Empirical analysis on Exponential distribution and Central Limit Theorem

Alessandro Rotatori

27 giugno 2017

Synopsis

The objective of the following empirical analysis is to describe the characteristics of the exponential distribution with respect to the simple statistics, mean and variance. Furthermore, the validity of Central Limit Theorem for estimating the population mean is evaluated.

Analysis

The Empirical Analysis is performed by analysing the behavior of the exponential distribution, more in details analyzing the properties of 40 X_i, random variables $\sim Exp(\lambda)$.

The parameters relevant for the analysis, therefore, are the following:

- $\lambda = 0.2$, as provided in the assignment, is the *rate* parameter;
- n, number of values sampled from the theoretical distribution;
- *sim*, number of simulations.

```
lambda <- 0.2
n <- 40
sim <- 1000
mn <- 1 / lambda
sd <- 1 / lambda
```

Therefore the theoretical mean and variance of the exponential distribution under analysis are as follows:

```
theo.val <-c(Mean = mn, Variance = sd^2)
theo.val
```

```
##     Mean Variance
##        5       25
```

In order to evaluate the Exponential distribution with $\lambda = 0.2$ under the Central Limit Theorem, as previously stated, the empirical analysis to be performed requires to evaluate the distribution of means of 40 iid random variables $\sim Exp(\lambda)$. Therefore, 1000 simulation of the mean of 40 random variables were performed in order to create a sufficiently large distribution. The R code is as follows:

```
mns <- numeric(sim)
set.seed(10)
for (i in 1:sim) mns[i] <- mean(rexp(n, lambda))
```

Remember that the *set.seed* function is required in order to guarantee the replicability of the analysis. The vector *mns* has the values of the 1000 means of 40 iid Exponential. We know that the mean and variance of the distribution of means are a proxy for the theoretical mean and variance (divided by the number of observation n) of the original distribution. The following code shows the evidence:

```
data.frame(Estimated = c(mean(mns), var(mns)), Theoretical = c(mn, sd^2/n),
          row.names = c("Mean", "Variance"))
```

```
##          Estimated Theoretical
## Mean     5.0450596       5.000
## Variance 0.6372544       0.625
```

Moreover, what the Central Limit Theorem (CLT) states is that the distribution of averages of iid variables (normalized) becomes that of a standard normal as the sample size increases, $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/2)$. Therefore, by the CLT, we can say that the empirical distribution, the vector *mns*, is $\sim \mathcal{N}(5, 5^2/40)$ as $n = 40$ increases. The following code shows the evidence, by comparing the distribution of the means of the 40 iid Exponentials with the distribution of $n\ x\ sim = 40000$ iid Exponential random variables. First the simulation of the empirical distribution of 40000 iid Exponential random variables:

```
set.seed(11)
sim.vector <- rexp(sim * n, lambda)
```

then the result is as follows:

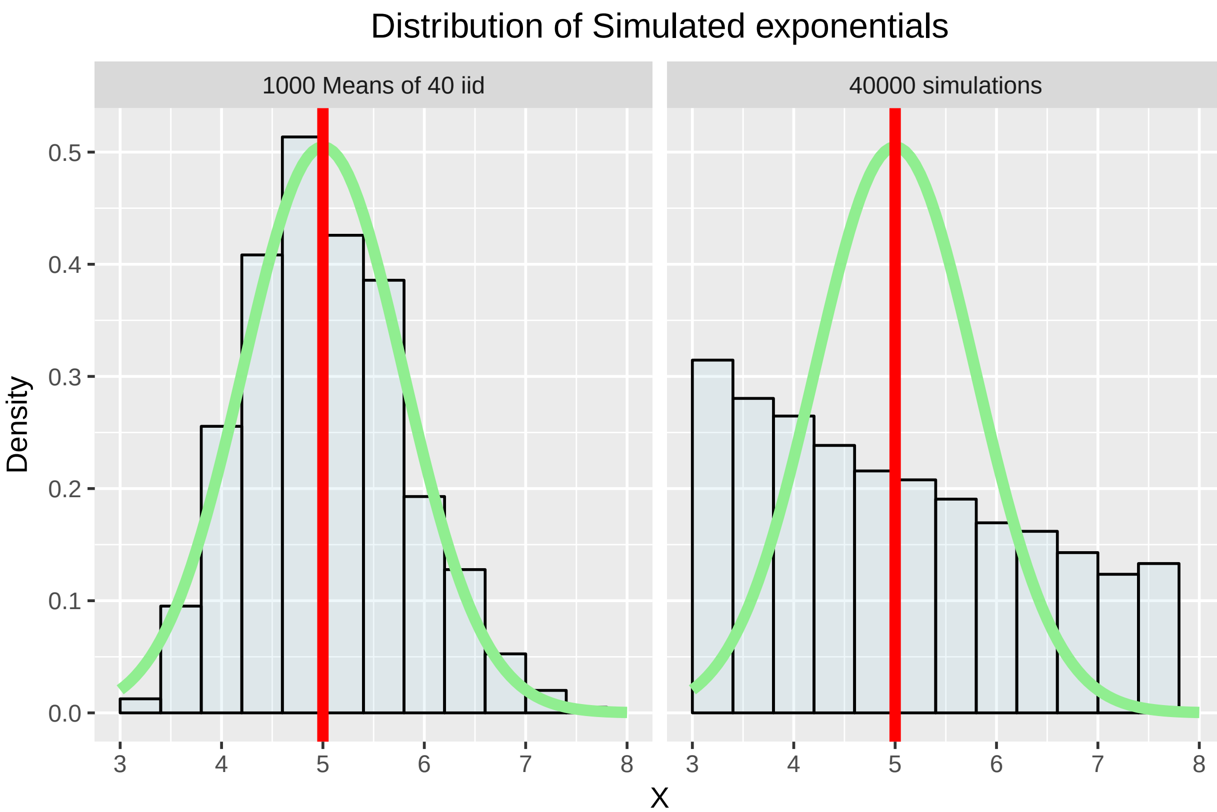

Figure 1

The first of the two graph shows the distribution of the means of the 40 iid Exponential variables, with imposed the:

- expected normal distribution, light green line;
- the theoretical mean, red vertical line.

As expected the distribution of means is centered on the theoretical mean and it is far more normal than the distribution of the 40000 iid Exponential (rightern graph).

As last step of the analysis, we could test the value of the empirical mean of the distribution of means against the associated normal distribution defined with the CTL.

Appendix

The following is the code used to generate the plot in figure 1.

```
library(ggplot2)
dat <- data.frame(x = c(sim.vector, mns), test = c(rep("40000 simulations", sim * n), rep("1000 Means o
g <- ggplot(dat, aes(x = x, fill = test))
g <- g + geom_histogram(alpha=.20, color = "black", fill = "lightblue", binwidth = 0.4, aes(y = ..densi
g <- g + stat_function(fun = dnorm, size = 2, args = list(mean = mn, sd = sd/sqrt(n)), color ="lightgre
g <- g + geom_vline(xintercept = mn, size = 2, colour = "red")
g <- g + facet_grid(. ~ test)
g <- g + labs(caption = paste("Figure ", 1), title = "Distribution of Simulated exponentials", y = "Dens
g <- g + theme(plot.title = element_text(hjust=0.5))
g <- g + xlim(3, 8)
g
```