Statistical Inference - Simulation

Rafael Lavagna

5 de abril de 2018

Overview

In this project we will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The CLT states that the distribution of averages of iid variables (properly normalized) becomes that of a standard normal as the sample size increases. Moreover, it states that the average distribution is centered at the population mean and its standard deviation is equal to the standard error of the mean.

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. We will set lambda = 0.2 for all of the simulations. We will investigate the distribution of averages of 40 exponentials by doing do a thousand simulations.

Simulations

We are going to build a matrix with 1000 rows each of them containing 40 exponentials. We are going to calculate the mean of each row and store them in a vector named "means".

```
n <- 40
mu.theoric <- 1/0.2
sd.theoric <- 1/0.2/sqrt(n)
set.seed(1899)
exp <- rexp(40000,0.2)
M <- matrix(exp,1000,40)
means <- apply(M,1,mean)
```

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

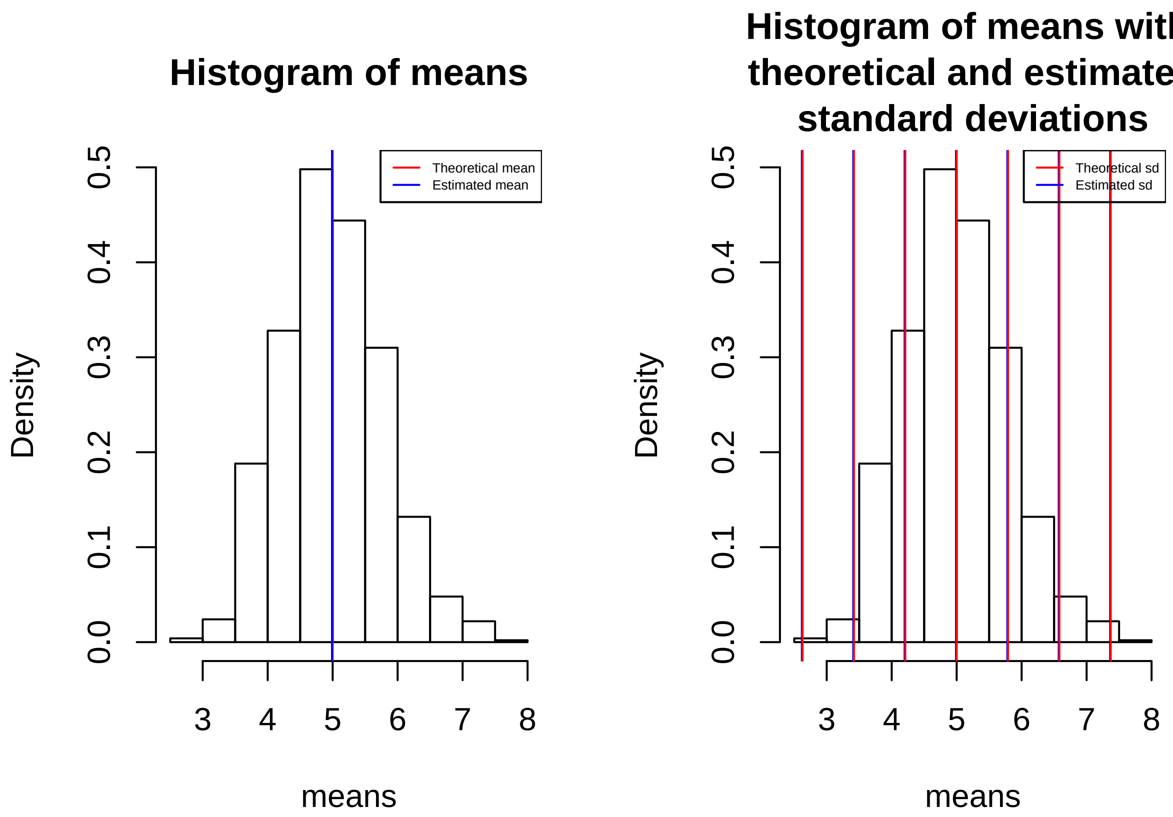
In this part, we will draw a histogram with the means (left image). In the picture of the right we are going to plot the same histogram but this time we are going to draw vertical lines at the points where theoretical and estimated +/- 3,2 and 1 standard devaitions are.

```
est.mu.population <- mean(means)
sd.estimated <- seq(est.mu.population - 3 * sd(means), est.mu.population + 3 * sd(means),
by = sd(means))
sd.theo <- seq(mu.theoric - 3 * sd.theoric, mu.theoric + 3 * sd.theoric, by = sd.theoric)

par(mfrow=c(1,2))

myhist <- hist(means, freq = FALSE)
abline(v=mu.theoric, col="red")
abline(v=est.mu.population, col="blue")
legend("topright", legend=c("Theoretical mean", "Estimated mean"),
       col=c("red", "blue"), lty = c(1,1), cex=0.45)

myhist2 <- hist(means, freq = FALSE, main = "Histogram of means with
theoretical and estimated
standard deviations")
abline(v=sd.estimated, col ="blue")
abline(v=sd.theo, col ="red")
legend("topright", legend=c("Theoretical sd", "Estimated sd"),
       col=c("red", "blue"), lty = c(1,1), cex=0.45)
```



We can clearly observe in the plot that the estimated mean (in blue) is really close to the theoretical value (in red).

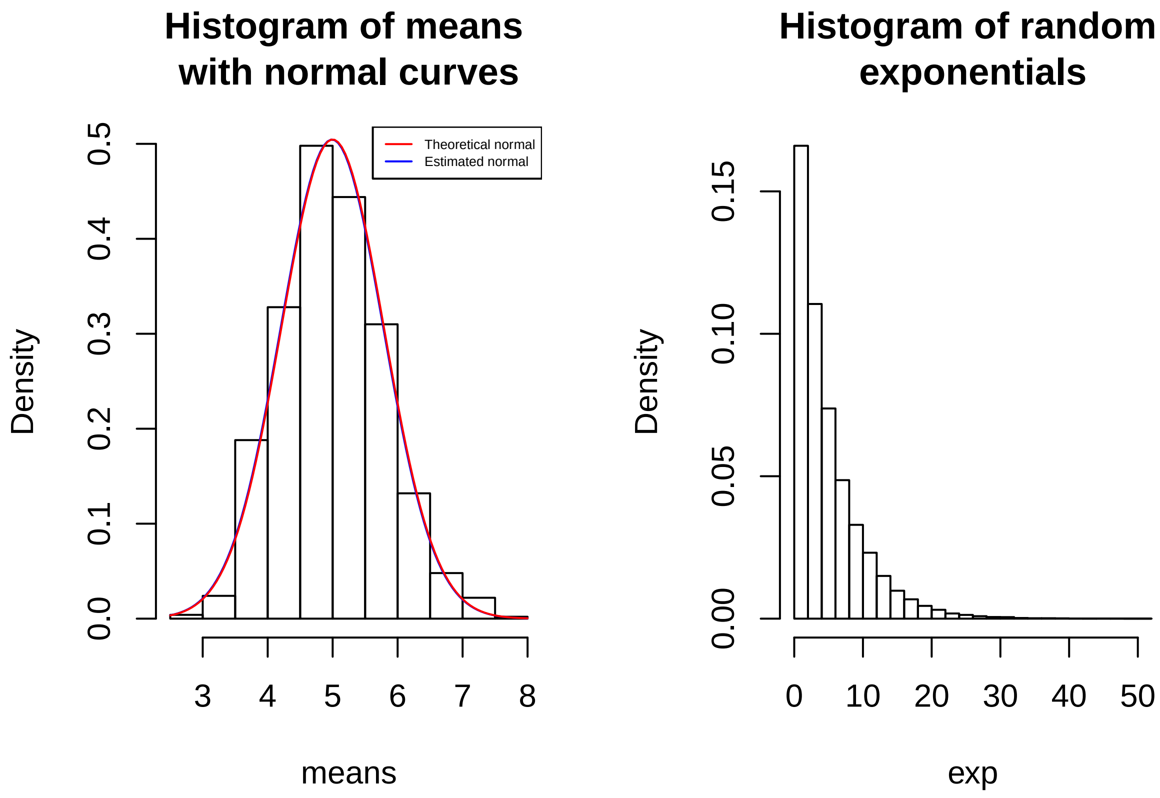
We can also observe in the picture on the right that the CLT gives a very precise estimation for the standard deviation since the theoretical values (red lines) are really close to the estimated ones (blue lines).

In both cases, the estimations are so close to the theoretical values that it is hard to distinguish them in the plots.

3. Show that the distribution is approximately normal.

```
par(mfrow=c(1,2))
myhist3 <- hist(means, freq = FALSE, main="Histogram of means
with normal curves")
curve(dnorm(x,mean=est.mu.population, sd=sd(means)),col="blue",add=TRUE)
curve(dnorm(x,mean=mu.theoric, sd=sd.theoric),col="red",add=TRUE)
legend("topright", legend=c("Theoretical normal", "Estimated normal"),
       col=c("red", "blue"), lty = c(1,1), cex=0.45)
```

```r
hist(exp, freq=FALSE, breaks = 20, main="Histogram of random
exponentials")
```



In this figure we observe two important things. First of all, how the average distribution (the histogram on the left) approximates to a normal by drawing with the histogram two normal curves, the red one with the theoretical parameteres and the blue one with the estimated ones. Once again, the estiamted values are so close to the theoretical values that both curves are almost overlapped in the plot. Another interesting aspect of this figure is to clearly observe the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.