Exponential Averages Distribution Analisys

Mateus Melo

14/09/2020

Project Overview

This report aims to investigate the exponential averages distribution. Simulating the mean of the distribution by collecting a large sample of averages, we will be able to see how it compares with the Central Limit Theorem.

The Central Limit Theorem

The Central Limit Theorem states that "the distribution of averages of iid variables becomes that of a standard normal as the sample size increases". To be precisely, the distribution of averages will be close to a standard normal and it will be possible to infer the population mean and the standard deviation of the population mean.

The distribution of averages will be in the form of a normal curve centered at the population mean and with the standard deviation equals to the standard deviation of the population mean divided by the square root of the sample size.

Simulation

To see this in practice, let us simulate the distribution of a thousand averages of 40 exponentials. We are going to use a lambda (rate parameter) of 0.2. Therefore, the mean and it's standard deviation will be equal to 5 (mean = sd = 1/lambda).

```r
exp_avg <- vector("numeric",1000)
for(i in 1:1000){
        exp_avg[i]<-sum(rexp(40,0.2))/40
}
exp_mean<-exp_sd<-5
exp_var<-25
exp_sample_mean<-mean(exp_avg)
exp_sample_sd<-sd(exp_avg)
exp_sample_var<-var(exp_avg)
cat("Difference of means: ",exp_mean-exp_sample_mean)
```

```
## Difference of means:  -0.02878052
```

```r
cat("Difference of standard deviations: ",exp_sd-exp_sample_sd*sqrt(40))
```

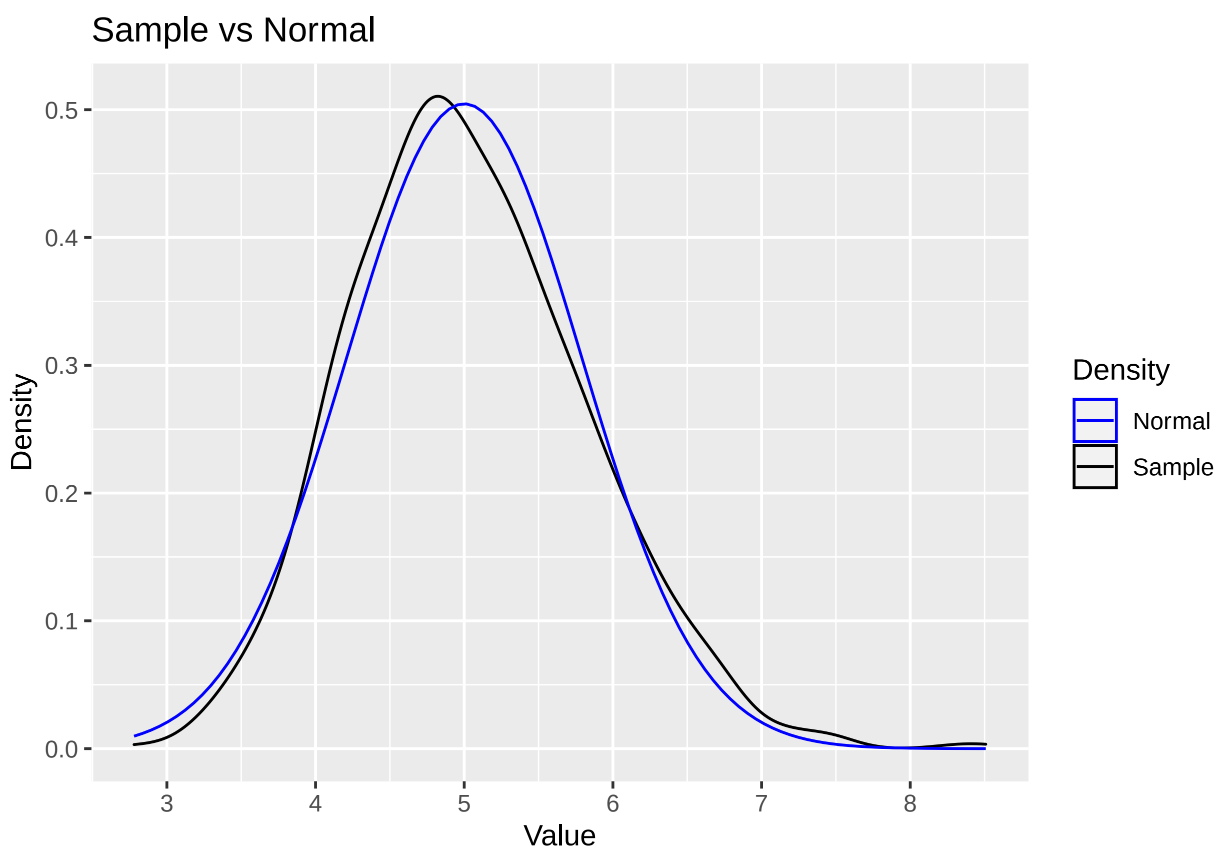
```
## Difference of standard deviations:  -0.02538638
```

```r
cat("Difference of variances: ",exp_var-exp_sample_var*40)
```

```
## Difference of variances:  -0.2545083
```

All the differences were below than 1%. Let us compare how the density os the samples compares with a normal curve centered at the population mean and with standard deviation set to the standard deviation divided by the squared root of the sample size.

```r
library(ggplot2)
ggplot()+geom_density(aes(exp_avg,colour="Sample"), )+
        stat_function(fun=dnorm,
        args=list(mean=exp_mean,sd=exp_sd/sqrt(40)), aes(colour="Normal"))+
        scale_color_manual("Density", values=c("blue","black"))+
        labs(title="Sample vs Normal", x="Value", y="Density")
```

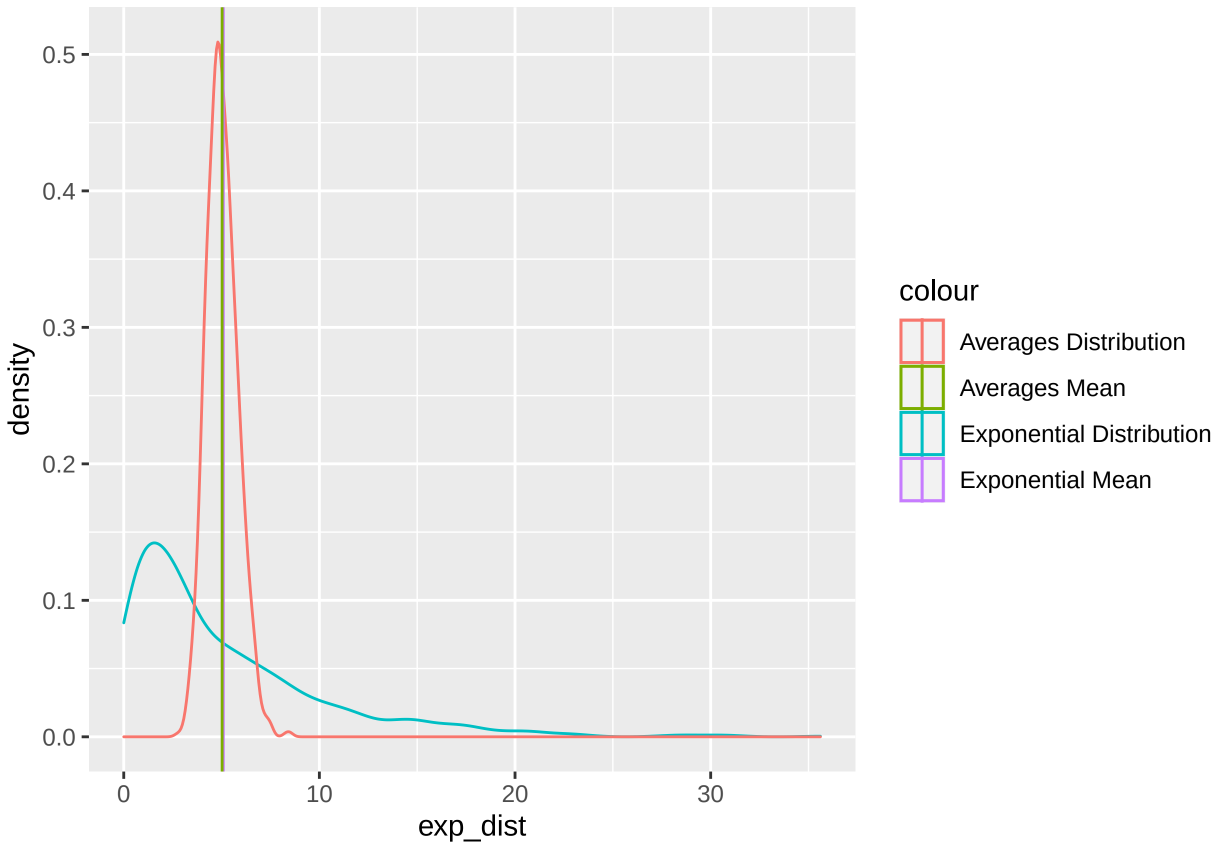


Pretty close. We were able to see in the plot above that the distribution of averages indeed behaves as the Central Limit Theorem predicted. Therefore, we could estimate with an error below than 1% the population mean and the standard deviation.

It is important to notice the difference between the distribution of averages of the exponential and the exponential distribution. Let us compare them both.

```r
exp_dist<-rexp(1000,0.2)
ggplot()+geom_density(aes(exp_dist, colour="Exponential Distribution"))+
        geom_density(aes(exp_avg, colour="Averages Distribution"))+
```

```
    geom_vline(aes(xintercept=mean(exp_dist),colour="Exponential Mean"))+
    geom_vline(aes(xintercept=mean(exp_avg),colour="Averages Mean"))
```



The means are so close that we are barely not able to distinguish them. However, the curves are very different. This happens because the exponential density is far more variable than the average density. To understand this, we have to recall that the exponential variance is 40 times larger than the averages variance. The standard deviation of the distributions of averages is also called standard error.