Central Limit Theorem applied to Exponential Distribution

Leandro GARCIA

OVERVIEW

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem.

PROCEDURE

The exponential distribution is simulated in R with `rexp(n, lambda)` where lambda is the rate parameter. The mean (μ) of exponential distribution is $\frac{1}{\lambda}$ and the standard deviation (σ) is also $\frac{1}{\lambda}$. For all the simulation, lambda is set to 0.2. We will simulate 1000 sample means, using 40 exponentials.

THEORETICAL VALUES

In this section we calculate the theoretical values for the exponential distribution distribution

Theoretical Mean: ($\mu_t = \frac{1}{\lambda}$)

```
lambda_t <- 0.2 # Value for simulation
mu_t <- 1/lambda_t
mu_t # 5.0
```

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

Theoretical mean ($\mu_t = 5.0$)

Theoretical Standard Desviation ($\sigma_t = \frac{1}{\lambda}$)

```
# lambda_t <- 0.2
sigma_t <- 1/lambda_t
sigma_t # 5.0
```

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

Theoretical Standard Deviation: ($\sigma_t = 5.0$)

Theoretical Variance ($Var_t = \sigma^2$)

```
# sigma_t <- 5.0
var_t <- sigma_t^2
var_t # 25.0
```

```
## [1] 25
```

Theoretical Variance: ($Var_t = 25.0$)

Theoretical Standard Error of the mean : ($SE_{\overline{X}} = \frac{\sigma}{\sqrt(n)}$)

```
sample_size <- 40
SE_t <- sigma_t / sqrt(sample_size)
SE_t # 25.0
```

```
## [1] 0.7905694
```

Theoretical Standar Error of the mean : ($SE_{\overline{X}} = 0.79$)

PRACTICAL VALUES

In this section we calculate the random values of the distribution, and then we will proceed to retrieve the initial parameter (inference).

SIMULATING

Getting the values to Simulate 1000 times 40 $X_{1...40}$, and then obtain 1000 $\overline{X}_i$ or in other words $\overline{X}_{1...1000}$

```
# Parameters for simulation
set.seed(1234)
means_size <- 1000
sample_size <- 40

# Generate the samples
data_list = rexp(means_size * sample_size, rate = lambda_t) # to be used for plotting
data_matrix <- matrix( data = data_list , nrow = means_size, ncol = sample_size)

# Getting the means for all the simulation
means <- apply(X = data_matrix, MARGIN = 1, FUN = mean )

# Simulated means based on the parameters
str(means)
```

```
##  num [1:1000] 4.6 6.02 5.46 4.18 7.14 ...
```

Practical mean: ($\mu_p = \sum_{i=1}^{1000} \frac{\overline{X}_i}{n}$)

```
?sd
```

```
## starting httpd help server ... done
```

```
mu_p <- mean(means)
mu_p
```

```
## [1] 4.974239
```

Practical mean: ($\mu_p = 4.97$)

Practical Standard Error of the mean: ($SE_{\overline{X}} = \sigma_{\overline{X}}$)

```
?sd
SE_p <- sd(means)
SE_p
```

```
## [1] 0.7713431
```

Practical Standard Error of the mean: ($SE_{\overline{X}} = 0.77$)

Practical Population Standard Deviation: ($\sigma_p = SE_{\overline{X}} * \sqrt(n)$)

```
?sd
sigma_p <- SE_p * sqrt(sample_size)
sigma_p
```

```
## [1] 4.878402
```

Practical Population Standard Deviation: ($\sigma_p = 4.88$)

Practical Population Variance: ($Var_p = \sigma^2$)

```
?sd
var_p <- sigma_p ^ 2
var_p
```

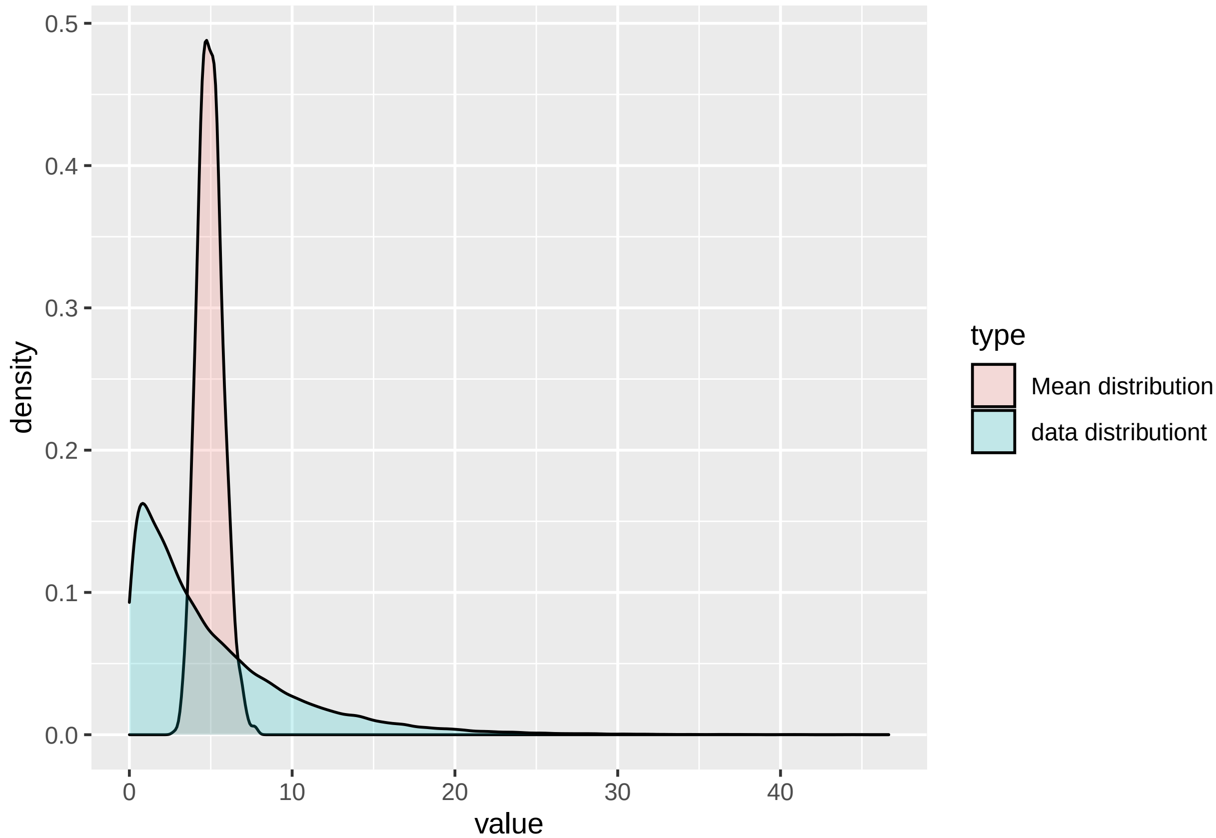
```
## [1] 23.79881
```

Practical Population Variance: ($Var_p = 23.8$)

THEORETICAL VALUES Vs. PRACTICAL VALUES

```
library(ggplot2)
means_df<- data.frame(value = means, type = rep("Mean distribution", length(means)))
data_df <- data.frame(value = data_list, type = rep("data distributiont", length(data_list)))
all_data <- rbind(means_df, data_df)

ggplot(data = all_data, aes(x = value, fill = type)) + geom_density(alpha = 0.2)
```



STATISTIC	THEORETICAL	PRACTICAL
μ	5.0	4.97
Var	25.0	4.88
σ	5.0	23.8
$SE_{\overline{X}}$	0.79	0.77

CONCLUSIONS:

Using random values of a exponential distribution, we were able to :

- Demonstrate the relationship that exist between the distribution of the average of the means and the population means.
- Prove that the distribution of the means is a *normal distribution* even when the population distribution is not.
- Confirm that the relationship between the population distribution and means distribution is dictated by the *Central Limit Theorem*.