Title

An Exercise about Exponential Distribution and Central Limit Theorem

Overview

With the help of simulation methods embeded in R, this exercise is to demonstrate certain properties of the distribution of the averages of samples with the size of 40 exponentials collected from the same exponential distribution. The exercise tries to fulfill the following three targets:

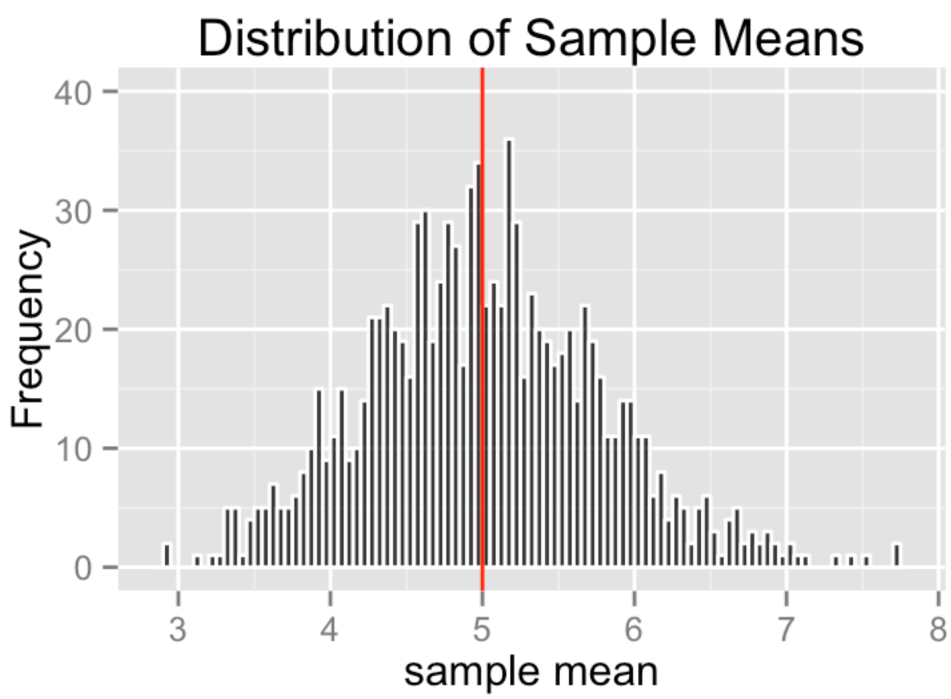
1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

Simulations

Following instructions in this exercise, I simulated an exponential distribution with its lambda being 0.2. I collect a sample of 40 exponentials from the distribution and calculate its mean value, repeating this process for 1,000 times. Using the 1,000 averages of 40 exponentials, I try to illustrate those topics mentioned above.

Sample Mean versus Theoretical Mean

I calculate the mean values of the 1,000 samples, and draw a historgram about them.



The histogram shows that the sample means forms an approximate normal distribution, symmetrically centering around the value of 5. This result is consistent with the fact that the theoretical mean of the exponential distribution is the inverse of the lambda value, which is 1/0.2 = 5.

Sample Variance versus Theoretical Variance

Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

By theory, the variance of an exponentional distribution is the inverse of the lambda squared, namely,

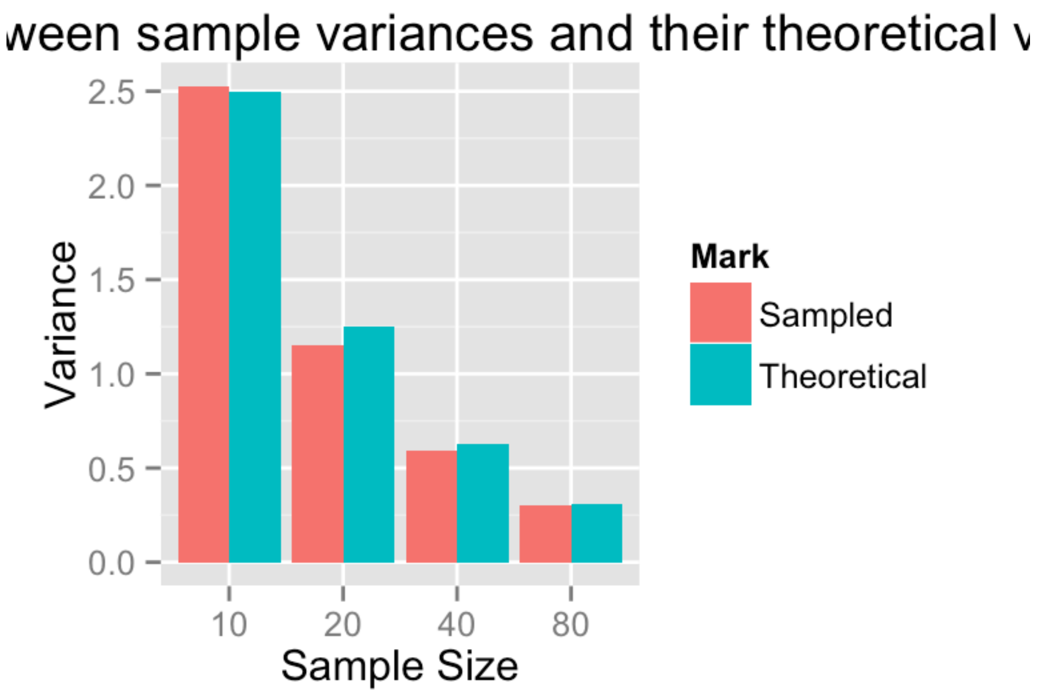
$$Var[X] = \frac{1}{\lambda^2}$$

The theoretical variance (S) of the averages of samples of size n is the variance of the original expoential distribution divided by sample size, namely,

$$S = \frac{Var[X]}{n}$$

To show the relationship between the variance of sample means of the simulated exponential distribution and its theoretical variance, I made four collections which contain 1,000 samples of size 10, 20, 40, and 80 respectively.

I calculated the variances of the four collection of samples, and their repective theoretical sample variances. To comparison the difference between the two group of values, I merge them into one data frame and use barplot to arrange them by group. By pairwise comparing sample variance and its theoretical value, we can clearly see the pattern in the figure.



By comparison, I find no significant difference between the simulated sample variances and their respective theoretical values. So the result is consistent with the theory of sample variances of an exponential distribution.

Distribution

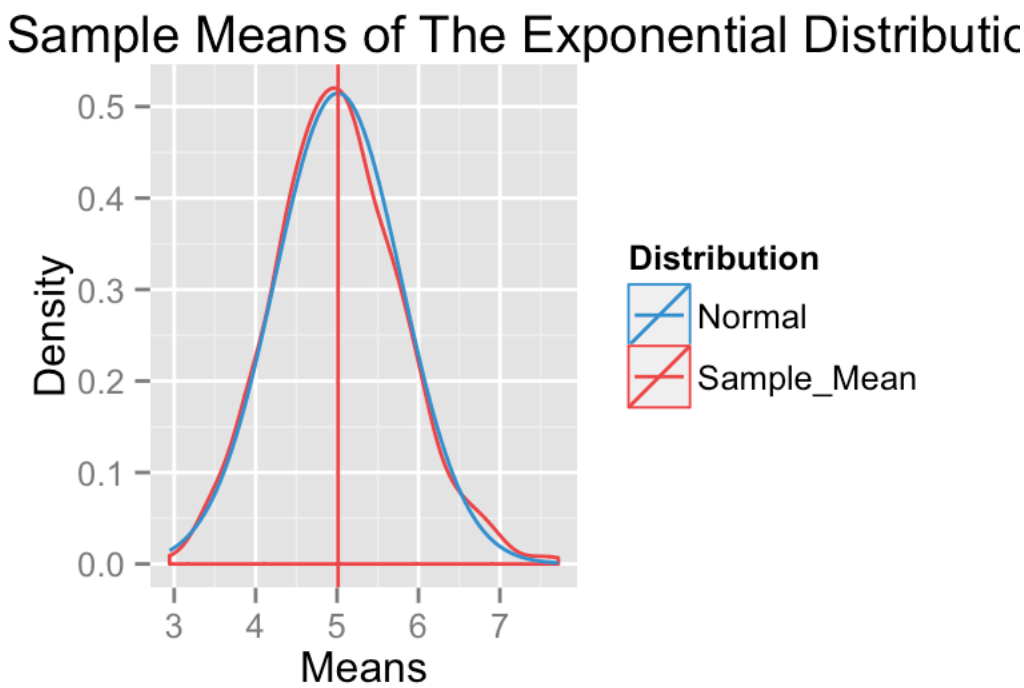
First, I calculate the average and standard deviation of the sample means of the collection of samples of 40 exponentials

```
## [1] 5.011911
```

```
## [1] 0.6004928
```

The average is 5.0119113, and the standard deviation is 0.6004928.

To demonstrate the distribution of means of samples of 40 exponentials from the exponential distribution, I plot a density distribution about those sample means, and superimpose a normal distribution with the same mean value and standard deviation as those of the collection of samples, namely, $N(\mu = 5.0119113, \sigma = 0.6004928)$.



From the plot we can see that the density distribution of sample means of the exponential distribution looks like the shape of bell curve, and approximately overlapps with the curve of normal ditribution which has the same mean value and standard deviation.