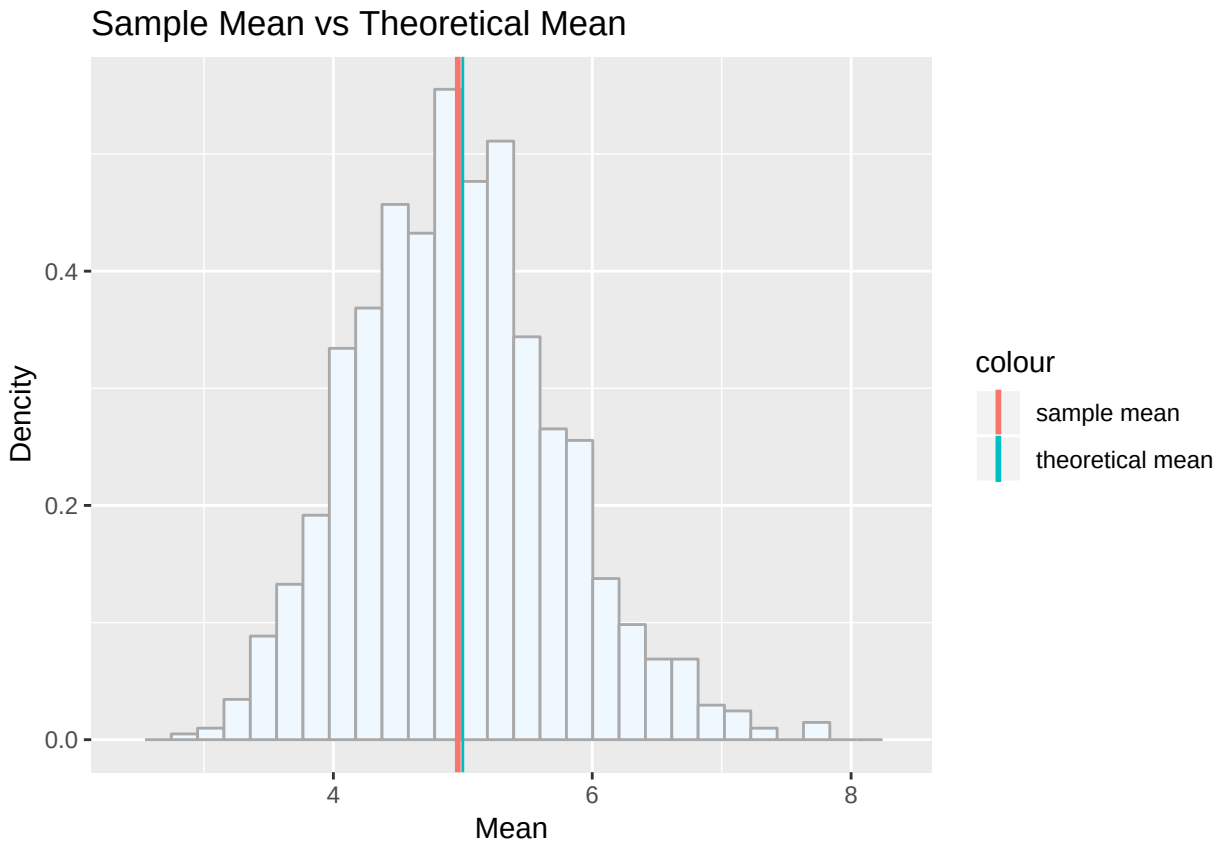
## Simulations

Simulate the distribution of 1000 averages of 40 random exponentials.

```
# set seed for reproducability
set.seed(3301)

n = 40 # number of exponentials
lambda = 0.2 # Set lambda = 0.2 for all of the simulations
samples <- 1000 # number of simulations

# do 1000 random simulations
mns = NULL
for (i in 1 : samples) {
        mns = c(mns, mean(rexp(n, lambda)))
}
```

## Sample Mean versus Theoretical Mean

```
# the means of simulations
sample_mean <- mean(mns)
# theoritical mean
theoretical_mean <- 1 / lambda

df <- data.frame(mns)
g <- ggplot(df, aes(x = mns))
g <- g + geom_histogram(aes(y = ..density..), fill = "aliceblue", color = "darkgrey")
g <- g + xlim(range(density(mns)$x))
g2 <- g + geom_vline(aes(xintercept = sample_mean, color="sample mean"), lwd = 1)
g2 <- g2 + geom_vline(aes(xintercept = theoretical_mean, color="theoretical mean"))
g2 <- g2 + labs(title = "Sample Mean vs Theoretical Mean", x = "Mean", y = "Dencity")
g2
```




```
c(sample_mean, theoretical_mean)
```

```
## [1] 4.961172 5.000000
```

Above shows us that our sample mean is 4.9611719 which is pretty close to our theoretical mean of 5.

### Sample Variance versus Theoretical Variance

```
# the variance of simulations
sample_var <- var(mns)
# theoritical variance
theoretical_var <- (1 / lambda) ^2 / n

c(sample_var, theoretical_var)
```

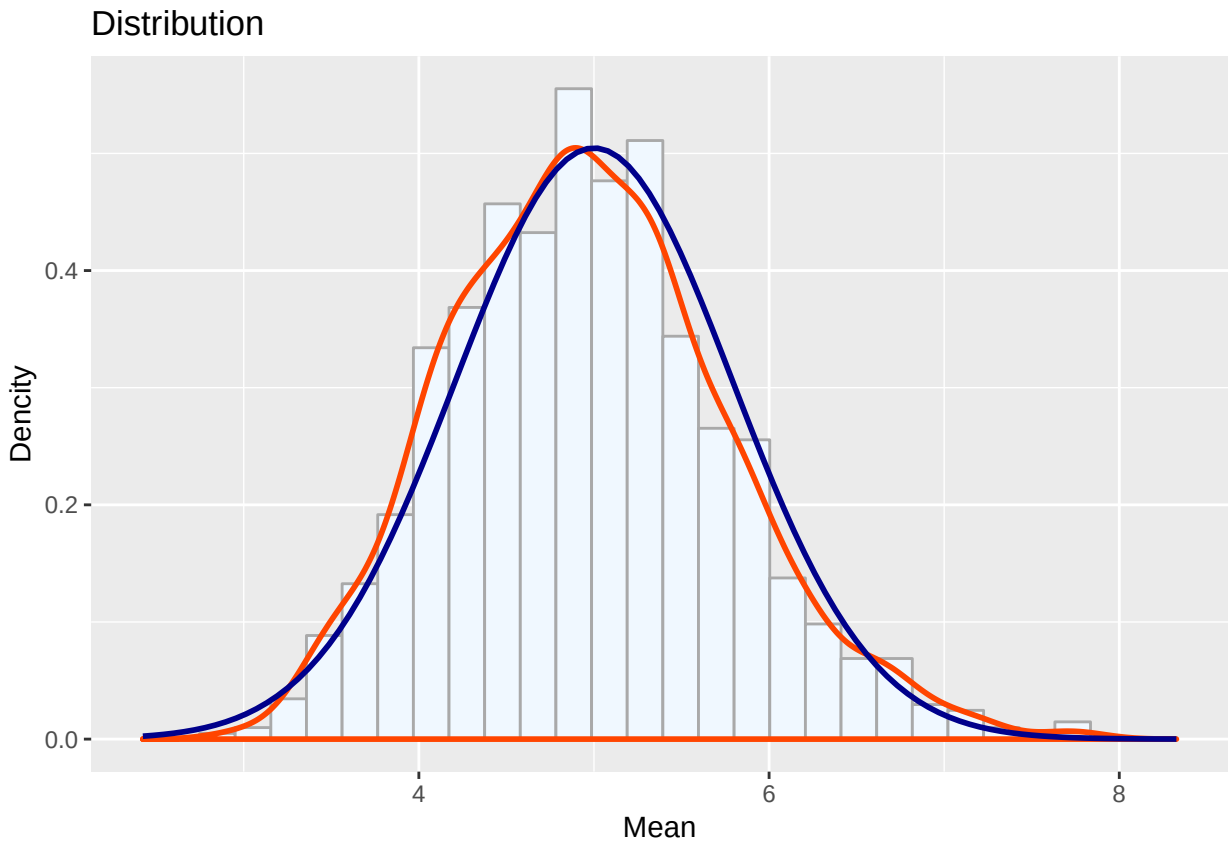
```
## [1] 0.6217743 0.6250000
```

This shows us that the variances are very close.

### Distribution

```
# theoritical standard deviation
theoretical_sd <- 1 / (lambda * sqrt(n))

g3 <- g + geom_density(aes(y = ..density..), color = "orangered", lwd = 1.0)
g3 <- g3 + stat_function(fun = dnorm, args = list(mean = theoretical_mean, sd = theoretical_sd), color =
```

```
g3 <- g3 + labs(title = "Distribution", x = "Mean", y = "Dencity")
g3
```




Above shows us that the distribution (red line) is approximately normal (blue line).