[STAT-12] The central limit theorem

Miguel-Angel Canela
Associate Professor, IESE Business School

The central limit theorem

The **central limit theorem** is another grand name of Mathematical Statistics. It grants that, for big samples, the normal distribution gives a good approximation to the distribution of many interesting sample statistics, such as means or variances, which are proportional to a sum of independent and equally distributed terms. There are several versions of the theorem, which differ on the amount of assumptions made. Of course, the less assumptions you make, the longer the proof. Since we are not concerned with the technicalities, I state the theorem in a loose and general form, which could be applied to various particular cases. You can find in many textbooks proofs of some particular cases, which have preceded the general theorem in the development of Mathematical Statistics. For instance, the application of the central limit theorem to the binomial setting is an older result, the **Laplace-De Moivre formula**, found in many elementary textbooks.

Suppose that X_1, X_2, ..., X_n, ... is a sequence of independent equally distributed variables, with mean μ and variance σ^2. I denote by $\bar{X}_n$ the mean of the first n terms of this sequence. The central limit theorem states that the CDF of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ converges to the standard normal CDF as $n \to \infty$. This means, in practice, that we can use the normal as an approximation to calculate probabilities related to $\bar{X}_n$.

An equivalent version of the theorem uses the sum instead of the mean, replacing μ and σ by $n\mu$ and $n\sigma$, respectively. Comments about a particular sampling distribution being **asymptotically normal** usually refer to the approximation of that distribution by a normal. When such approximations are available, it is customary to specify whether the true distribution or the normal approximation is used. The terms exact and asymptotic are commonly used to this purpose.

¶ The convergence granted by this theorem is not the same as that of the law of large numbers. Here, we do not say that the variables converge to a limit, but that the distributions do.

Example 1. One of the great things of the central limit theorem is that there is no restriction on the type of distribution from which we sample, as long as it has moments of first and second order. In particular, it can be used for discrete distributions. To see how this works, look at Figure 1. For the outcome of a regular die, the distribution is uniform, with probability 1/6. But the distribution of the mean of two dice (left), is no longer uniform, but triangular.

What happens when we increase the number of dice averaged? That the number of possible values increase, so that a continuous approximation makes sense, and the probability distribution gets closer to the bell shape of the normal (right).

Approximation by the normal

Example 2. By virtue of the central limit theorem, the normal can be seen as the limit of many distributions. For instance, a $\mathcal{B}(n, \pi)$ variable is the sum of n independent Bernouilli variables. Then,

$$\mathcal{B}(n, \pi) \approx \mathcal{N}\big(n\pi, n\pi(1 - \pi)\big).$$

Figure 2 (left) illustrates this approximation. For $\pi = 0.1$ and $n = 100$, the mean is $n\pi = 10$

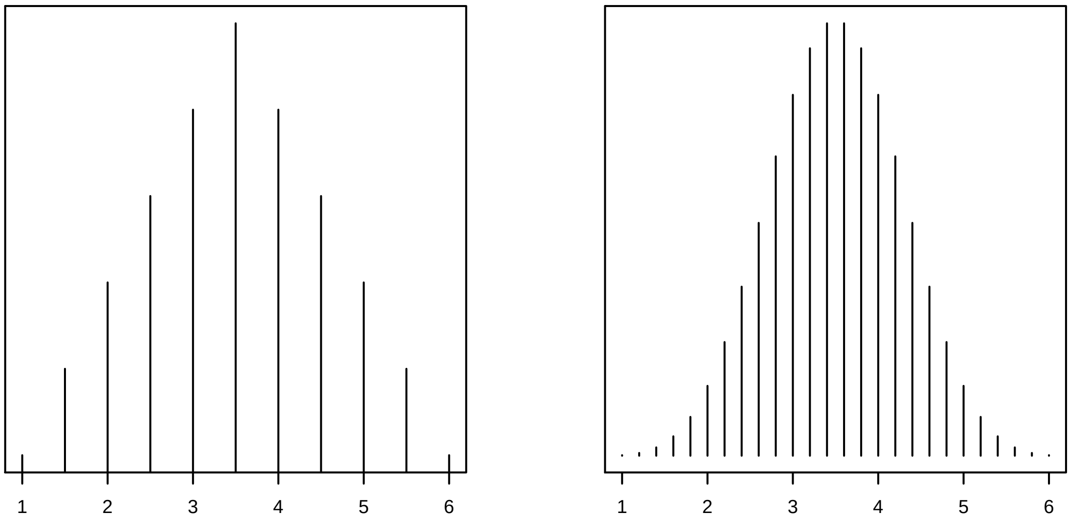

Figure 1. Distribution of the mean of two (left) and five (right) throws of a die

and the variance $n\pi(1-\pi) = 9$. In the figure, we see both the $\mathcal{B}(100, 0.1)$ probabilities and the $\mathcal{N}(10, 9)$ approximation. It is worth to remark, with respect to the approximation of the binomial, that the quality of the approximation depends, not only on n, but also on π, improving as π gets closer to 0.5. $\square$

Example 3. The Poisson distribution, with λ integer, provides another example. A Poisson variable with $\lambda = n$ can be seen as the sum of n independent Poisson variables with $\lambda = 1$. Then,

$$\mathcal{P}(n) \approx \mathcal{N}\big(n, n\big).$$

This is illustrated in Figure 2, which shows the $\mathcal{P}(50)$ probabilities together with the $\mathcal{N}(50, 50)$ approximation. $\square$

Example 4. Finally, the central limit theorem also applies to the χ^2 distributions. Indeed, a $\chi^2(n)$ variable is the sum of n independent $\chi^2(1)$ variables. So for a high n, the $\chi^2(n)$ distribution is asymptotically normal,

$$\chi^2(n) \approx \mathcal{N}(n, 2n).$$

Sample skewness and kurtosis

The sampling distributions of the sample skewness and kurtosis are also asymptotically normal and independent, with standard deviations $\sqrt{6/n}$ and $\sqrt{24/n}$, respectively. So, the magnitude of any of these statistics can be assessed by comparing the ratio of the actual value to the standard deviation with a critical value of the standard normal. The **Jarque-Bera (JB) statistic**, quite popular in Econometrics, is

$$JB = \frac{n}{6}\left(Sk^2 + \frac{K^2}{4}\right).$$

Under normality, JB is asymptotically $\chi^2(2)$ distributed.

Example 5. For the `amzn` data set, the 2013 daily returns have skewness $\mathrm{Sk} = 0.549$ and kurtosis $\mathrm{K} = 4.659$. With $n = 251$, we have a standard deviation of 0.155 of the sampling distribution of the skewness, and 0.309 for that of the kurtosis. So, the values of both statistics exceed, by

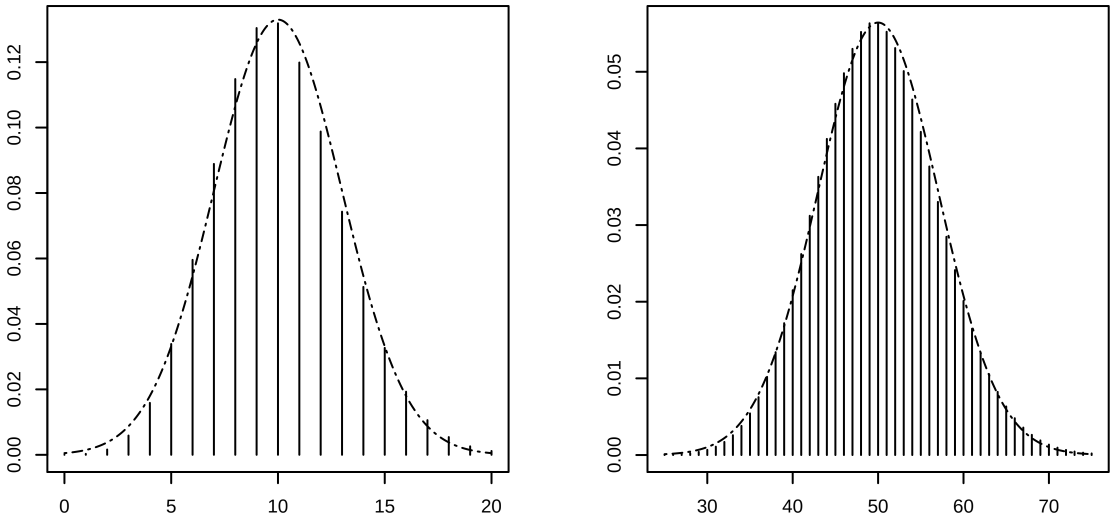

Figure 2. Normal approximation for the $\mathcal{B}(100, 0.1)$ (left) and $\mathcal{P}(50)$ (right)

large two standard deviations, and can be taken as evidence of a departure from the normal. The Jarque-Bera statistic is

$$JB = \frac{251}{6}\left(0.552^2 + \frac{4.720^2}{4}\right) = 239.6,$$

which is a very extreme value for a $\chi^2(2)$ distribution.

Homework

A. Simulate the sampling distribution of the mean of an exponential distribution, for sample sizes $n = 1$, $n = 5$, $n = 10$ and $n = 50$ and discuss the results.