# Part 1: Simulation Exercise

## Yunpeng Cheng

### 9/15/2020

## Overview

This is part 1 for Simulation Exercise. In this project I 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.

## Simulations:

Initializing parameters and set random seed firstly, then generate the distribution of the averages of 40 exponentials for 1000 simulations.

```r
lambda <- 0.2
nsample <- 40
nsimu <- 1000
set.seed(42)

samp_ave <- NULL
for (n in 1:nsimu){
  samp_ave <- c(samp_ave, mean(rexp(nsample, lambda)))
}
```

## 1 Sample mean and theoretical mean

Sample mean:

```r
mean(samp_ave)
```

```
## [1] 4.986508
```

Theoretical mean:

```r
1/lambda
```

```
## [1] 5
```

The sample mean and theoretical mean are almost same except a tiny difference.

## 2 Sample variance and theoretical variance

Sample variance:

```r
var(samp_ave)
```

```
## [1] 0.6344405
```

Theoretical variance:

```r
(1/(lambda*sqrt(nsample)))^2
```
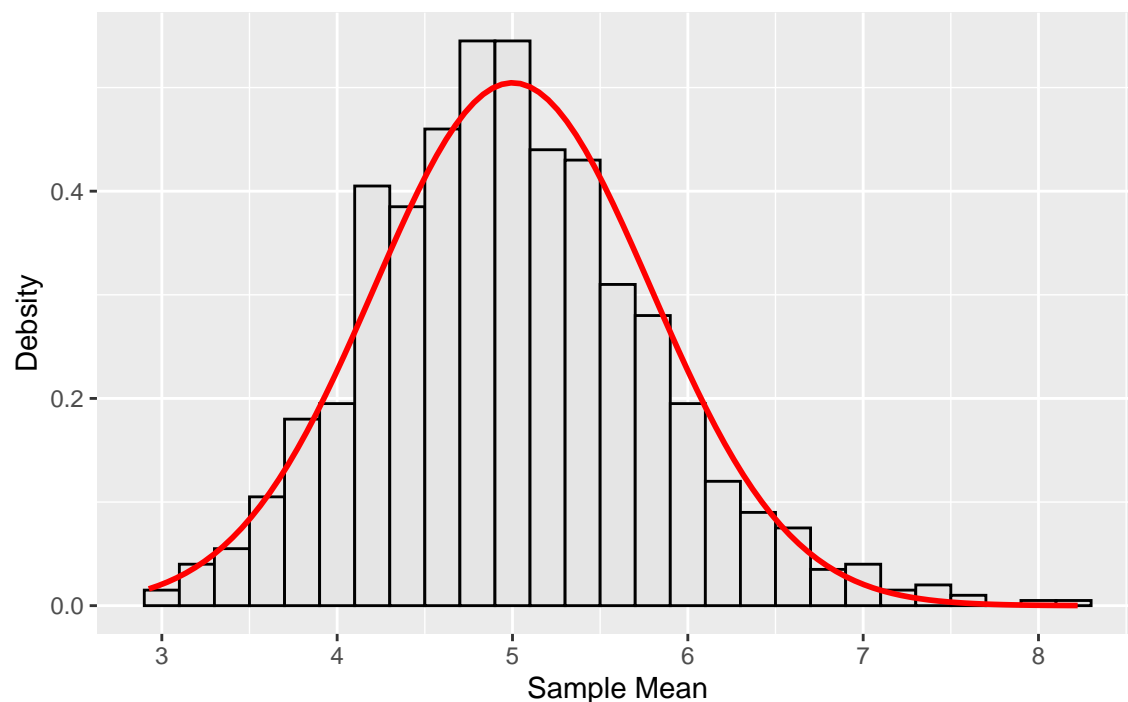
```
## [1] 0.625
```

The sample variance and theoretical variance are close.

## 3 Distribution

To see if the sample distribution is normal distribution, the histogram of the sample should be plotted and compared with the standard normal distribution.

```
## Warning: package 'ggplot2' was built under R version 3.6.1
```



Sample Means Distribution vs Normal Distribution

The gray histogram is for the sample means, the red line is for the normal distribution. From the figure, the distribution of sample means is approximately normal.

Quantile-quantile(QQ) plot is a better graphical way to see if the distribution of sample means is normal or not. The function qqnorm() can plot the sample means against the normal distributin. The function qqline() can add a referenc line to show the deviations.
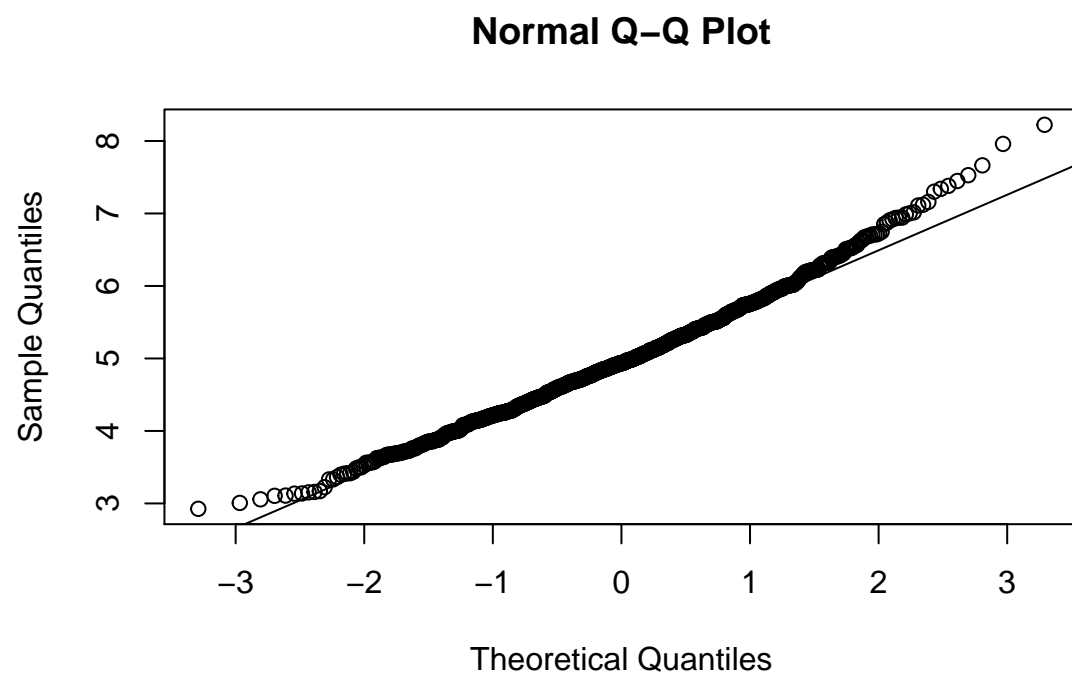
# Normal Q–Q Plot



The qqplot shows that the sample means of 40 exponentials is close to normal distribution.