Exponential Distribution Analysis

Trina Dey

11 July 2020

Overview

This report provides analysis of Exponential Distribution and compares how the distribution of mean of samples appears close to Normal distribution. We know that the exponential distribution has 2 parameters - n the population size and lambda the rate parameter. We also know that the mean and standard deviation of exponential distribution is 1/lambda. We will simulate some samples and analyse the data with respect to Central Limit Theorem.

Data Simulation and Calculation

For our simulation we are going to set n = 40 and lambda = 0.2 and calculate means of 1000 samples derived from exponential distribution and plot it on a graph.

Here is the data for generating 1000 samples of means of 40 exponential data.

```
set.seed(512)
mean_data <- NULL
for( i in 1:1000)
  mean_data <- c(mean_data , mean(rexp(40,0.2)))
```

Now we will compute mean and variance of the sampling distribution.

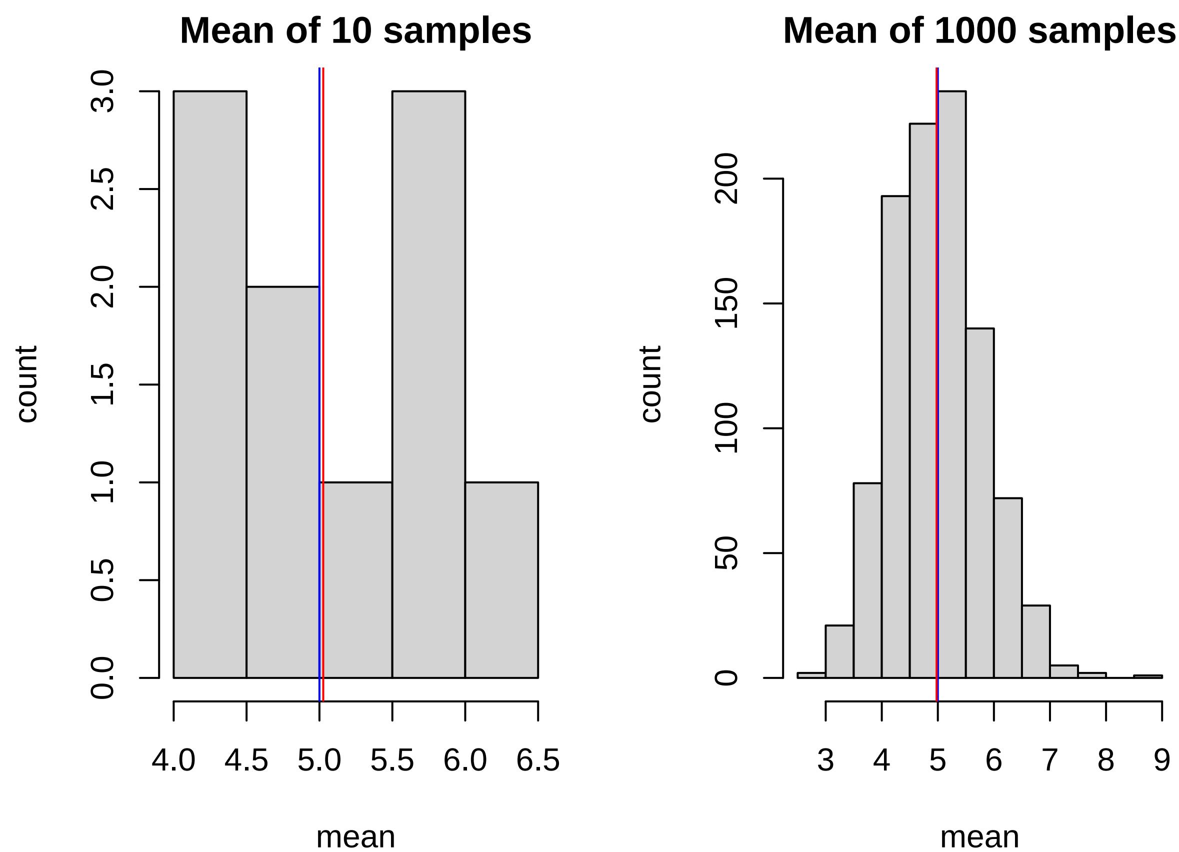
```
sample_mean <- mean(mean_data)
sample_var <- var(mean_data)
c(sample_mean,sample_var)
```

```
## [1] 4.9777902 0.6250162
```

Sample Mean vs. Theoretical Mean

The sample mean estimates the population mean. Population mean is 1/lambda which is equal to 5.0. So our sample mean should be same or close to our population mean. And that is what we found by our analysis in previous step. We will show this by plotting the sampling distribution. To show the difference of sample size and to verify the Central Limit Theorem, we will plot data for two different sample size say 10 and 1000.

```
par(mfrow = c(1,2) , mar = c(4,4,2,2))
hist(mean_data[1:10] ,xlab= "mean", ylab="count", main="Mean of 10 samples", border = "black")
abline(v=5.0, col="blue")
abline(v=mean(mean_data[1:10]), col="red")
hist(mean_data , xlab="mean", ylab="count", main = "Mean of 1000 samples",border = "black")
abline(v=5.0 , col="blue")
abline(v=sample_mean , col ="red")
```



The blue line shows the population mean which is 1/lambda i.e. 5.0 in our case and the red line shows sample mean which is closer to population mean. When the sample size is large enough, sampling distribution appears normally distributed and sample mean approaches population mean.

Sample Variance vs. Theoretical Variance

The variance of sampling distribution can be calculated by square of standard deviation or square of the standard error which is defined as population standard deviation divided by square root of population size.

```
lambda <- 0.2
population_sd <- 1/lambda
populationsize <- 40
theoretical_sample_sd <- population_sd/ sqrt(populationsize)
theoretical_var <- theoretical_sample_sd^2
c(sample_var,theoretical_var)
```

```
## [1] 0.6250162 0.6250000
```

So our calculated sample variance turns out to be 0.6250162 compared to theoretical value 0.625 which seems pretty close.

Conclusion

As the central limit theorem predicts that if we take sufficiently large samples from the population with replacement, the sampling distribution ie central limits of the sample (mean or median) will be approximately normally distributed. The plots clearly show how the distribution seems approximately normal when the sample size taken is sufficiently large (say 1000) vs not so normally distributed when the sample size is small (say 10). Also, with sample size 10, we can see a difference between the blue line of population mean and the red line of sample mean, however as the sample size grows large, the sample mean clearly approaches population mean.