Comparative study between Exponential Distribution and Central Limit Theorem

Raghav Pradhan

2024-12-07

Synopsis

This is a comparative study between Exponential Distribution (**Theoretical**) & CLT (**sample**) to check if there will be any difference between them.

The *central limit theorem* states that if you take sufficiently large samples from a population, the samples' means will be normally distributed, even if the population isn't normally distributed.

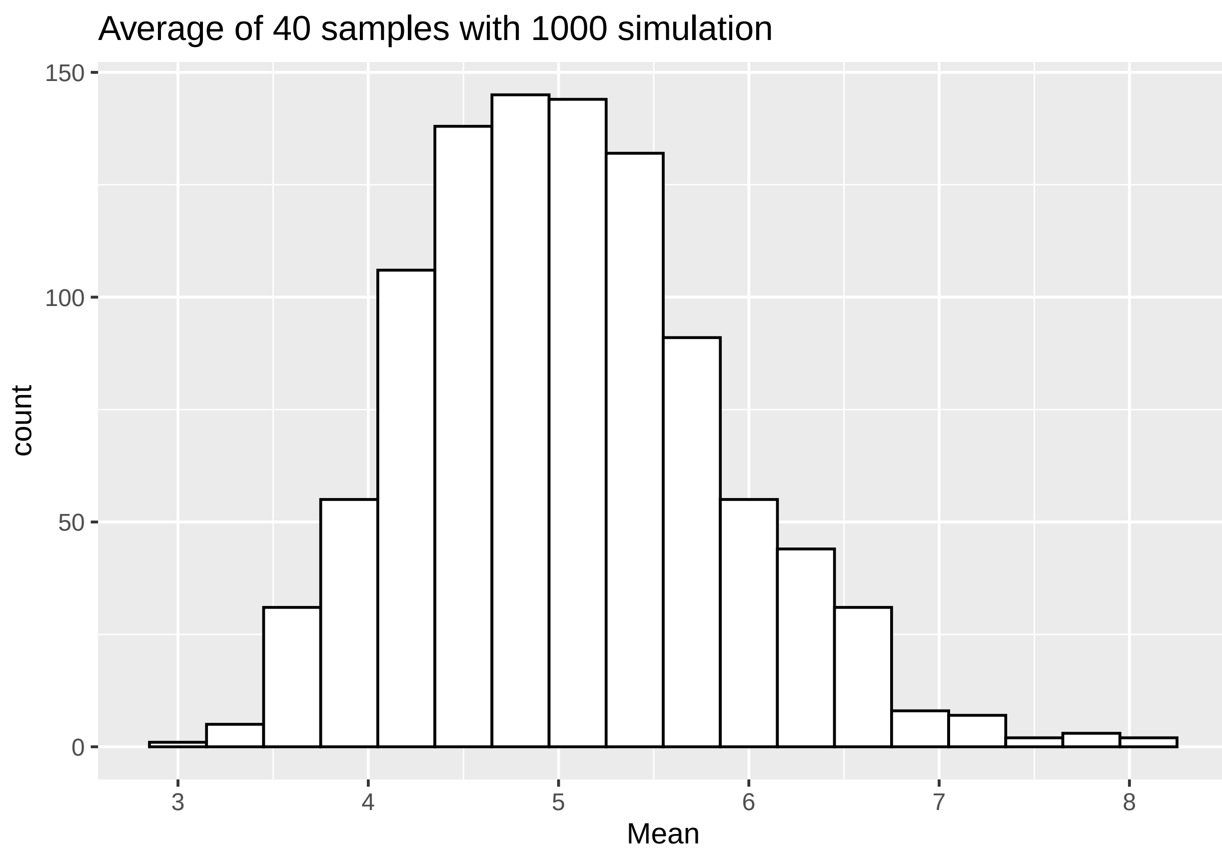
For the sample distribution I will be the using the distribution of averages of 40 exponential samples and simulate 1000 times for the CLT to work correctly,then I will also be comparing the **sample** and **theoretical** means & variances.

Simulation

```r
set.seed(10)  #seeding the values

 mns <- as.numeric()  #empty vector to store values in for loop

for (i in 1:1000)  {    #this loop samples 40 exponential distribution and calculates it's mean
  mns <- c(mns,mean(rexp(40,0.2)))  #then stores it in mns for 1000 times
}
```

Now let's plot a simple histogram to see how CLT works

```r
library(ggplot2)  #Using the GGplot package

ggplot(as.data.frame(mns),aes(x=mns)) +
  geom_histogram(color = "black",fill = "white",binwidth = 0.3) +
  labs(title = "Average of 40 samples with 1000 simulation") + xlab("Mean")
```



We can see that the distribution does indeed follow a normal distribution.

since we are done with simulating the sample distribution now let's compare it with Theoretical mean.

Comparison of mean & variances

Let us first calculate the **sample** mean and their variances.

```
s_mean <- mean(mns) #mean
s_var <- var(mns)  #variance
```

The **Theoretical** mean of an exponential distribution is $1/\ \lambda =$ `5` where λ *lambda* is 0.2, whereas the **Sample** mean is `5.04506`.

The **Theoretical** variance of an exponential distribution is $(1/\lambda\ \hat{}2)/\text{n} =$ `0.625`, and the **Sample** variance is `0.6372544`.

```
theoretical_m <- 5
theoretical_var <- 0.625


s_mean - theoretical_m   #calculating the diff b/w sample & theoretical mean
```

```
## [1] 0.04505959
```

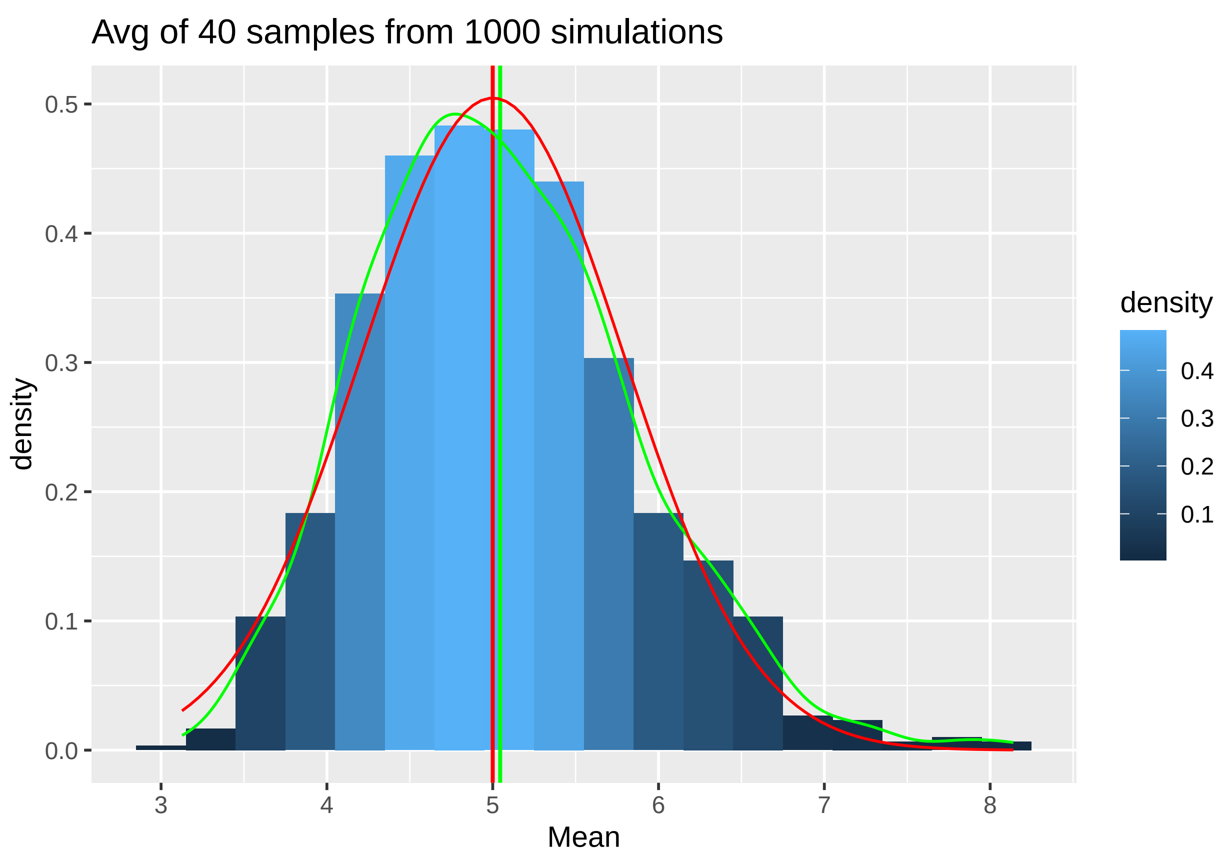
```r
s_var - theoretical_var  #calculating the diff b/w sample & theoretical variance
```

```
## [1] 0.01225439
```

Calculating the differences of the values show that the **sample** and the **Theoretical** means & variances are very much similar to one another.

Distributions

```r
ggplot(as.data.frame(mns),aes(x = mns)) +
  geom_histogram(aes(y = after_stat(density) ,fill = after_stat(density)),binwidth = 0.3) +
  labs(title = "Avg of 40 samples from 1000 simulations") +
  xlab("Mean") +
  geom_vline(xintercept = s_mean,color = "green",linewidth = 0.7) + #sample mean vertical line
  geom_vline(xintercept = theoretical_m,color = "red",linewidth = 0.7) + #Theoretical mean line
  geom_density(color = "green") + #density line for sample distribution
  stat_function(fun = dnorm,args = list(mean = theoretical_m,sd = sqrt(theoretical_var)),
               color = "red") #writing a function for density line of Theoretical distribution
```



In the above plot The Green lines represent **sample** mean & it's density while the red plot represents the **Theoretical** mean & it's density.The above distribution has a bell shaped curve which shows that it's approximately normal or *gaussian* distribution.

Conclusion

With the above simulation,plots we can conclude that the **Theoretical** & **sample** are very much approximately normal(due to *Central Limit Theorem*) even the Means & variances were also similar.

Thank you