Statistical Inference: Simulating Exponential Distributions

Henrique Souza (github.com/htssouza)

Overview

This report is intended to investigate the exponential distribution in R and compare it with the Central Limit Theorem (CLT).

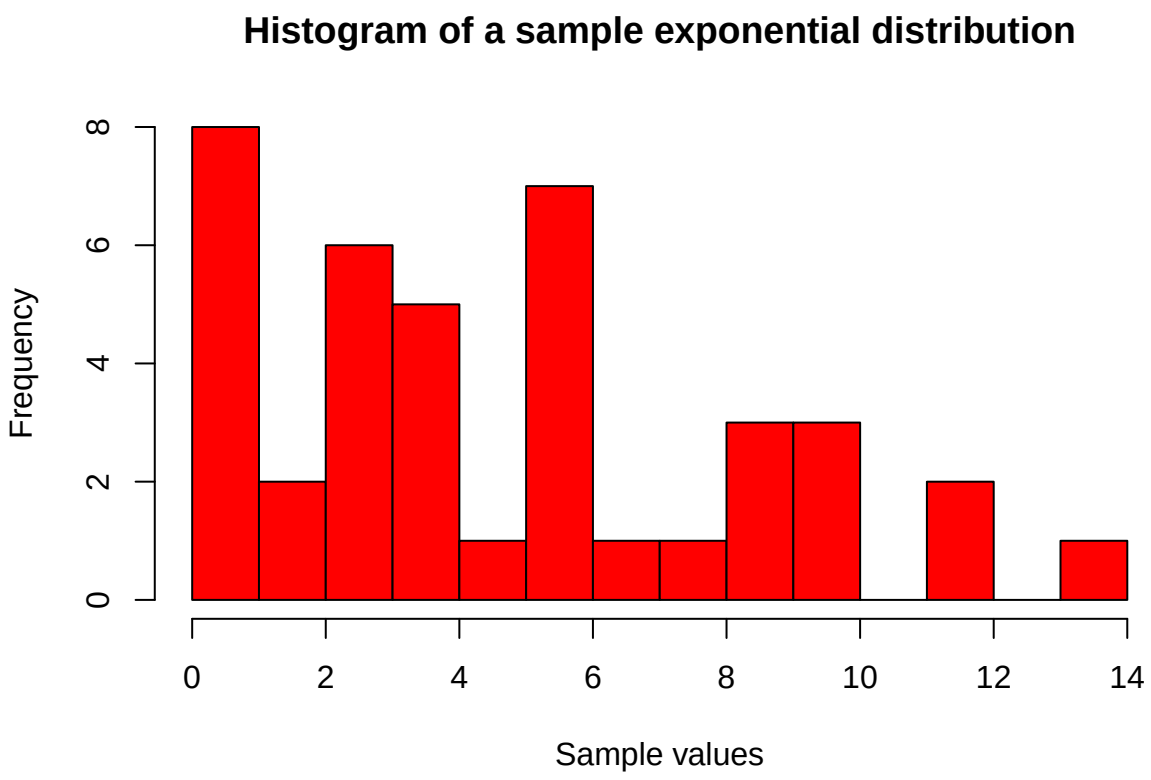
Simulations

For all the simulations lambda will be equal to 0.2. A seed will also be set.

```
lambda <- 0.2
set.seed(123456)
```

Now lets generate some sample values and plot their histogram.

```
sample <- rexp(40, lambda)
```



Sample Mean versus Theoretical Mean

```
sample_mean <- mean(sample)
theoretical_mean <- 1 / lambda
sample_mean
```

```
## [1] 4.671228
```

```
theoretical_mean
```

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

You can see that their values are pretty close.

Let's test with a larger sample size (1000) and see if it gets even closer:

```
larger_sample <- rexp(1000, lambda)
larger_sample_mean <- mean(larger_sample)
larger_sample_mean
```

```
## [1] 5.227839
```

Theyr are not even closer.

Sample Variance versus Theoretical Variance

Let's compute the sd of our sample.

```
sample_sd <- sd(sample)
theoretical_sd <- 1 / lambda
sample_sd
```

```
## [1] 3.638052
```

```
theoretical_sd
```

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

Not so close to the expected value.

Let's now repeat the comparison with the larger sample.

```
larger_sample_sd <- sd(larger_sample)
larger_sample_sd
```

```
## [1] 5.228888
```

Which is much better.

Distribution

Computing a thousand exponential distributions and the same number of normal distributions.

```
means = NULL
for (i in 1 : 1000) means = c(means, mean(rexp(40, lambda)))
ndist <- rnorm(1000, mean = mean(means), sd = sd(means))
```

Plotting both so they can be compared.