Central Limit Theorem: An Easy Demostration

Ian Castillo Rosales

November, 21

In this short paper we are going to study the **Central Limit Theorem (CLT)**. This theorem gives you the ability to measure how much the means of various samples will vary, without having to take any other sample means to compare it with. By taking this variability into account, you can use your data to answer questions about a population

We simulated forty times a number with a exponential distribution with a parameter equal to 0.2 (exp(0.2)), then we took the average of the sample and repeated this process thounsand of times to obtain the distributión of averages of this samples (40 exp(0.2)s).

We did this in the best easy way, first, we set the parameters

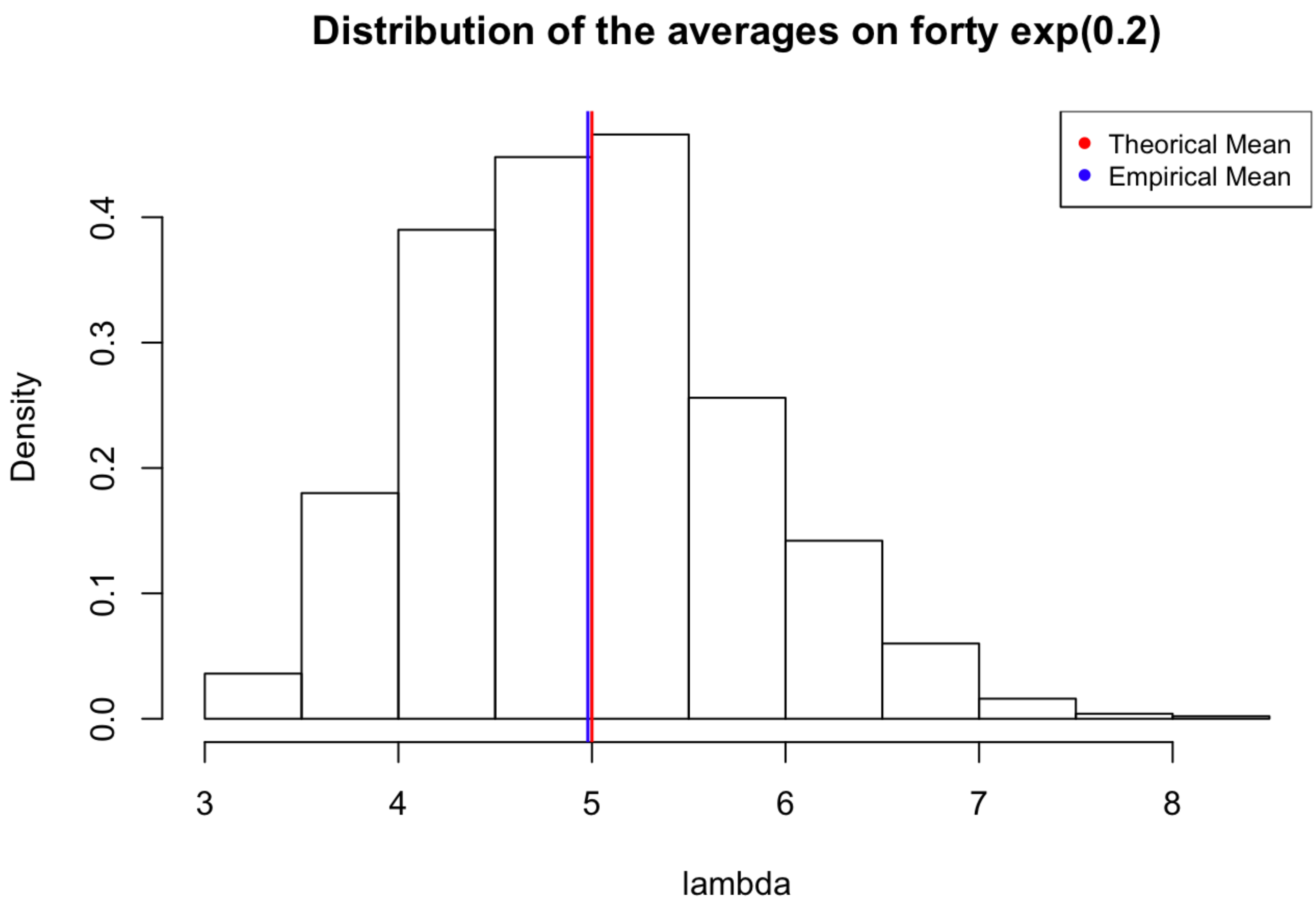
```
lambda <- 0.2
n_sim <- 40
n <- 1000
media <- 1/lambda
sd <- 1/lambda
```

Then que simulate the means of 40 exp(0.2)s

```
medias <- NULL
for(i in 1:n) medias <- c(medias, mean(rexp(n_sim, lambda)))
```

Plot the distribution of the averages and show the Theorical Mean ($\frac{1}{\lambda}$) and the Emprical Mean

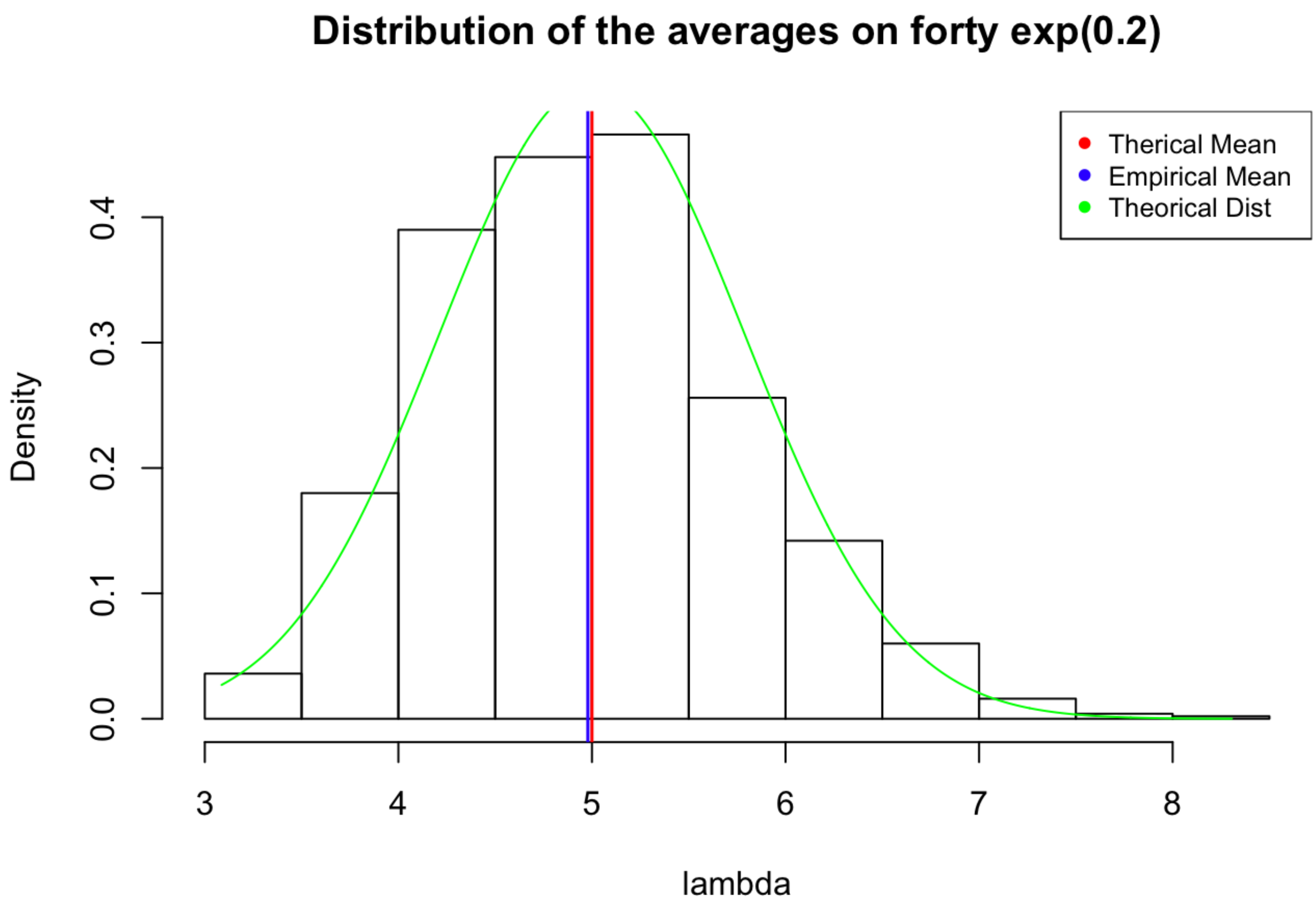
```
hist(medias, main = "Distribution of the averages on forty exp(0.2)", xlab = "lambda", freq =
 F)
abline(v = 1/lambda, col = "red", lwd = 1.5)
abline(v = mean(medias), col = "blue", lwd = 1.5)
legend("topright", legend = c("Theorical Mean", "Empirical Mean"), col = c("red", "blue"), ce
x = 0.8, pch = 16)
```



To compare the distribution with the theorical distribution (a normal distribution) we put the line of a shell distribution

```
hist(medias, main = "Distribution of the averages on forty exp(0.2)", xlab = "lambda", freq =
 F)
abline(v = 1/lambda, col = "red", lwd = 1.5)
abline(v = mean(medias), col = "blue", lwd = 1.5)

x <- seq(min(medias), max(medias), length = n)
normales <- dnorm(x, mean = media, sd = sd/sqrt(n_sim))
lines(x, normales, col = "green")
legend("topright", legend = c("Therical Mean", "Empirical Mean", "Theorical Dist"), col = c("
red", "blue", "green"), cex = 0.8, pch = 16)
```



Moreover, we know that the variance of the theorical distribution should be $(\frac{1}{\lambda}^2$, and from the samples we can compare this with $variance(means) \cdot number\ of\ simulations$, then, the absolute error of the difference between the Thoerical Variance and the Empirical Variance is:

```
teo_var <- (1/lambda)^2
emp_var <- var(medias)*n_sim
(var_err <- abs(teo_var - emp_var))
```

```
## [1] 0.9913
```

This error is expresed in square units, the error of standar deviation is:

```
sqrt(var_err)
```

```
## [1] 0.9956
```