Statistical Inference Course Project Part 1 : Central Limit Theorem using Simulation

Jonathan Lok-Chuen Lo

October 25, 2015

Overview

In this project, I will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution is used and simulated in R with `rexp(n, lambda)` where lambda is the rate parameter. I will investigate the distribution of 40 exponentials and set lambda = 0.2. The simulated mean and standard deviation will be compared against the theoretical values to illustrate the Central Limit Theorem.

Simulations

Generate the simulated data from exponential distribution with lambda = 0.2 and number of exponentials = 40. A thousand simulations is produced, and hence total of 40,000 simulated data is required.

```
# Initialise the seed to ensure reproducibility
set.seed(8)

# Parameters for the simulation
n <- 40
lambda <- 0.2
numSimulations <- 1000

# Generate the simulated data and compare the sample mean and variance
simulatedData <- matrix(rexp(n * numSimulations, lambda), numSimulations, n)
simulatedDataMean <- apply(simulatedData, 1, mean)
simulatedDataVar <- apply(simulatedData, 1, var)
```

Sample Mean versus Theoretical Mean

Calculate the average sample mean and compare it against the theoretical mean of exponential distribution, which is equal to 1/lambda.

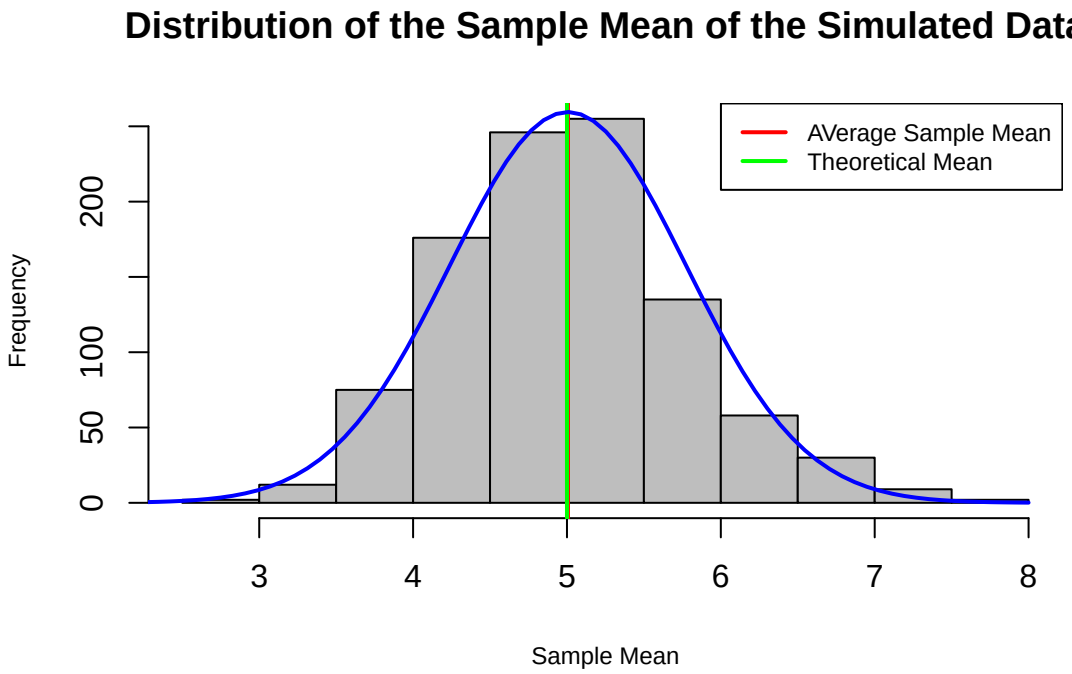
```
sampleMean <- mean(simulatedDataMean)
theoreticalMean <- 1 / lambda

# Print the results
cbind(sampleMean, theoreticalMean)
```

```
##      sampleMean theoreticalMean
## [1,]   5.006442              5
```

Tha sample mean is **5.0064** while the theoretical mean is **5**.

Following is a plot of the distribution of the sample mean distribution. This shows that the sample mean is a very good estimate of the theoretical mean. The corresponding normal distribution is plotted in blue as a reference and the sample mean does appear to be close to normal.



Sample Variance versus Theoretical Variance

Calculate the average sample variance and compare it against the theoretical variance of exponential distribution, which is equal to (1/lambda)^2.

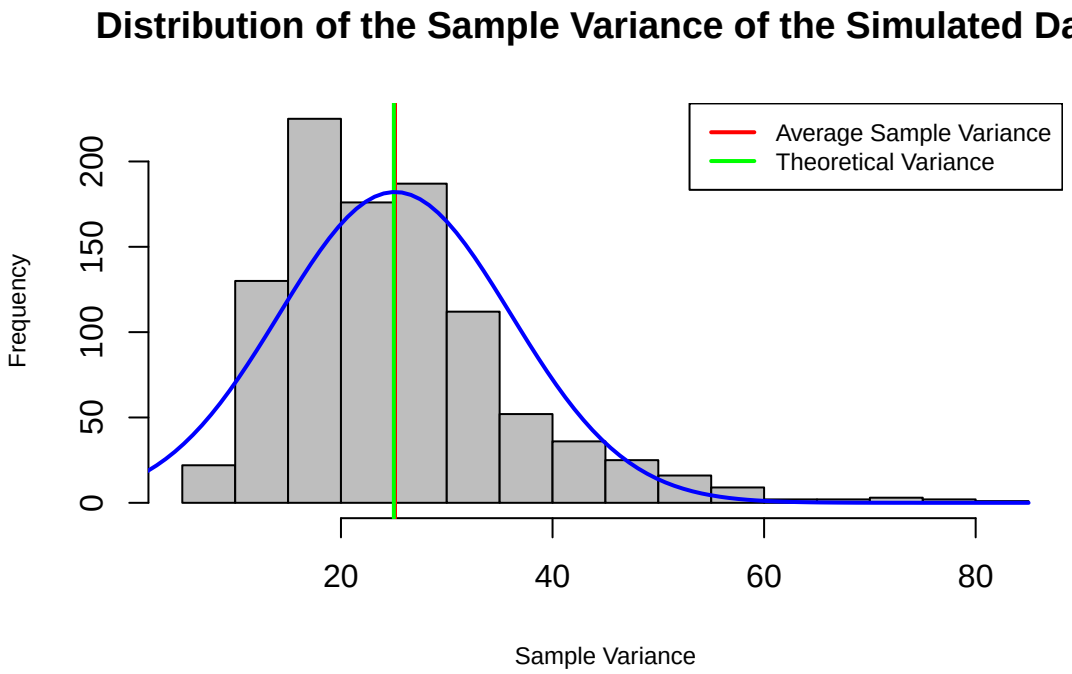
```
sampleVar <- mean(simulatedDataVar)
theoreticalVar <- (1 / lambda)^2

# Print the results
cbind(sampleVar, theoreticalVar)
```

```
##      sampleVar theoreticalVar
## [1,]  25.10708             25
```

Tha sample variance is **25.1071** while the theoretical variance is **25**.

Following is a plot of the distribution of the sample variance. This again shows that the sample variance is a very good estimate of the theoretical variance. The corresponding normal distribution of the sample variance is plotted in blue and the sample variance does appear to be close to normal.



Distribution

Below is a plot of the exponential distribution with number of exponentials = 40 and lambda = 0.2. It can be seen that although the distribution is very different to a normal distribution, the distribution of the sample mean and sample variance are close to normal as shown earlier.