Statistical Inference Course Project 1

Xiaoyi Leo Liu

Overview

In this project, 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 thousand simulations.

To illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials, we will:

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.

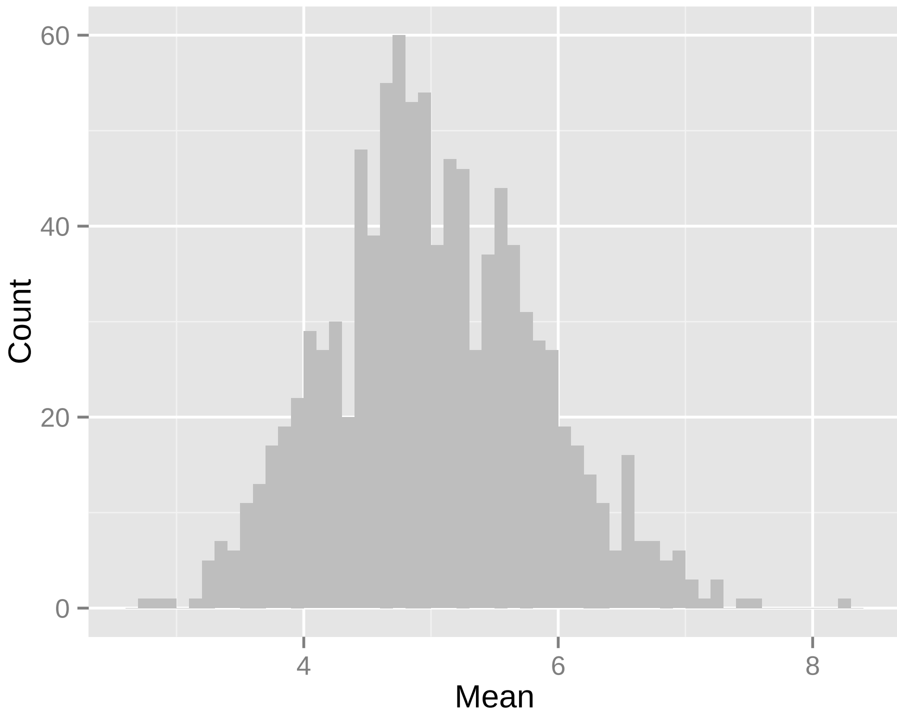
Additionally, we will focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of **40** exponentials.

Simulations

```r
# Parameters
lambda = 0.2;
Num_Exponentials = 40;
Num_Simulations = 1000;

# Generate Exponential Distributions
Expo_Dist = matrix(rexp(Num_Simulations * Num_Exponentials, lambda), nrow = Num_Simulations);
Row_Mean_Expo_Dist = apply(Expo_Dist, 1, mean);

# Hist of Mean
library(ggplot2)
Row_Mean_Expo_Dist = data.frame(Row_Mean_Expo_Dist);
Hist_Mean = ggplot(aes(x = Row_Mean_Expo_Dist), data = Row_Mean_Expo_Dist) + geom_histogram(binwidth = 0
Hist_Mean
```



Sample Mean versus Theoretical Mean

The expected mean of an exponential distribution with rate $\lambda = 0.2$ is $\mu = \frac{1}{\lambda} = 5$.

```
mu = 1/lambda
mu
```

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

The average sample mean of the 1000 simulations of 40 randomly sampled exponential distributions is

```
Sample_Mean = mean(Row_Mean_Expo_Dist[, 1]);
Sample_Mean
```

```
## [1] 5.028531
```

It can be observed that the sample mean is very close to the theoretical mean.

Sample Variance versus Theoretical Variance

The expected variance of an exponential distribution with rate $\lambda = 0.2$ is

```
Theoretical_Var = 1/(lambda^2 * Num_Exponentials);
Theoretical_Var
```

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

The variance of the average sample mean of those 1000 simulations is

```
Sample_Var = var(Row_Mean_Expo_Dist[, 1]);
Sample_Var
```

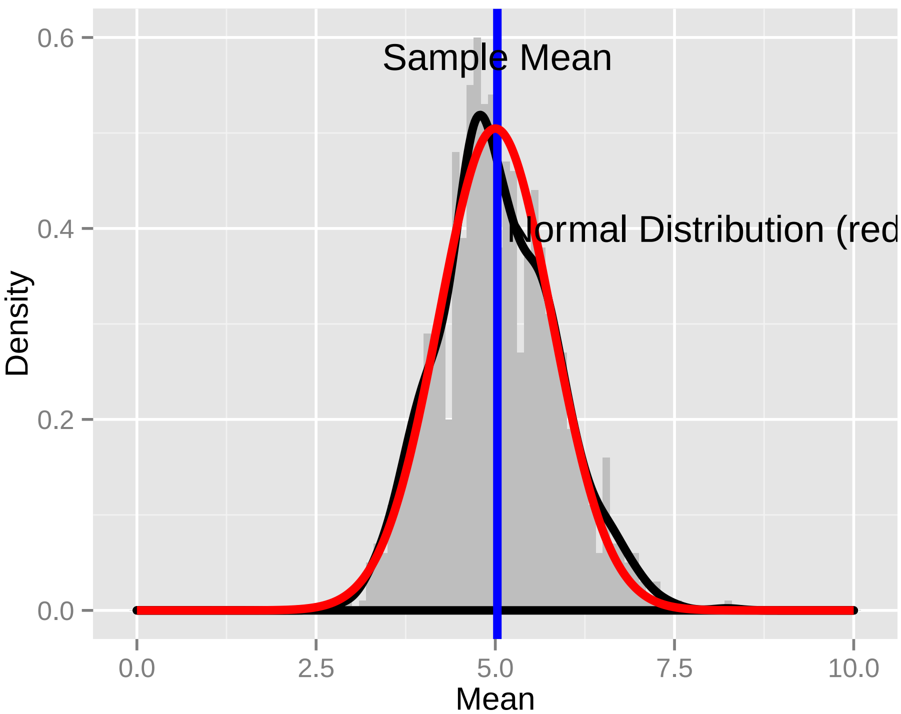
```
## [1] 0.6687454
```

It is also seen that the sample variance of the 1000 simulations and the theoretical variance are very close.

Distribution of Means versus Normal Distribution

```
# Generate Normal Distribution
x = seq(0, 10, len = 1000)
y = dnorm(x, mean = mu, sd = sqrt(Theoretical_Var))
Norm_Dist = data.frame(x, y)

Dist_Mean = ggplot(aes(x = Row_Mean_Expo_Dist), data = Row_Mean_Expo_Dist) + geom_histogram(aes(y = ..de
Dist_Mean
```



As we can see from the plot, the calculated distribution of means of randomly sampled exponantial distributions overlaps nicely with the normal distribution.