# The Investigation of the Exponential Distribution

*Tianxing Li*

*Friday, February 20, 2015*

## Overview

This is the project for the statistical inference class. This report investigated the exponential distribution in R and compared it with the Central Limit Theorem.

## Simulations

1. The simulation runs on the following parameters.

```
loop <- 1000
lambda <- 0.2
n <- 40
seed <- 100
```

2. Run the simulation and format the result to a matrix.

```
set.seed(seed)
rawResult <- rexp(loop * n, lambda)
result <- matrix(rawResult, loop)
```

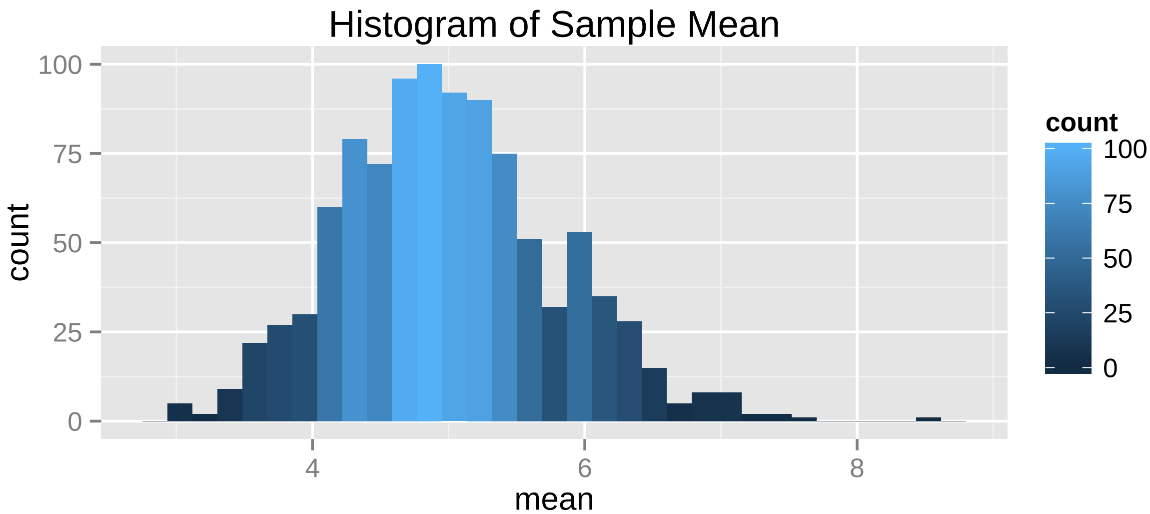
3. Calculate the sample mean and variance.

```
sampleMean <- rowMeans(result)
sampleVar <- apply(result, 1, var)
```

## Sample Mean versus Theoretical Mean

Plot the histogram of the sample mean as below.

```
#hist(sampleMean)
library(ggplot2)
m <- qplot(sampleMean, geom="histogram", main="Histogram of Sample Mean",
           xlab="mean", ylab="count")
m + geom_histogram(aes(fill = ..count..))
```

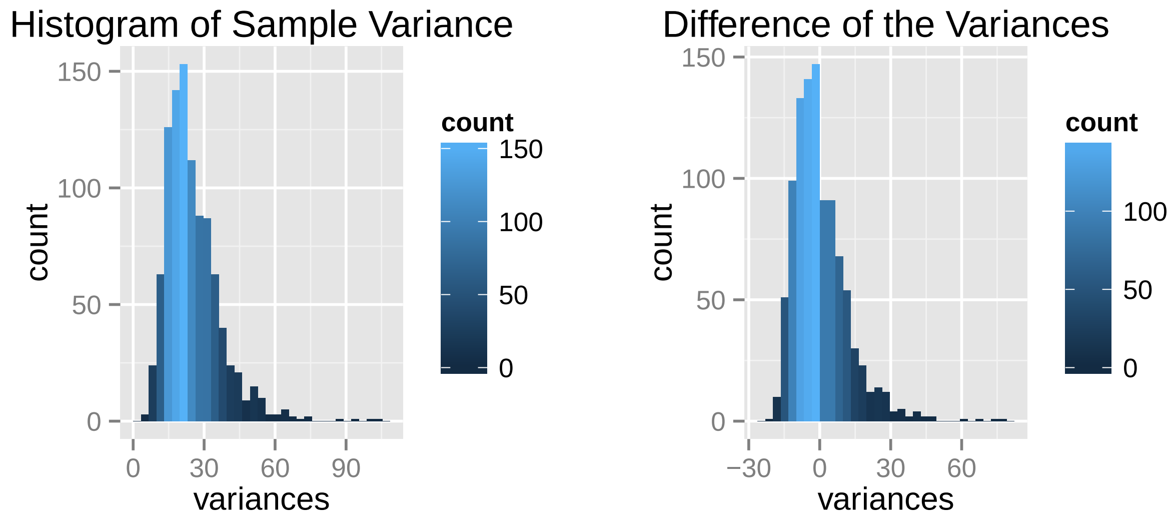



In the simulation, we set $\lambda$ to 0.2. The mean of exponential distribution is $1/\lambda$, which is 5. The simulated data sample has values for mean of 4.9997019, which is close to the expected value.

### Sample Variance versus Theoretical Variance

In the simulation, we set $\lambda$ to 0.2. The mean of exponential distribution is $(1/\lambda)\hat{}2$, which is 25. Plot the histogram of the sample variance as below, as well as the difference between the sample Variance and theoretical Variance, which seems to obey gamma distribution.

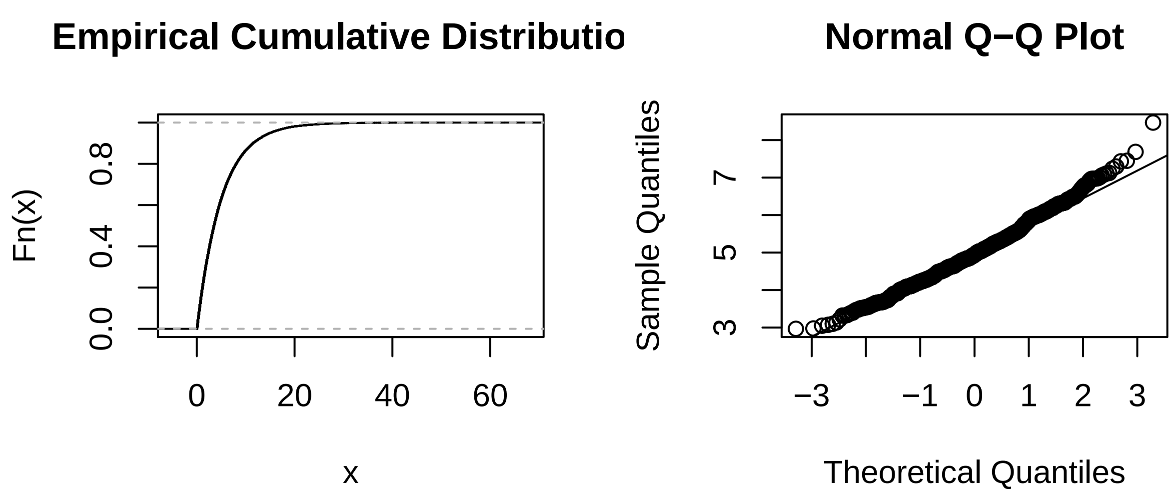
```
library(gridExtra)
m1 <- qplot(sampleVar, geom="histogram", main="Histogram of Sample Variance",
           xlab="variances", ylab="count")
m1 <- m1 + geom_histogram(aes(fill = ..count..))

diff <- sampleVar - (1 / lambda)^2
m2 <- qplot(diff, geom="histogram", main="Difference of the Variances",
           xlab="variances", ylab="count")
m2 <- m2 + geom_histogram(aes(fill = ..count..))
grid.arrange(m1, m2, ncol=2)
```




### Distribution

We can tell the distribution is approximately normal by examining the empirical cumulative distribution and Quantile-Quantile (Q-Q) plot. R allows to compute the empirical cumulative distribution function by `ecdf()` and R also provides `qqnorm()` to get Quantile-Quantile (Q-Q) plot in order totest the goodness of fit of a gaussian distribution.

```
defaultPar <- par(mfrow=c(1, 2))
plot(ecdf(rawResult), main="Empirical Cumulative Distribution")
qqnorm(sampleMean)
qqline(sampleMean)
```




```
par(defaultPar)
```