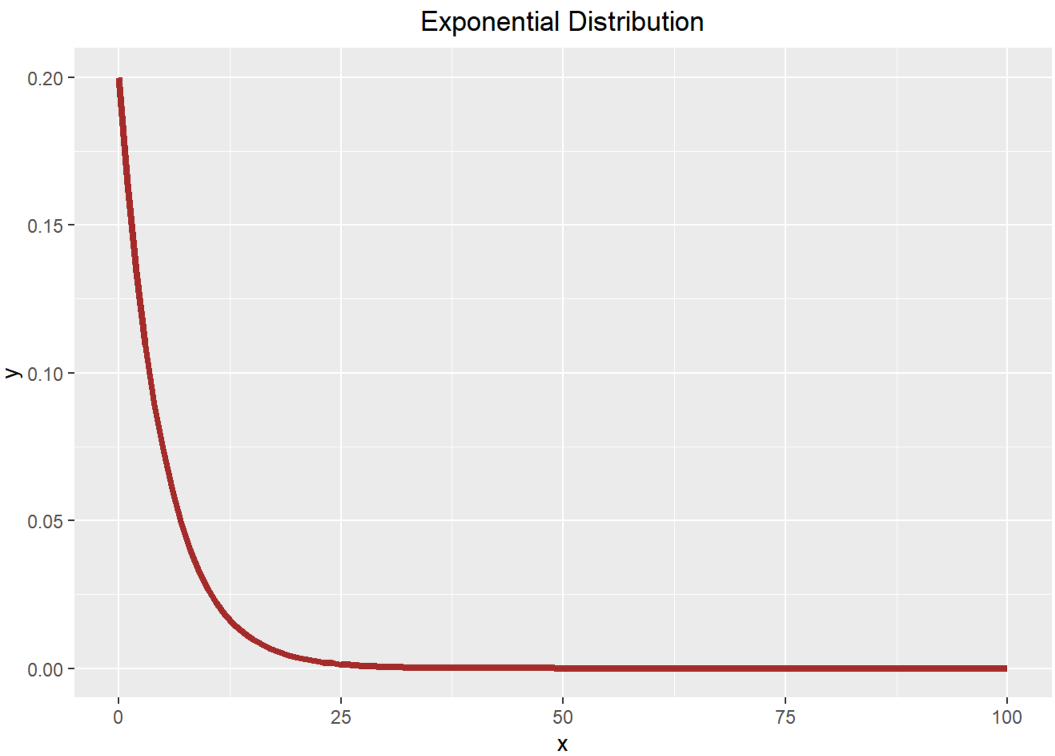
Simulation of CLT with Sample Means from Distribution of Exponential Distribution

by KSW

2023-10-22

Overview

Central Limit Theorem(also as known as CLT) states that the distribution of sample means approximates a normal distribution (i.e., a “bell curve”) as the sample size becomes larger, regardless of the population’s actual distribution shape. In this report, we will explore how the theorem appears in real life by simulating the samples from Exponential Distribution. Exponential Distribution is shown as follows when the lambda to ‘0.2’.



Simulations

So, let’s make samples for our simulation! we will make a thousand of samples of size 40. And calculate 1000 sample means. With this, we will look at the mean, variance of the distribution and distribution itself in comparison to the theoretical(CLT) counterpart of each of them. Let’s start!

```
n=40
B=1000

set.seed(123)
randomexp <- rexp(n*B, rate = 0.2)
samples40 <- matrix(randomexp, nrow=B, ncol = n)
```

In the above code, we set seed and make a thousand of samples of size 40 each. Here are a few example of our samples. Each row is equivalent to one sample.

Now we will calculate means of each rows to make 1000 sample means.

```
samplemeans <- apply(samples40, MARGIN = 1, mean)
samplemeans <- data.frame(x=samplemeans)
```

And here are some examples of our sample means.

```
head(samplemeans)
```

```
##          x
## 1 4.438396
## 2 5.698263
## 3 6.963634
## 4 4.702773
## 5 4.106572
## 6 4.665833
```

Let's calculate mean and variance of our distribution!

```
sammean <- mean(samplemeans$x)
samvar <- var(samplemeans$x)
```

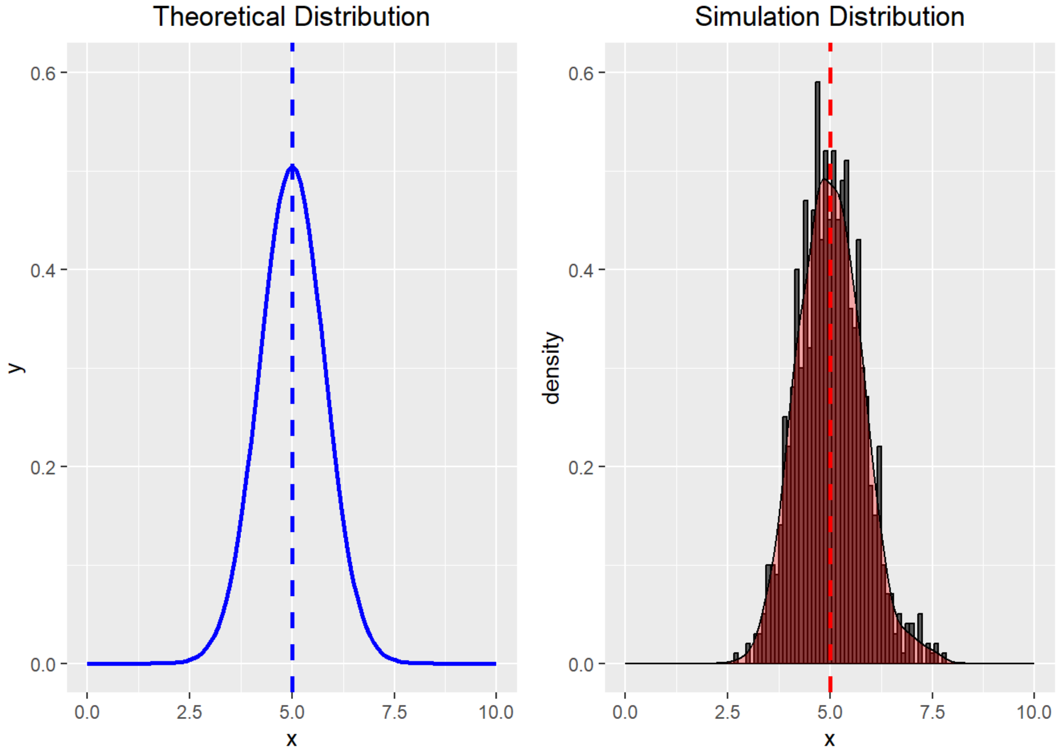
Mean and variance of our sample means distribution is 5.0119113 and 0.6088292 each.

Sample Mean versus Theoretical Mean

So, let's compare our simulated mean to the mean that CLT states. Theoretically, CLT states that as the sample size becomes larger, the mean of sample means approximates a population mean. Since the mean of Exponential Distribution(population distribution) is '1/lambda' and we set lambda to 0.2, the population mean is 1/0.2 = 5. So our mean of sample means has to approximate 5. Previously we confirm that mean value is 5.0119113. It is pretty close!

And Let's check it visually. We will plot the distribution of sample means with the theoretical distribution that it has to follow together. Look at how similar those two distribution are.

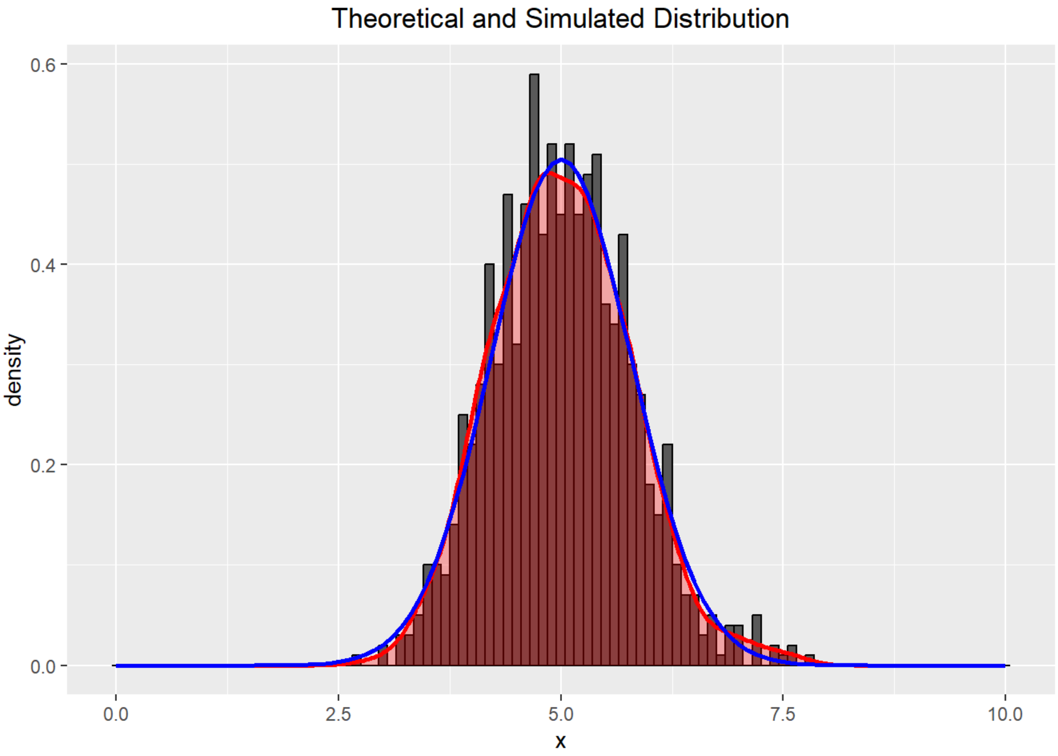
In according to CLT, if the sample size is large enough, the distribution of it approximates the normal distribution with same mean with the population and variance 'population variance/sample size'. The standard deviation of Exponential Distribution is 1/lambda, so when lambda is 0.2, the standard deviation is 5, variance is 25. So we can expect that our sample means distribution approximates normal distribution with mean 5, variance 0.625. Let's check it!



In above graph, the dashed lines represent the mean value of those two distribution each. As expected, simulation distribution is similar with theoretical normal distribution and both are centered at about 5, as the theorem says.

Sample Variance versus Theoretical Variance

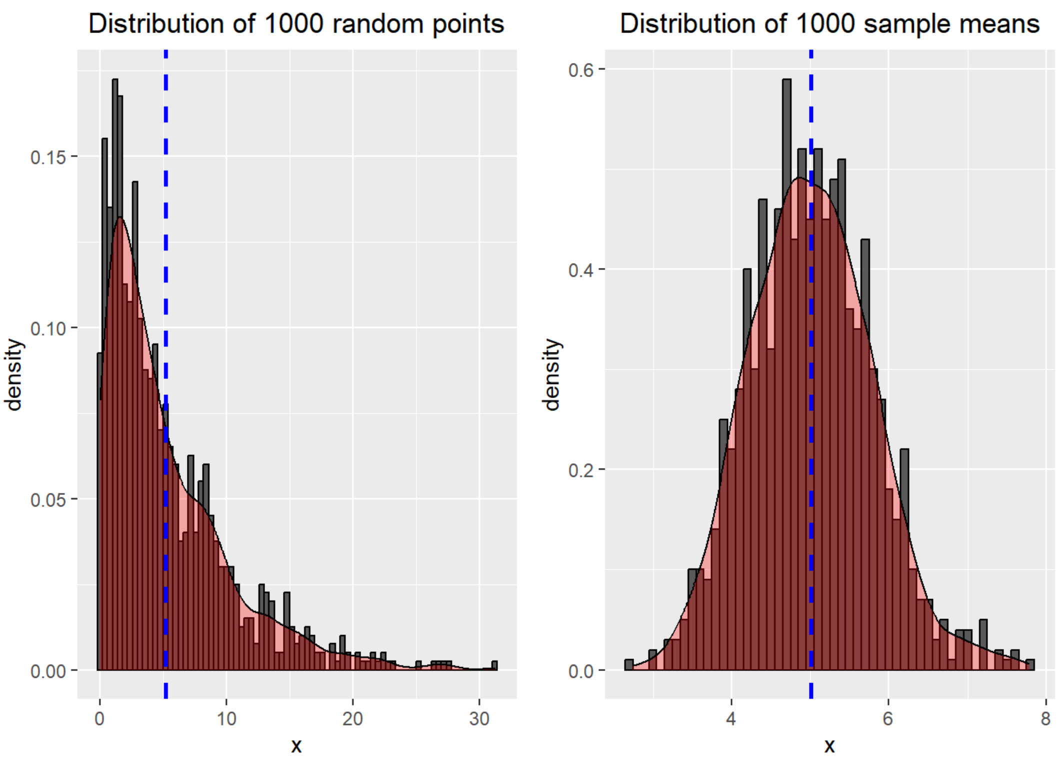
And next, Let's compare theoretical and simulation variance. As said before, the theoretical variance of sample mean distribution is 'population variance/sample size', so 25/40, 0.625 . In our set of sample means, variance is 0.6088292. They are quite similar! If we plot the 2 distribution in a panel together, it is as follows.(The 'blue' line is theoretical distribution and the 'red' line is simulated distribution.)



Despite their similarity, there is a slight difference between them. It can be assumed that it's because our sample size 40 is not large enough to reduce the difference fully.

Distribution

Lastly, Let's check if the distribution of our sample means is approximately normal. For this, we will plot 2 graphs. One is the distribution of 1000 random points from population(Exponential Distribution) and the other is the distribution of our sample means.



Again, CLT says that as the sample size becomes larger, the distribution of sample means approximates a normal distribution (i.e., a "bell curve"). In the above graph, the right graph(sample mean distribution) is symmetrical bell-shape and centered at about 5 although the distribution of 1000 sample points that approximates the population distribution is not normal. So we can say that our sample means distribution is approximately normally distributed.