Statistical Inference Course Project

Humbert Costas

22 de agosto de 2015

Overview

In this project I will **investigate the exponential distribution** in R **and compare it with the Central Limit Theorem**. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. I will ***set lambda = 0.2*** for all of the simulations. I will investigate the distribution of averages of 40 exponentials.

```
set.seed(666)    # Reproducibility purposes
n <- 40          # Number of samples
lambda <- .2      # Set lambda to 0.2
nsims <- 1000    # Number of simulations
```

Simulations

Illustration via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials.

```
# The sample simulation
sam = NULL
for (i in 1:nsims) sam = c(sam, mean(rexp(n,lambda)))

# Mean and Variabce of the exponential distributions
s_mean <- mean(sam)
s_sd <- sd(sam)
s_var <- var(sam)

# Theoretical Mean and Variance
t_mean <- 1 / lambda
t_var <- ((1 / lambda) ^ 2) / n
t_sd <- sqrt(t_var)
```

Results

1. Comparision of the sample mean and the theoretical mean of the distribution:

- Sample Mean (s_mean): 4.9878184
- Theoretical Mean (t_mean): 5

They are really close.

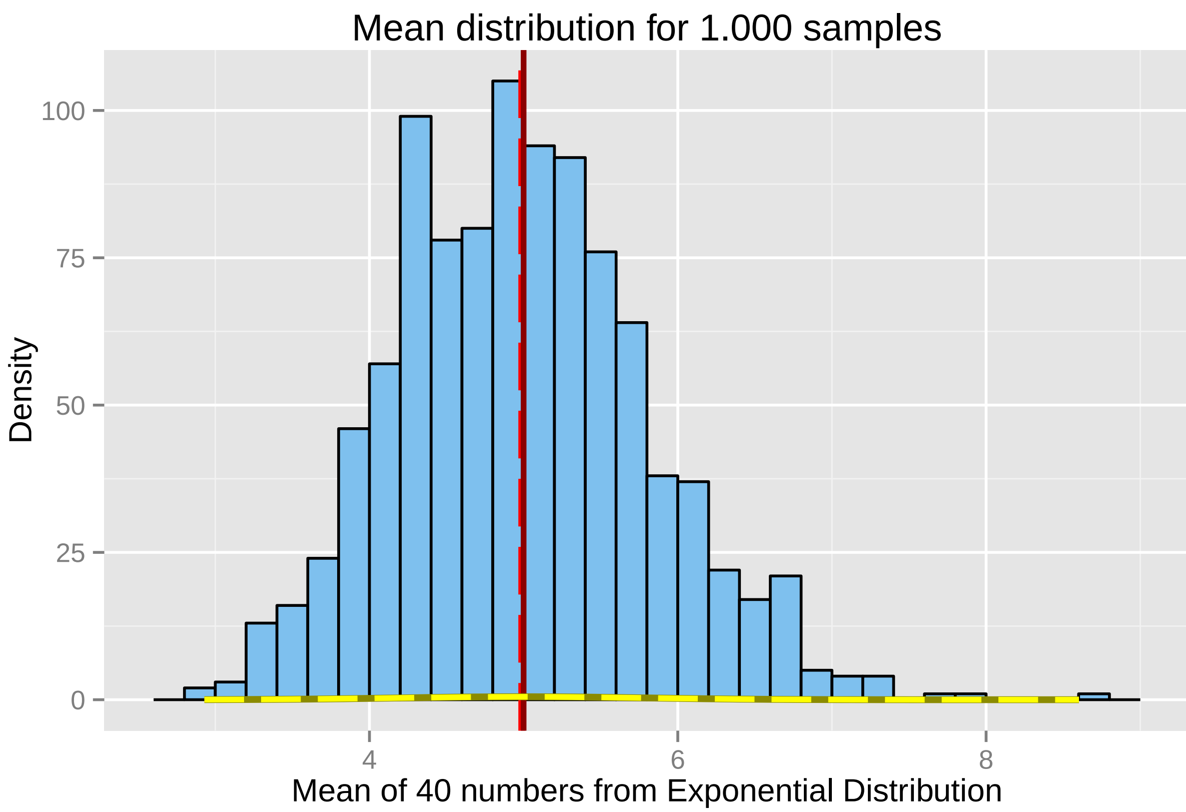
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

- Sample Variance (*s_var*): 0.6646822
- Theoretical Variance (*t_var*): 0.625

They are really close.

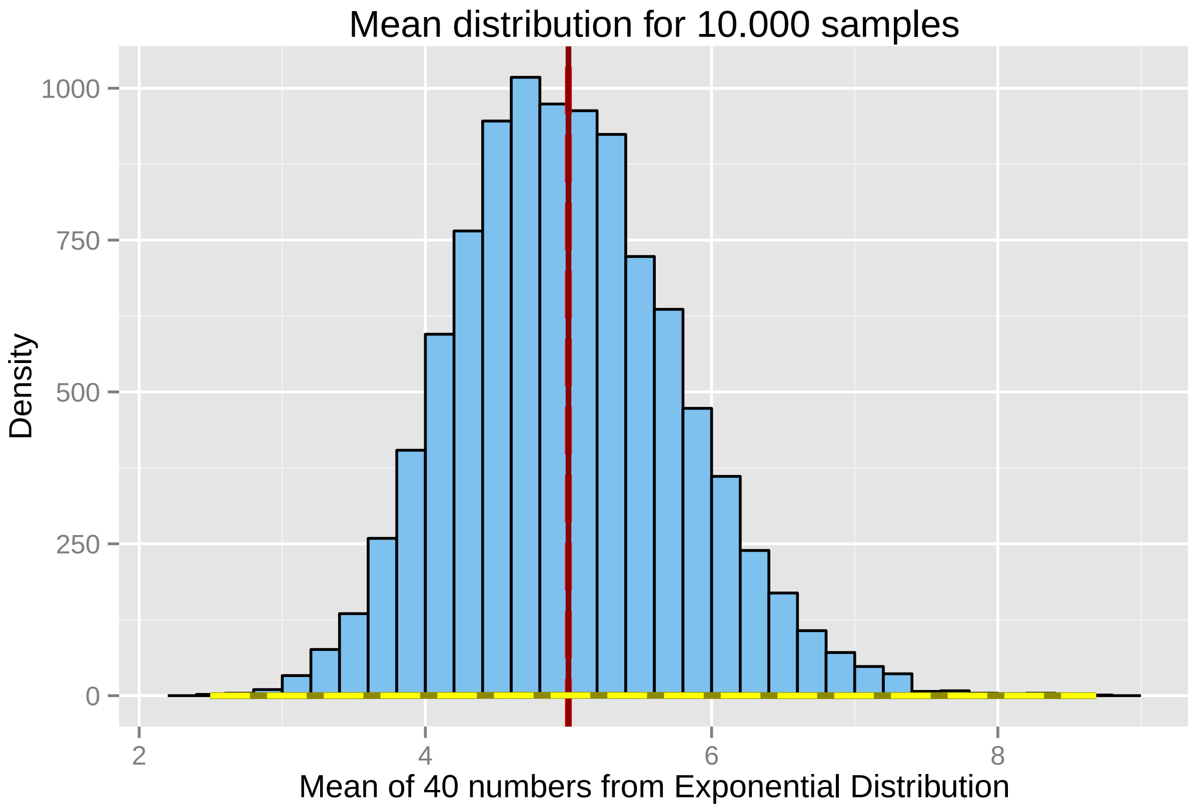
3. Show that the distribution is approximately normal.

```
library(ggplot2)
show_distribution <- function(sam, title)
{
    df_sample_means <- data.frame(sam)
    g <- ggplot(data = df_sample_means, aes(x = sam))
    g <- g + geom_histogram(binwidth = lambda, fill = "skyblue2", color = "black")
    g <- g + geom_vline(xintercept = s_mean, size = 1.2, color = "red", linetype = "longdash")
    g <- g + geom_vline(xintercept = t_mean, size = 1, color = "red4")
    g <- g + stat_function(fun = dnorm, args = list(mean = t_mean, t_sd), color = "yellow4", size = 1.2
    g <- g + stat_function(fun = dnorm, args = list(mean = s_mean, s_sd), color = "yellow", size = 1, l
    g <- g + labs(title = title, x = "Mean of 40 numbers from Exponential Distribution", y = "Density")
    g
}
print(show_distribution(sam, "Mean distribution for 1.000 samples"))
```



The results shows that the sample means distribution (*red dashed*) is very close to the theoretical center of the distribution (*strong red*).

It seems that the distribution is approximately normal. Let's see what happens if we simulate 10.000 instead of 1.000 to show a better normal distribution.



As we can see, it's approximately a normal distribution.

Conclusion

This is an example that supports the demostration that the Central Limit Theorem apply to the exponential distribution. The means for an exponential distribution is approximately a normal distribution with sample mean and sample variance equals to population mean and variance.