Statistical Inference Assignment - Part One

Stephanie Lum

January 31, 2016

Overview

This project investigates exponential distributions simulated in R and compares them to the Central Limit Theorem. The investigation will include the distribution of averages of 40 exponentials and a thousand simulations will be performed.

This report will show the following simulations and comparisons:
1. The sample mean vs. the theoretical mean
2. The sample variance vs. the theoretical variance
3. The distribution of a large collection of random exponentials vs. the distribution of a large collection of averages of exponentials

Simulations

The code below demonstrates the method of finding the mean of 40 exponential distribution values over 1000 simulations.

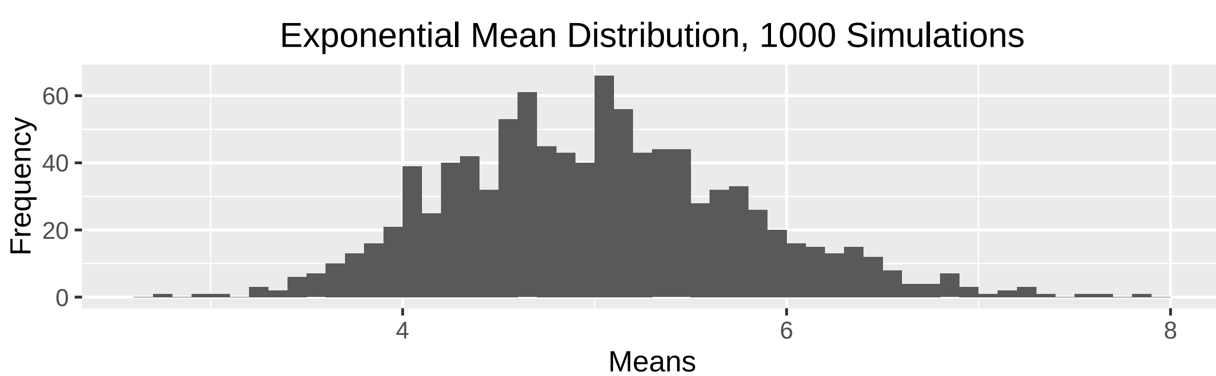
```r
# Load packages
library(ggplot2)

# Define variables
NumberSim <- 1000  # number of simulations
lambda <- 0.2  # rate parameter
n <- 40 # number of exponentials

# Set the seed to ensure reproducibility
set.seed(8675309)

# Perform 1000 simulations of finding the means of 40 exponential distribution values
SimValues <- matrix(rexp(NumberSim * n, rate = lambda), nrow = NumberSim, ncol = n)
SimMean <- rowMeans(SimValues)
```

A histogram of the means can be plotted.



Sample Mean versus Theoretical Mean

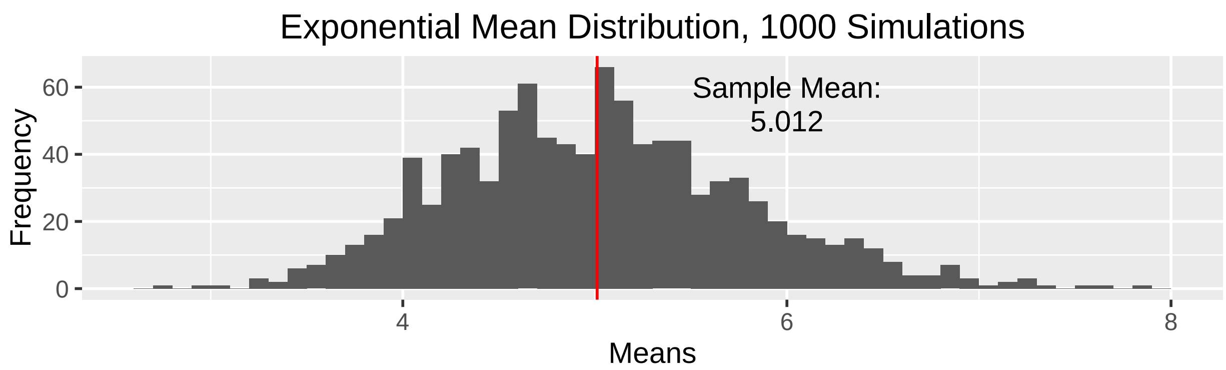
The theoretical mean of the exponential distribution is $\mu = 1/\lambda$ where λ is 0.2.

```
mu <- 1/lambda
mu
```

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

Therefore, μ should equal 5.

The sample mean can be seen in the following plot.



The sample mean is found to be 5.012, which is very close to the theoretical mean of 5.

Sample Variance versus Theoretical Variance

The standard deviation of the sample mean of the exponential distribution is $\sigma = \frac{1/\lambda}{\sqrt{n}}$. Therefore, the theoretical variance of the sample mean is $Var = \sigma^2$.

```
# Determine the theoretical variance
sigma <- (1/lambda)/sqrt(n)
Theoretical_Var <- sigma^2
Theoretical_Var
```

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

The theoretical variance is calculated to be 0.625.

```
# Determine the sample variance
Sample_Var <- round(var(SimMean), 3)
Sample_Var
```

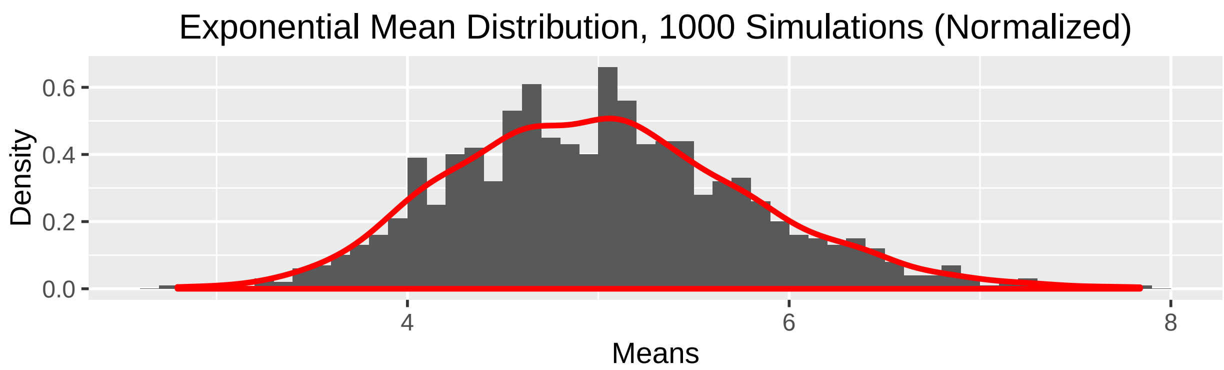
```
## [1] 0.603
```

The sample variance is found to be 0.603, which is also very close to the calculated theoretical variance.

Distribution

The Central Limit Theorem states that with a large enough sample size, the distribution of averages of iid variables should become that of a standard normal distribution.

A plot of the exponential mean distribution with 1000 simulations is shown below.



From the plot, it can be seen that the curve of the densities follows the curve of a standard normal distribution.

An additional way to show whether or not this exponential distribution is normal is to test its 95% confidence interval.

```
# Calculate the theoretical confidence interval
theoretical_conf_int <- mu + c(-1,1)*1.96*sigma/sqrt(n)

# Calculate the sample confidence interval
sample_conf_int <- mean(SimMean) + c(-1,1)*1.96*sd(SimMean)/sqrt(n)
```

The theoretical 95% confidence interval is calculated to be [4.755, 5.245]. The sample 95% confidence interval is found to be [4.77, 5.25], which is very close to the theoretical interval.

This is evidence that the exponential distribution with a λ of 0.2 is approximately normally distributed.