Exponential distribution investigation

Federico Calore

24 Oct 2015

This document investigates the exponential distribution using R simulations, and compares it with the distribution of averages of 40 exponentials.
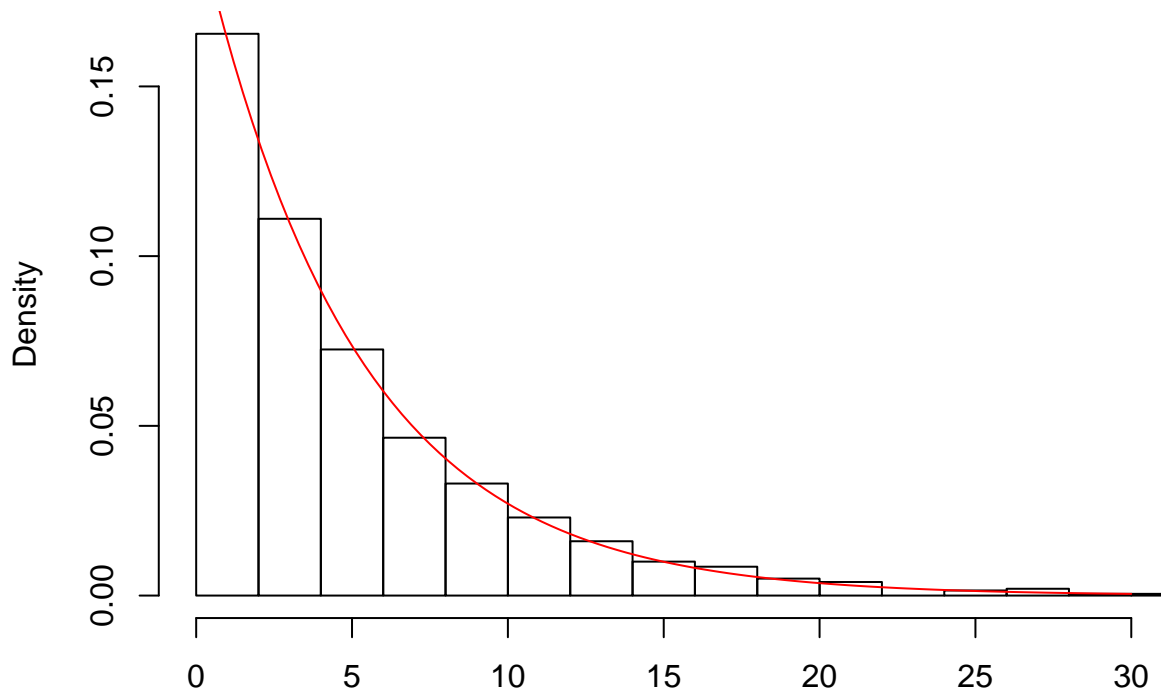
Exponential distribution density

The exponential distribution can be simulated in R with **rexp(n, lambda)**.
This code generates a sample n = 1000 random exponential, and plots the frequency with the **hist()** function.

```
randvar <- rexp(1000, 0.2) # generates 10000 random data points
hist(randvar)              # plots the histogram of frequencies
```

Its density plot can be overlayed with the **curve()** function, passing the **dexp()** density expression.
Hereby the resulting plot combining the histogram of the simulation and the theoretical distribution:

Exp distribution with lambda = 0.2

Mean and standard deviation

The mean and standard deviation of the exponential distribution equal to 1/lambda (here $1/0.2 = 5$); these are the theoretical values. We can calculate the sample mean and standard deviations from the generated data:

```r
mean(randvar)
## [1] 5.04053
sd(randvar)
## [1] 5.045832
```

The **Law of the Large Numbers** (LLN) says that the average of the samples limits to the population statistics.
In fact, thanks to the large n = 1000, the empirical values are quite close to the theoretical 5.

We can prove the LLN with simulations of sampled data. The following charts shows how the cumulative mean and standard deviations of the samples tends to the theoretical value of the population distribution, as n grows larger:
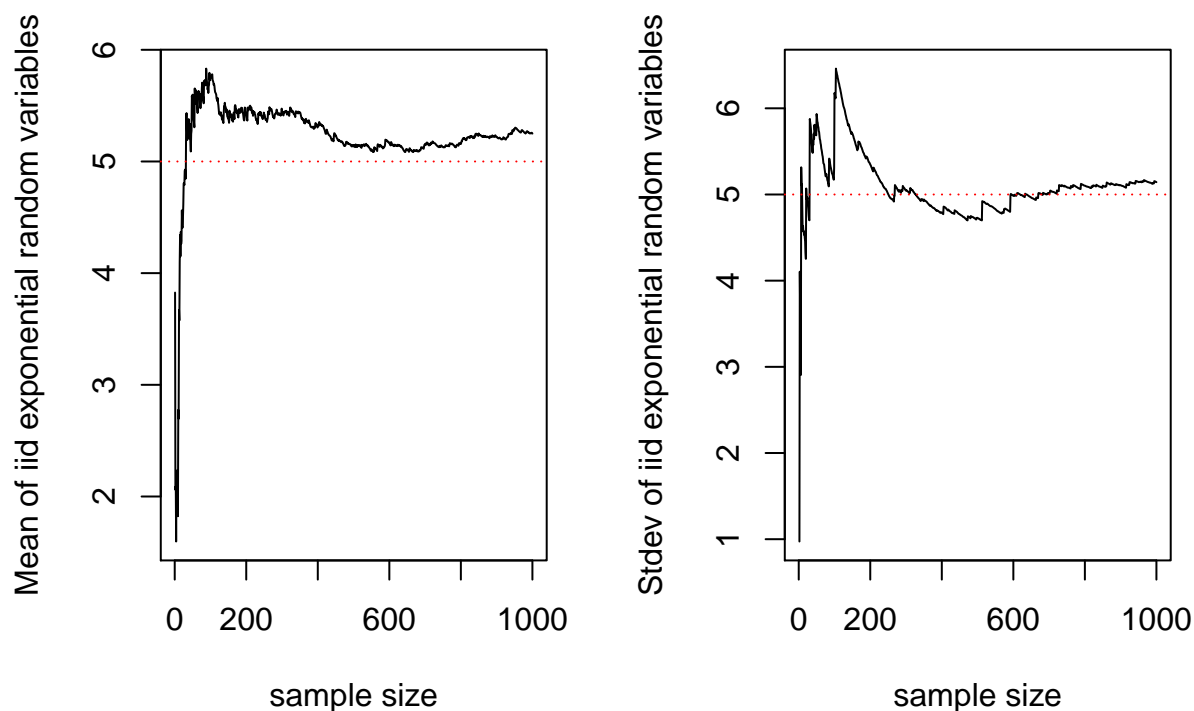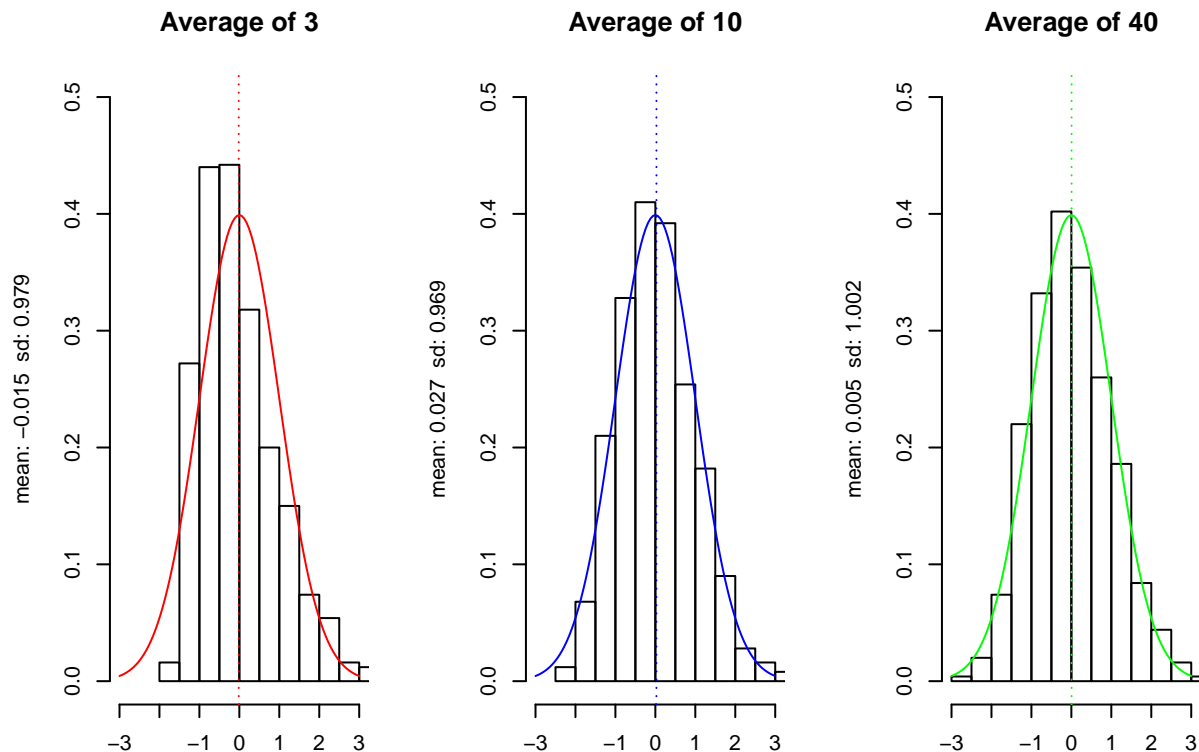
Distribution of averages

Indipendent samples from an exponential distribution will have the same distribution as the original population; this is how we plotted the histogram on the first page. But average of iid will distrubute approximately normal.

The **Central Limit Theorem** (CLT) says that the distribution of averages of iid normalized variables tends to a standard normal distribution, as the sample size increases.

We can average samples of 3, 10, and 40 exponentials and compare their frequency histograms with n = 1000 samples.

The results are scaled and centered on the mean, and compared with a standard normal distribution $N(0,1)$ (overlayed as a curve on the histograms).

Already with the average of three observations, the resulting distribution is not exponential anymore.

It appears clearly that the larger the sample that we average, the less skewed and the more normal the distribution of averages appears.

Their mean and standard deviations are very close to a standard normal N(0, 1), thanks to the LLN seen above.