

# Statistical\_Inference\_Project\_Part\_1

*freestander*

*Saturday, July 25, 2015*

## Problem Description

We 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. We will investigate the distribution of averages of 40 exponentials with a 1000 simulations.

The project is divided into 3 sections:

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.

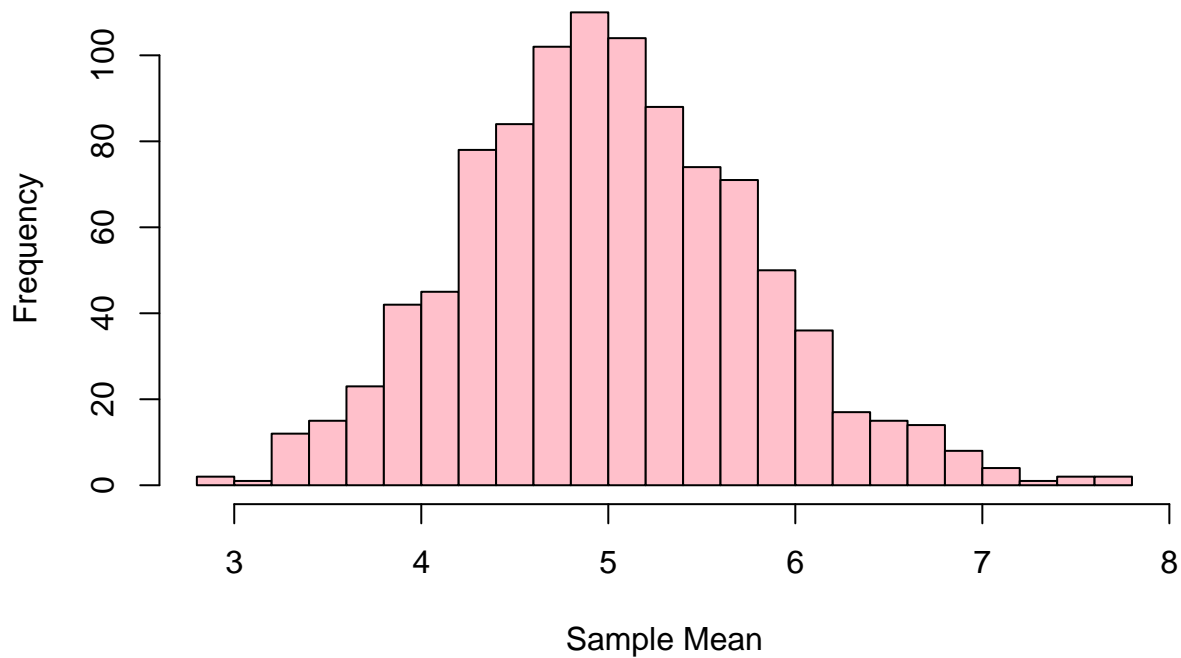
## Run Simulations

```
# set up parameters for the simulation
sim_number <- 1000
lambda <- 0.2
n <- 40
# set up the seed to get the same simulation result
set.seed(123)
# run the simulation for 1000 times for the average of 40 exponentials
sim_mean = NULL
for (i in 1 : sim_number) sim_mean = c(sim_mean, mean(rexp(n, rate=lambda)))
```

Plot the result of the simulation.

```
hist(sim_mean, breaks = 20, col = "pink", xlab = "Sample Mean", main = "Distribution of Sample Mean")
```

## Distribution of Sample Mean



## Sample Mean and Theoretical Mean Comparison

We can calculate the sample mean and theoretical mean of the distribution.

```
# sample mean
sample_mean <- mean(sim_mean)
sample_mean
```

```
## [1] 5.011911
```

```
# theoretical mean
theory_mean <- 1/lambda
theory_mean
```

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

The sample mean based on the simulation is 5.011911 while the theoretical mean is 5, which are very close.

## Sample Variance and Theoretical Variance Comparison

We can calculate the sample variance and theoretical variance of the distribution.

```
# sample variance
sample_var <- var(sim_mean)
sample_var
```

```
## [1] 0.6004928
```

```
# theoretical variance
theory_var <- (1/lambda)^2/n
theory_var
```

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

The sample variance based on the simulation is 0.6004928 while the theoretical variance is 0.625, which are close as well.

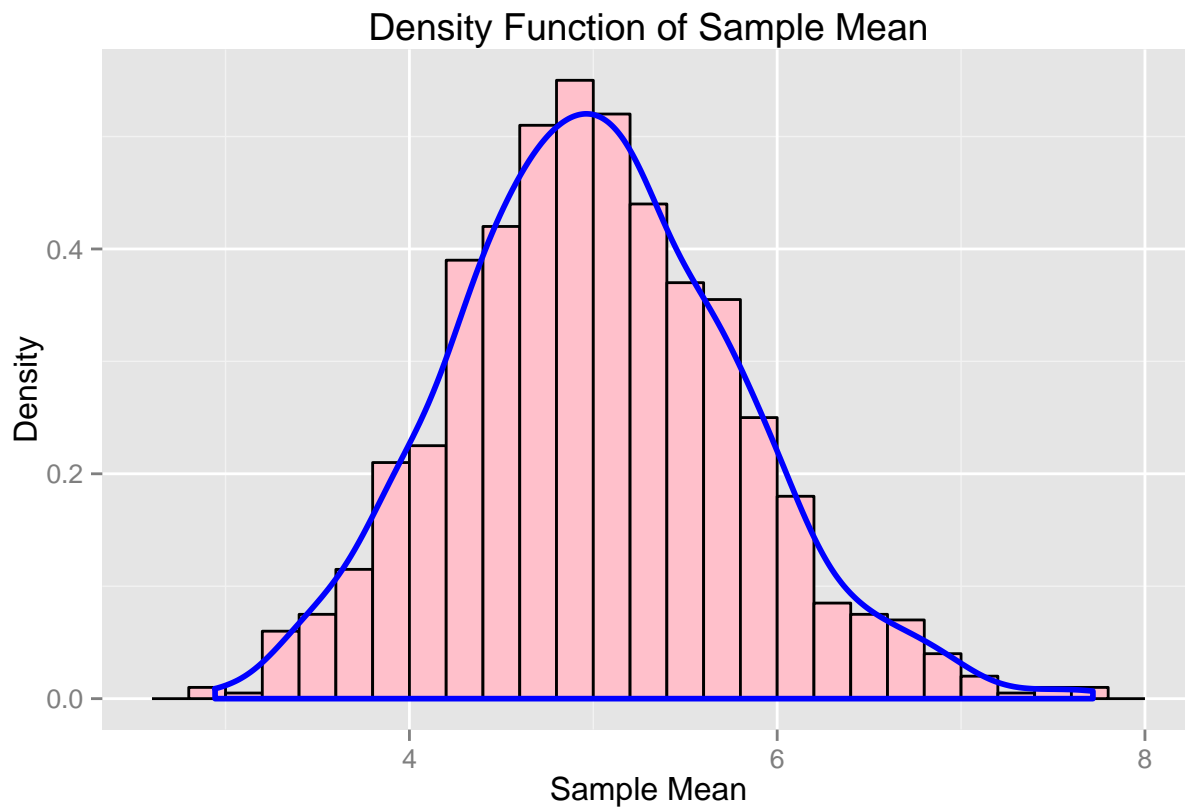
As a result, we can see that the sample distribution is close to the normal distribution based on the values mean and variance.

## Testing Normal Distribution Approximation

To show that the distribution is indeed approximately normal, we can try various tests, as shown below.

### Fitness of the Normal Distribution

```
# load the plotting library
library(ggplot2)
# draw the density with ggplot
sample_data <- data.frame(sim_mean);
m <- ggplot(sample_data, aes(x = sim_mean))
m <- m + geom_histogram(aes(y = ..density..), colour = "black", fill = "pink", binwidth = 0.2)
m <- m + geom_density(colour = "blue", size = 1)
m <- m + xlab("Sample Mean") + ylab("Density") + ggtitle("Density Function of Sample Mean")
m
```



From the plot we believe the sample distribution is very close to normal distribution.

### 95% Confidence Interval Comparison

```
# sample confidence interval of 95%
sample_CI <- round (mean(sim_mean) + c(-1,1)*1.96*sd(sim_mean)/sqrt(n),3)
sample_CI
```

```
## [1] 4.772 5.252
```

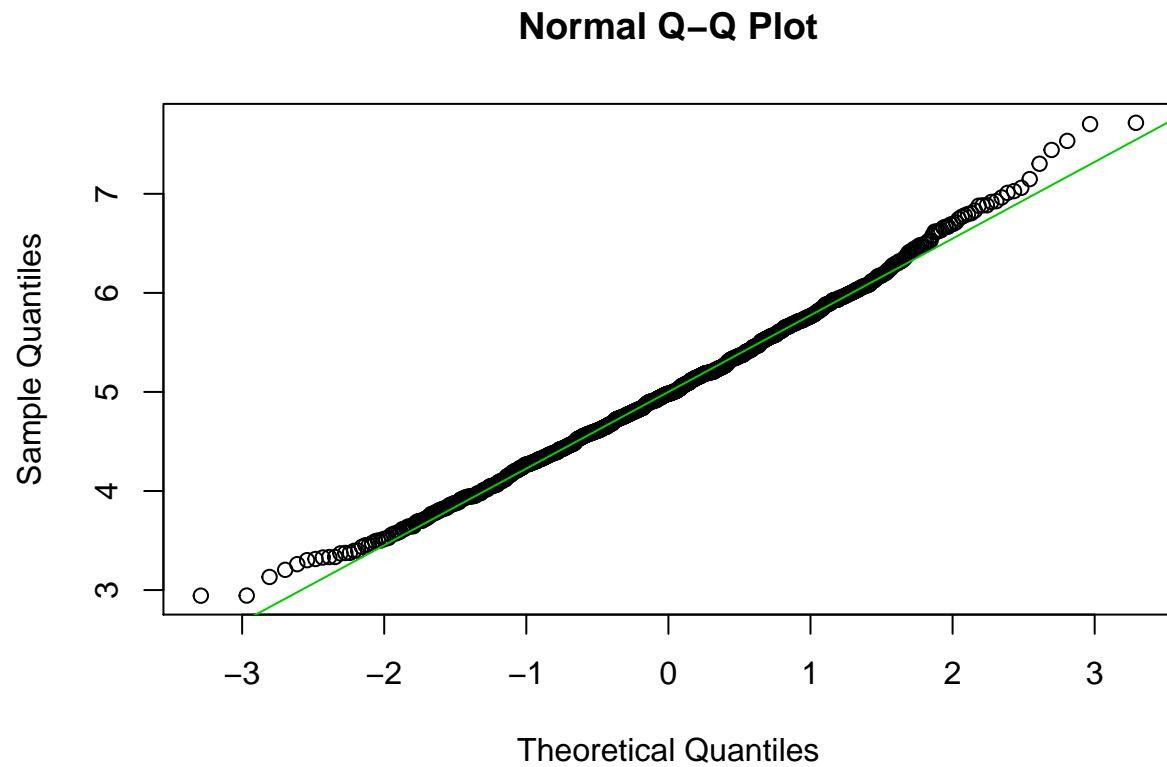
```
# theoretical confidence interval of 95%
theory_CI <- theory_mean + c(-1,1)*1.96*sqrt(theory_var)/sqrt(n)
theory_CI
```

```
## [1] 4.755 5.245
```

95% confidence interval for sample is [4.772, 5.252]. Theoretical 95% confidence interval [4.755, 5.245], which are also close.

### Q-Q Plot of Quantiles

```
qqnorm(sim_mean, main = "Normal Q-Q Plot")  
qqline(sim_mean, col = "3")
```



Sample quantiles also match well with theoretical quantile. As a result, we believe that the sample distribution is approximately normal distributed.