

Application of the Central Limit Theorem on exponential distribution samples

Istvan Andras Horvath

November 9th, 2018

Synopsis

This analysis gives a simple example of the *Central Limit Theorem (CLT)* in work. First we do a simulation where a bunch of samples are drawn from an exponential distribution, then we take the mean of these samples, and investigate the distribution of these (i.e. the sampling distribution of sample means).

Simulation

Our simulation takes **1000 samples** of **40 observations** drawn from the **exponential distribution** with $\lambda = 0.2$. The samples are stored in a matrix, where each row is a sample, and then we take the mean of each sample and store it in a vector, which gives us the sampling distribution.

Note that the mean of the parent distribution is 5 and its variance is 25 ($1/\lambda$ and $1/\lambda^2$, respectively).

The first 10 observations from the first 6 samples of the total of 1000:

- **1:** 9.9400499, 3.2826425, 10.1411389, 9.5359762, 9.9790733, 10.021609, 3.3380234, 2.5568815, ...
- **2:** 1.852186, 6.3798372, 0.3644684, 3.7490035, 12.9094397, 1.0001554, 1.9420843, 7.2751347, ...
- **3:** 0.3626897, 0.690116, 1.6967716, 0.3439116, 1.7755715, 4.6599877, 2.6835379, 1.4357562, ...
- **4:** 2.0105746, 5.1834819, 11.0493176, 0.4793081, 0.0084416, 0.172083, 3.8642995, 2.3376573, ...
- **5:** 0.2795998, 4.8557262, 8.9952741, 5.1614842, 0.0402059, 0.045647, 2.6974713, 0.6728431, ...
- **6:** 3.079352, 1.0549501, 5.4862602, 0.6861427, 0.5439516, 13.2570887, 7.129205, 1.4701634, ...

The mean of the sampling distribution

An important conclusion of the CLT is that the mean of the sampling distribution of sample means approximates the mean of the parent distribution, or in other words the mean of the underlying population distribution. (Note: this is valid for other population parameters as well, not only for the mean.)

By calculation, the sampling distribution mean is 5.043053, while the parent distribution mean, as we've seen earlier, is 5. These values fall quite close to each other.

To illustrate this, we plot the parent distribution mean (in red) against the sampling distribution mean (in blue) in the figure below. These are all drawn onto the sample means, just to give an overview of those values as well.

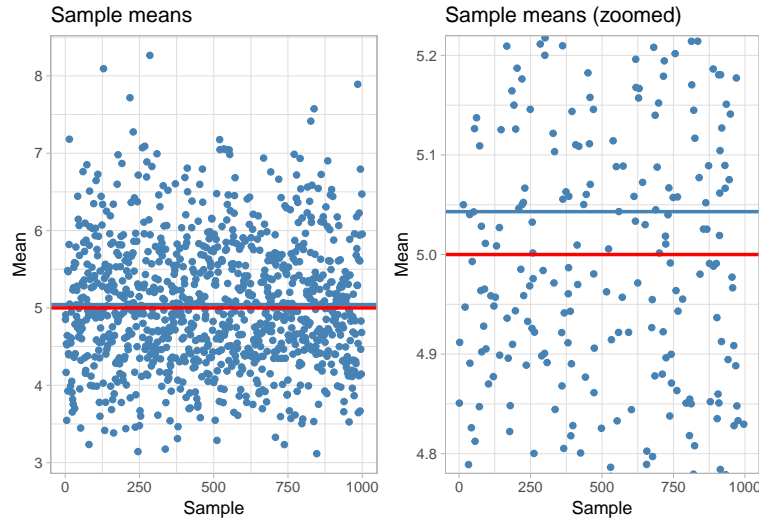


Figure 1: Population mean vs. sampling mean

The left part of the figure shows the closeness of the two means: the two horizontal lines almost overlap. The right part of the figure zooms in on the plot around the means, just for a better view.

The variability of the sample means

By definition, the theoretical variance of the sampling distribution is the variance of the parent population divided by the sample size (σ^2/n), which can be estimated by the variance of the samples divided by the sample size (S^2/n).

Based on these the theoretical sample variance is 0.625 while the estimated sample variance is 0.6787078. Similarly, the standard deviations are 0.7905694 and 0.8238373 for theoretical and estimated, respectively.

The estimations are quite close to the theoretical values.

Parent distribution vs. sampling distribution

Another important conclusion of the Central Limit Theorem is that the sampling distribution of a statistic is approximately normal. (More precisely, as the sample size n increases, the sampling distribution becomes more and more normal shaped.)

To illustrate this, the following figure shows two histograms side-by-side:

- *left plot*: the histogram of 1000 random values taken from the exponential distribution ($\lambda = 0.2$)
- *right plot*: the histogram of the means of the 1000 samples

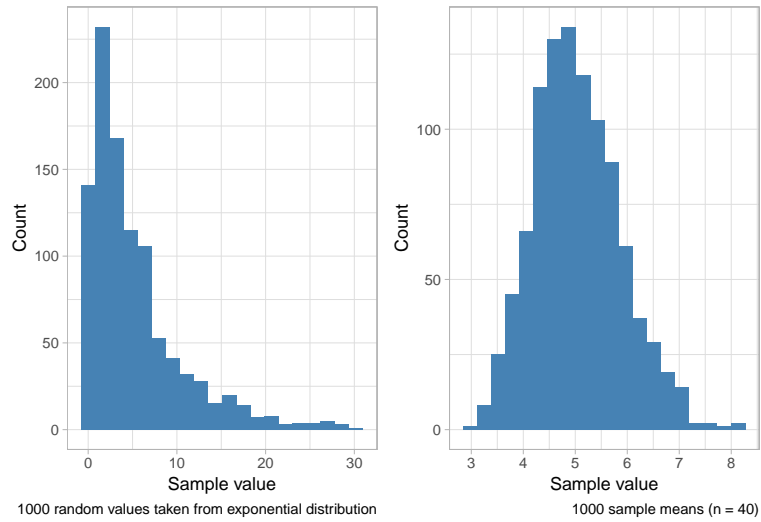


Figure 2: Parent distribution vs. sampling distribution

It is clearly visible that even though the observations taken directly from the parent distribution tend to distribute exponentially, the means of the samples taken from the same parent distribution tend to distribute normally: it has an almost symmetric bell shape around the mean.