Statistical Inference Project - Exponential Distribution and the CLT

Felipe Godoi Rosario

Overview

In this report I will investigate the exponential distribution and compare it with the Central Limit Theorem.

The Central Limit Theorem (CLT) is one of the most important theorems in statistics. It states that the distribution of averages of *iid* variables (properly normalized) becomes that of a standard normal as the sample size increases.

The comparisson will be made between the distribution of averages of 40 exponentials (doing a thousand simulations) and the exponential distribution. The mean of the exponential distribution is 1/lambda and the standard deviation is also 1/lambda, I'll use a fixed lambda = 0.2 for all simulations.

Sample Mean versus Theoretical Mean

The theoretical mean of the exponential distribution is 1/lambda. As our lambda = 0.2, our theoretical mean is:

```
tm <- 1/0.2
```

```
## [1] 5
```

If we calculate the mean of a simulated exponential distribution of 1,000 cases we'll see that it's very close to 5:

```
set.seed(10)
mean(rexp(1000,0.2))
```

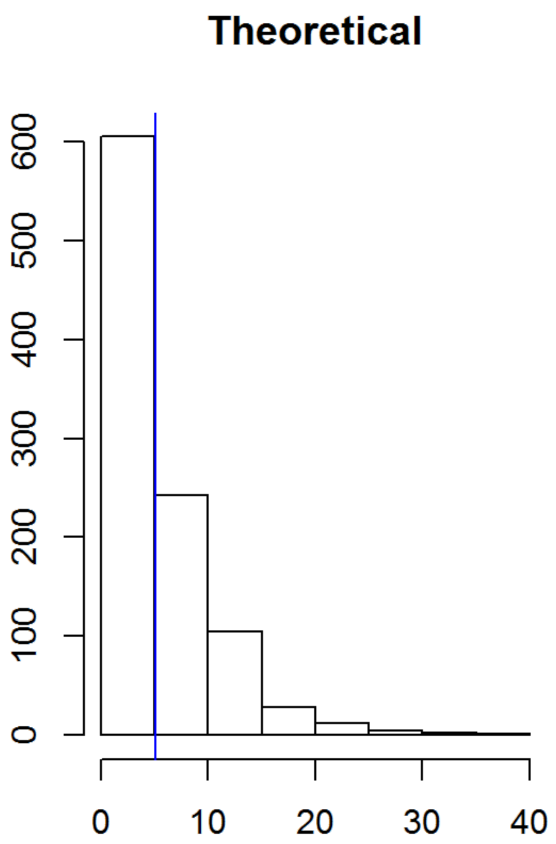
```
## [1] 5.11491
```

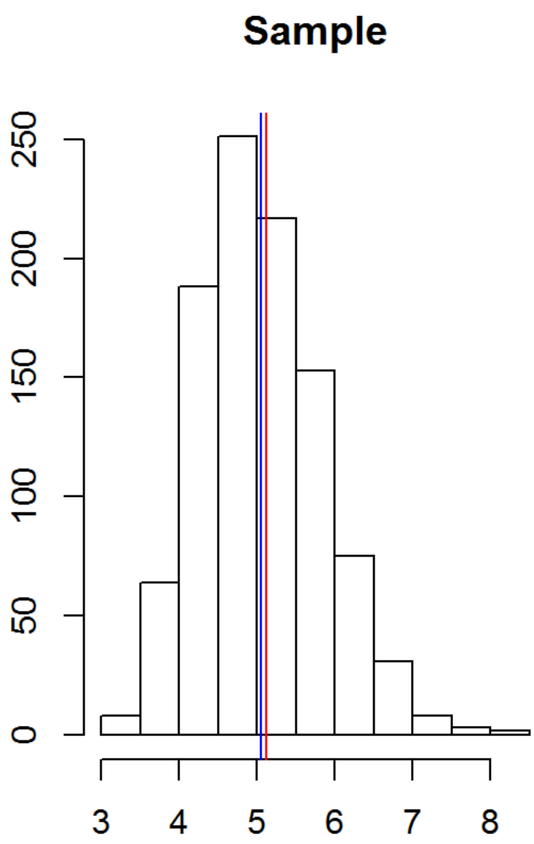
Now let's simulate 1,000 averages of 40 exponentials, find the mean of the distribution of these averages and compare to the theoretical mean:

```
set.seed(10)
vmeans <- NULL
for (i in 1 : 1000) vmeans = c(vmeans, mean(rexp(40,0.2)))
mean(vmeans)
```

```
## [1] 5.04506
```

As we can see, the two means are nearly equal. The following plots show the exponential distribution, the sample of averages distribution and their means, for comparisson. The theoretical mean is colored in red and the sample mean, blue:





The variances, on the other hand, are going to be different. The variance of the distribution of averages, if normalized, must be close to 1. The following code shows this is true:

```
set.seed(10)
var(rexp(1000,0.2))
```

```
## [1] 22.6982
```

```
set.seed(10)
vnmeans <- NULL
for (i in 1 : 1000) vnmeans = c(vnmeans, ((mean(rexp(40,0.2))-(1/0.2))/((1/0.2)/sqrt(40))))
var(vnmeans)
```

```
## [1] 1.019607
```

Finally, to show that the distribution of averages is approximately Normal, let's simulate a standard Normal distribution, calculate its mean and variance and compare with the distribution of averages from above.

```
set.seed(10)
mean(rnorm(1000))
```

```
## [1] 0.01137474
```

```
var(rnorm(1000,1))
```

```
## [1] 1.089982
```

```
set.seed(10)
vnmeans <- NULL
for (i in 1 : 1000) vnmeans = c(vnmeans, ((mean(rexp(40,0.2))-(1/0.2))/((1/0.2)/sqrt(40))))
mean(vnmeans)
```

```
## [1] 0.05699637
```

```
var(vnmeans)
```

```
## [1] 1.019607
```

We can see that the values are very close to each other. The histograms of the 2 distributions are very similar, too: