The Central Limit Theorem and Inferential Analysis

Byron Rodriguez

Overview

The Central Limit Theorem states that drawing a sample from a set of averages of multiple random distributions, no matter what kind of, their average will estimate the mean of the whole population being studied. This is a very important characteristic, since it is possible to estimatete the statistical parameters of the population in study, even with a few samples.

Simulation with Exponential Distribution

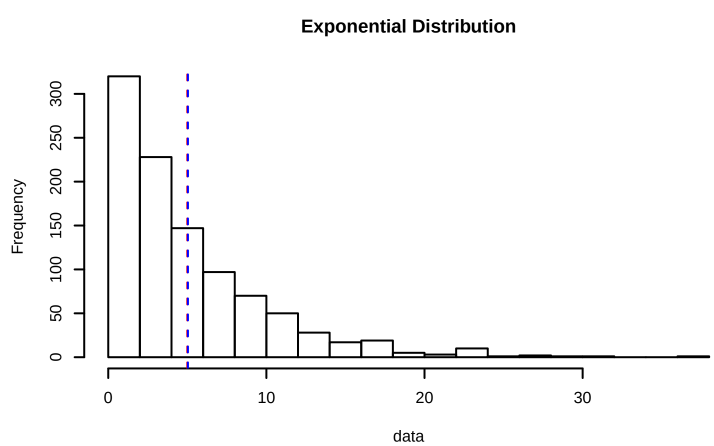
The exponential distribution has a parameter lambda, whose inverse is equal to the theoretical mean and the theoretical variance of this distribution.

The following R code generates 1000 observations distribuited exponentialy with lambda 0.2.

```
lambda = 0.2; data = rexp(1000,lambda);
```

Then, it is possible to visualize the distribution through an histogram, and compare graphically the theoretical mean against the average of the data of the simulation.

```
t_mean = 1/lambda;r_avg = mean(data);
hist(data,breaks=25,main="Exponential Distribution");
abline(v=t_mean,col='red',lty=2);abline(v=r_avg,col='blue',lty=2);
```



The red dashed line shows the tehoretical mean, while the blue doashed line shows the average of the data. With 1000 observations the two values are colse. In fact, the difference is:

```
t_mean - r_avg
```

```
## [1] -0.03781135
```

The variance of the data is:

```r
var(data)
```

```
## [1] 24.74621
```

The difference between the theoretical standard deviation and the standard deviation of the sampled data is:

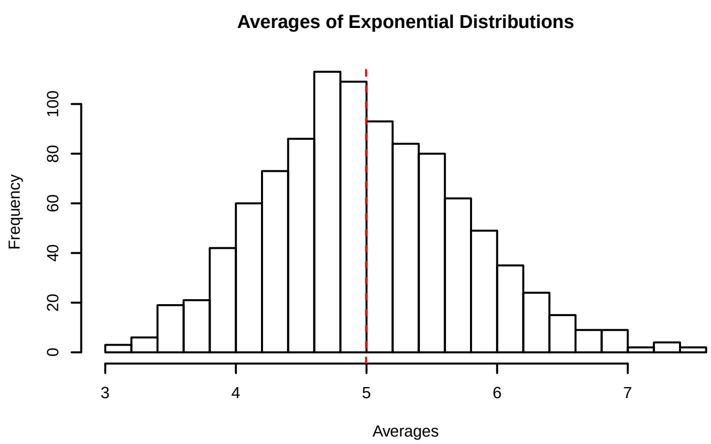
```r
(1/lambda)^2 - var(data)
```

```
## [1] 0.253788
```

It is evident that the value of the real variance is close to the theoretical variance.

In order to analyse the Central Limit Theorem, we are going to take the averages of 1000 simulations of exponentials with 40 observations each one.

```r
mean_s<-c();for(i in 1:1000){mean_s <-c (mean_s , mean(rexp(40,lambda)));}
hist(mean_s,breaks=25,main="Averages of Exponential Distributions",xlab="Averages");
abline(v=mean(mean_s),col='red',lty=2);
```



In this graphic, the red dashed line shows the average of the data, whose value is:

```r
mean(mean_s)
```

```
## [1] 4.996761
```