

# Statistical Inference - Course Project 1

*K H CHEN*

*20 Feb 2015*

## About This Project

This is to 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.

## PART 1: Simulation

```
# 1. Setting Environment #
library(ggplot2)
set.seed(10)
lambda <- 0.2

# 2. Generating Simulations #
samplesize <- 40
simulation <- 1000

# 3. Storing Simulations #
simulated_exponential <- matrix(rexp(simulation*samplesize, rate=lambda), simulation, samplesize)

# 4. Getting the average of 40 exponentials from the simulations
row_means <- rowMeans(simulated_exponential)
```

## PART 2: Inferential Data Analysis

### A. Sample Mean versus Theoretical Mean

```
# 1. Calculating the Mean of Distribution #
mean(row_means)

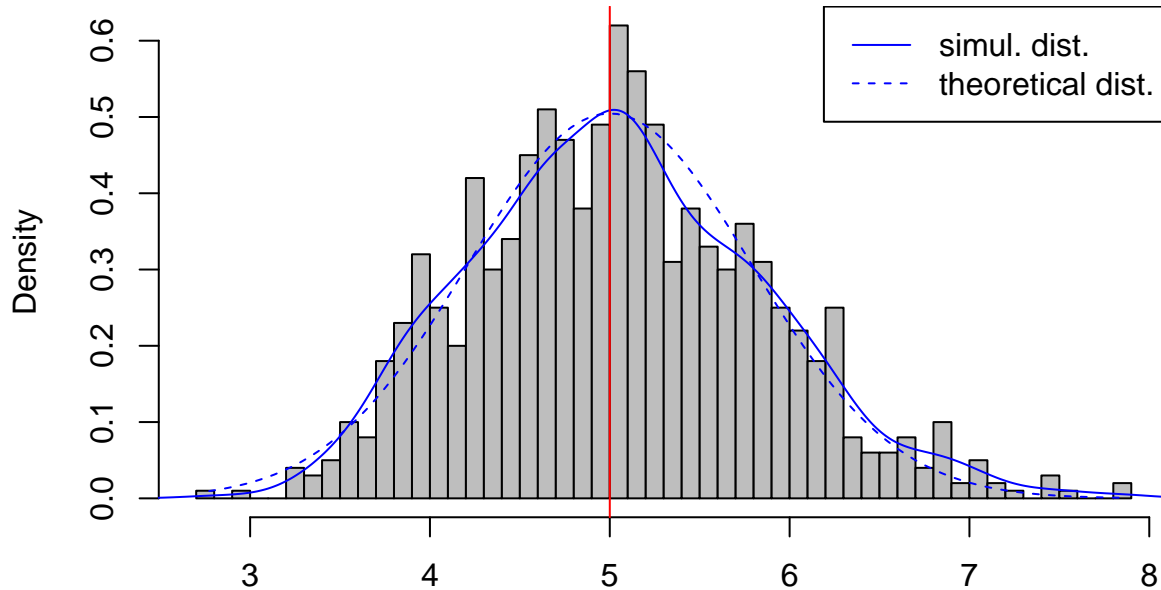
## [1] 5.04506

# 2. Cuoutputing the Mean of Analytical Expression #
1/lambda

## [1] 5
```

The distribution of sample means is shown below:

### Distribution of the mean of samples, generated from exponential distribution with lambda=0.2



The distribution of mean of 40 exponentials is centered at 5.0450596 and it is shown to be close to the theoretical center of the distribution,  $(1/\lambda^{-1} = 1/\lambda)$ .

### B. Sample Variance versus Theoretical Variance

```
# 1. Calculating the Standard Deviation of Mean #
sd(row_means)
```

```
## [1] 0.80881
```

```
# 2. Calculating the Standard Deviation from Theoretical Distribution Expression #
(1/lambda)/sqrt(samplesize)
```

```
## [1] 0.7905694
```

```
# 3. Calculating the Variance of the Sample Mean #
var(row_means)
```

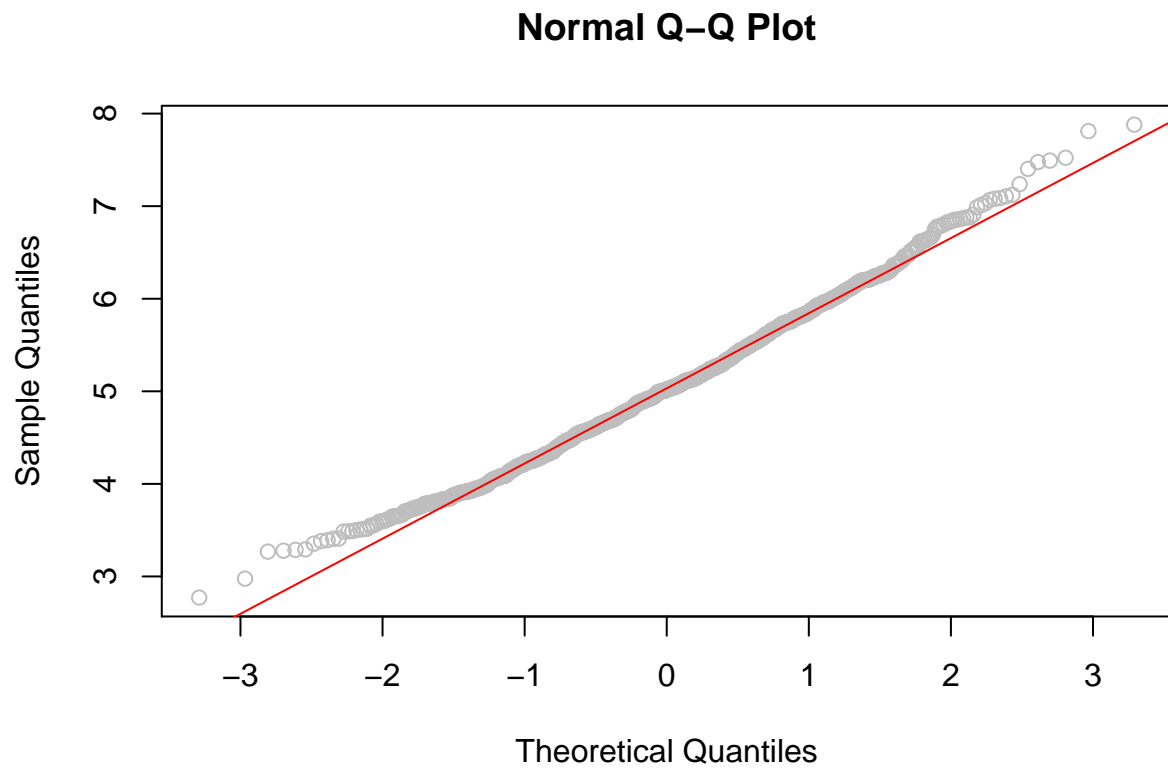
```
## [1] 0.6541736
```

```
# 4. Calculating the Theoretical Variance of the Distribution #
1/((0.2*0.2) * 40)
```

```
## [1] 0.625
```

This shows that the simulated variance of the 40 exponential distribution is close to the theoretical variance of the distribution.

### C. Distribution



The average of the generated samples follows the normal distribution based on the Central Limit Theorem (CLT). It also shows that the distribution of averages of 40 exponentials is approximated to close to a normal distribution.