Statistical Inference Project I: Central Limit Theorem

Andreas Hadimulyono

Overview

This document investigates the exponential distribution in R and compare it with the Central Limit Theorem. In this analysis the the sample mean and variance acquired from simulations shall be compared against their theoretical counterparts.

Simulations

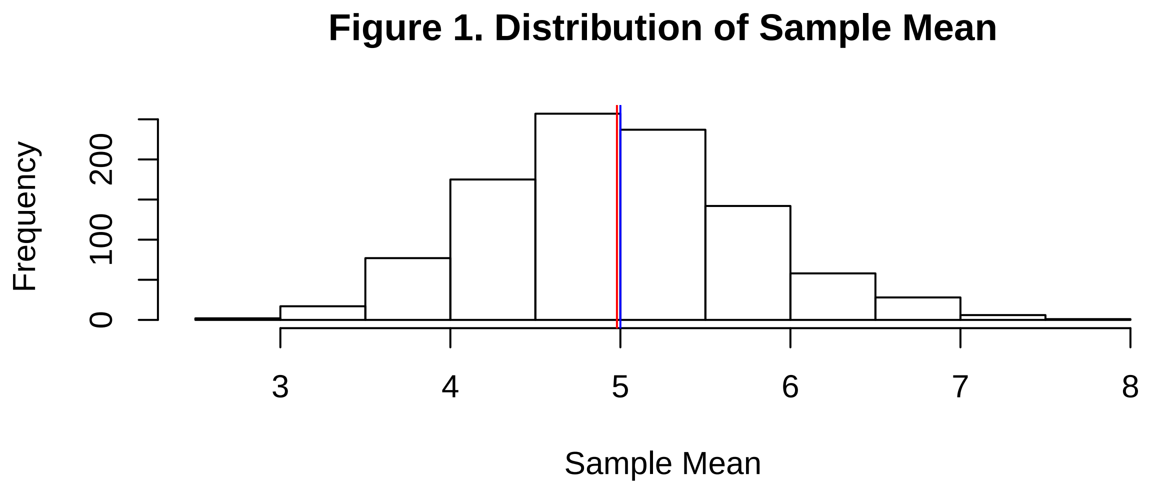
The simulation for exponential distribution is done by using R function rexp. In this analysis the value of lambda used is 0.2 as specified in the Project description. The simulation will first generate 40 random numbers with exponential distribution and take the mean of the 40 numbers. The random number generation process will be repeated 1000 times, and the means generated will be stored in a vector.

```
# set seed so that the analysis can be repeated
set.seed(123856);
# lambda value, number of samples, and number of experiments
lambda <- 0.2; numberOfSamples <- 40; numberOfExperiments <- 1000;
```

```
# generate all means and variants
allMeans <- NULL;
for (i in 1 : numberOfExperiments) {
    allMeans = c(allMeans, mean(rexp(numberOfSamples, lambda)));
}
allVars <- NULL;
for (i in 1 : numberOfExperiments) {
    allVars = c(allVars, var(rexp(numberOfSamples, lambda)))
}
```

SAMPLE VS THEORETICAL MEAN

```
hist(allMeans,
     main="Figure 1. Distribution of Sample Mean",
     xlab="Sample Mean", ylab="Frequency");
abline(v=1/lambda, col="blue")
abline(v=mean(allMeans), col="red")
```

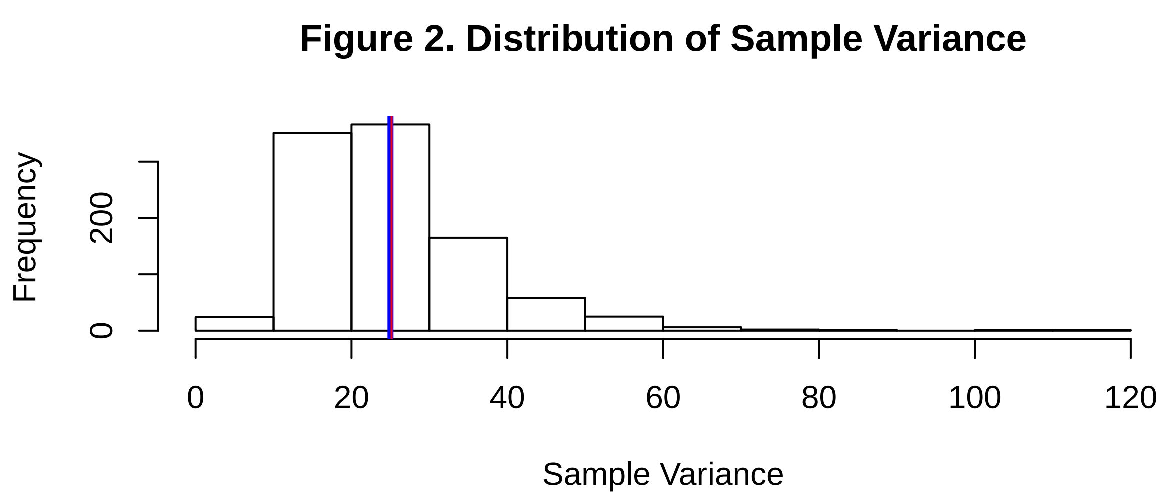


The theoretical mean of the exponential distribution is 1/lambda, which is 5. The Law of Large Number states that as sample sizes grow larger, the average of the sample mean will closely estimate the population mean. The average of the sample mean is 4.9797109 which is very close to the theoretical population mean.

In Figure 1, the average of sample mean is marked with red vertical line. The theoretical population mean is drawn with blue line. As observed in Figure 1, the theoretical mean and average of sample mean are very close to each other.

SAMPLE VS THEORETICAL VARIANCE

```
hist(allVars,
     main="Figure 2. Distribution of Sample Variance",
     xlab="Sample Variance", ylab="Frequency");
abline(v=1/(lambda^2), col="blue", lwd=3) ## make it thicker so it can be seen in the graph
abline(v=mean(allVars), col="red")
```



The theoretical variance of the exponential distribution is 1/lambda^2, which is 25. The Law of Large Number states that as sample sizes grow larger, the average of the sample variance will closely estimate the population variance. The average of the sample variance is 25.1483239 which is close to the theoretical population variance.

In Figure 2, the average of sample variance is marked with red vertical line. The theoretical population variance is drawn with blue line. It may not be clear here as the line almost overlaps, but the two lines are very close to each other.

DISTRIBUTION

According to Central Limit Theorem, the distribution of mean of IID variable (properly normalized) become that of standard normal as the sample size increases. The mean for this population (exponential distribution) is 1/lambda, which is 5 and variance is 1/lambda^2, which is 25. The standard error is then the square root of variance divided by n (number of samples), which is 0.7905694. We will use this standard error and population mean to normalize the sample mean we have calculated in the previous step, and then plot the histogram against theoretical normal distribution. Figure 3 shows the plot, with histogram as the distribution density and the red line as the normal distribution. As Figure 3 shows, the distribution of the sample mean closely follows that of normal distribution

```
meansNormalized <- (allMeans - (1/lambda))/sqrt((1/(lambda^2))/numberOfSamples)
hist(meansNormalized, ylim = c(0,.5), prob=TRUE,
     main="Figure 3. Distribution of Sample Mean Against Normal Distribution",
     ylab="Density",
     xlab="Normalized Sample Mean")
curve(dnorm(x), add=TRUE, col="red", lwd=2)
```