[STAT-06] Sampling distributions

Miguel-Angel Canela
Associate Professor, IESE Business School

Sample mean

Let X_1, ..., X_n be a (statistical) sample from a distribution, that is, a collection of independent variables following that distribution. A function $G = g(X_1, \dots, X_n)$ is called a **statistic** (singular). The obvious example is the **sample mean**, defined as

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

Note that, here, $\bar{X}$ is not a number, but a random variable, with a probability distribution. To distinguish this distribution from the distribution from which we draw the sample, we call it a **sampling distribution**. In general, statistics are interesting when their sampling distributions have nice properties. I'll be more specific about this in the next lecture.

Let us assume that the distribution we are sampling from has mean μ and standard deviation σ. It is not difficult to prove that

$$\mathrm{E}\big[\bar{X}\big] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i] = \mu, \qquad \mathrm{var}\big[\bar{X}\big] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}[X_i] = \frac{\sigma^2}{n}\,.$$

The first formula means that, on the average, the sample mean is right as an approximation of the population mean μ. Also, since the variance is a measure of the variation about the expectation, the second formula tells us that the approximation improves as the sample size increases.

Note that the independence of the observations is needed for the formula of the variance but not for that of the mean. When sampling from a $\mathcal{N}(\mu, \sigma^2)$ distribution, we know something more on the sampling distribution of the mean, that $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$. If the distribution is not normal, we still have the central limit theorem, discussed in the next lecture.

Limit theorems

From the expression of the variance, we see that $\mathrm{var}\big[\bar{X}\big] \to 0$ as $n \to \infty$. Since the mean of $\bar{X}$ is equal to μ irrespective of the sample size, we can say that $\bar{X}$ converges to μ as $n \to \infty$. This statement, called the **law of large numbers**, is one of the great theorems of Mathematical Statistics.

Although the idea of the law of large numbers is clear enough, a comment on **limit theorems** is worth here. In Statistics textbooks, a lot effort is put on explaining the distinctions among the different types of convergence. Why? First, because the proofs of limit theorems can be more or less difficult depending on that. Second, because, although the definition of the limit of a sequence of numbers has nothing to hide, because numbers are simple things, but a random variable carries a lot on its back. What does converge, the values of the variables, the densities, parameters like means and variances?

There are different types of convergence, and developing them here will take more space than allowed. So, this discussion is quite short. The law of large numbers is easily proved if we formulate

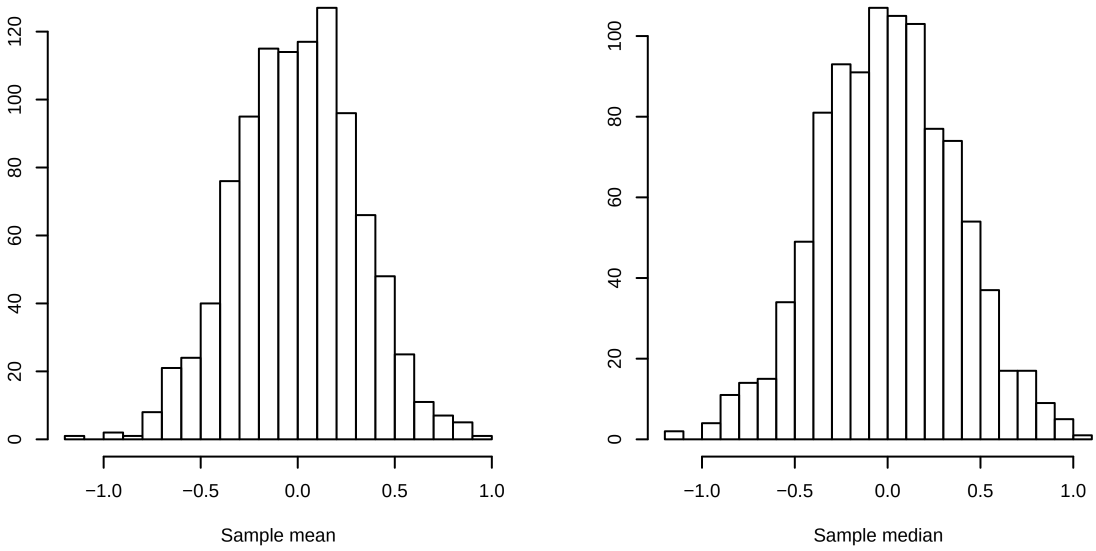

Figure 1. Sampling distributions of the mean and median

it in terms of **convergence in probability**. It can be stated as: for every number $\epsilon > 0$,

$$\lim_{n\to\infty} \mathrm{p}\big[|\bar{X} - \mu| > \epsilon\big] = 0.$$

Example 1. The simulation of their sampling distributions helps to understand the properties of statistics. In this example, I sample from the $\mathcal{N}(0,1)$ distribution, with size $n = 10$. I generate 1000 samples, saving the means, medians and variances. The resulting data set contains 1,000 observations of each of these three statistics.

The histograms of the sample mean and the sample median are shown in Figure 1. The means are –0.0039 and 0.0012, respectively, close to the zero population mean. The standard deviation of the sample mean is 0.3112, close to the theoretical value $1000^{-1/2} = 0.3162$. The standard deviation of the sample median is 0.3652, a bit higher. This agrees with the theory, and supports the preference for the mean.

Chi square distribution

Let $Z_1, \ldots, Z_n$ be independent $\mathcal{N}(0,1)$ variables, and define $X = Z_1^2 + \cdots + Z_n^2$. The distribution of X is one of the classics of Statistics, the **chi square distribution with n degrees of freedom**, in short $\chi^2(n)$. The χ^2 distribution is used in many tests in practical statistical analysis, specially with maximum likelihood estimation (not covered by this course).

It follows directly from the properties of the expectation that the mean and the variance of the $\chi^2(n)$ distribution are n and $2n$, respectively. We see in Figure 1 three χ^2 density curves. The formula of the PDF is not a friendly one,

$$f(x) = \frac{x^{n/2-1}\, e^{-x/2}}{2^{n/2}\,\Gamma(n/2)}\,, \quad x > 0,$$

but the calculations can be easily managed in the computer. The denominator comes from the normalization condition. The notation $\chi^2_\alpha(n)$ is consistent with the notation proposed for the standard normal in the preceding lecture (z_α).

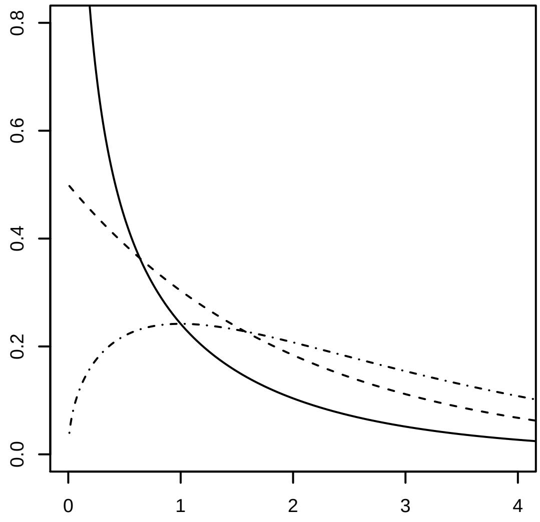

Figure 2. $\chi^2(1)$, $\chi^2(2)$ **and** $\chi^2(3)$ **density curves**

Γ denotes the Gamma function, which appears in many probability formulas. It is a positive, increasing function on $(0, +\infty)$, satisfying $\Gamma(n+1) = n!$ for any integer n. Although the Gamma function is defined by an intimidating integral formula, which I skip here, it is easily managed in the computer. Note that, since $\Gamma(1) = 1$, the $\chi^2(2)$ distribution is the same as the exponential distribution.

Sample variance

The **sample variance** is defined as

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \big(X_i - \bar{X}\big)^2,$$

The expectation of the sampling distribution is σ^2. Indeed, since all the terms of the sum on the right side of this formula have the same expectation (I assume here $\mu = 0$, to shorten the equations),

$$\begin{aligned} \mathrm{E}\big[S^2\big] &= \frac{n}{n-1}\,\mathrm{E}\big[\big(X_1 - \bar{X}\big)^2\big] = \frac{n}{n-1}\Big(\mathrm{E}\big[X_1^2\big] + \mathrm{E}\big[\bar{X}^2\big] - 2\,\mathrm{E}\big[X_1\bar{X}\big]\Big) \\ &= \frac{n}{n-1}\left(\sigma^2 + \frac{\sigma^2}{n} - \frac{2}{n}\sum_{i=1}^{n}\mathrm{E}\big[X_1X_i\big]\right) = \frac{n}{n-1}\left(\sigma^2 + \frac{\sigma^2}{n} - 2\,\frac{\sigma^2}{n}\right) = \sigma^2. \end{aligned}$$

This explains why the definition of S^2 with $n-1$ in the denominator is favored by most statisticians (this is why I'm giving you the detail of the argument). The formula for $\mathrm{var}[S^2]$ is more complex, involving the kurtosis. Nevertheless, under normality, the distribution of the sample variance can be related to the χ^2 model (the proof involves some matrix algebra). More specifically,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

In particular, the variance of the sample variance is

$$\mathrm{var}\big[S^2\big] = \frac{2\,\sigma^4}{n-1}\,.$$

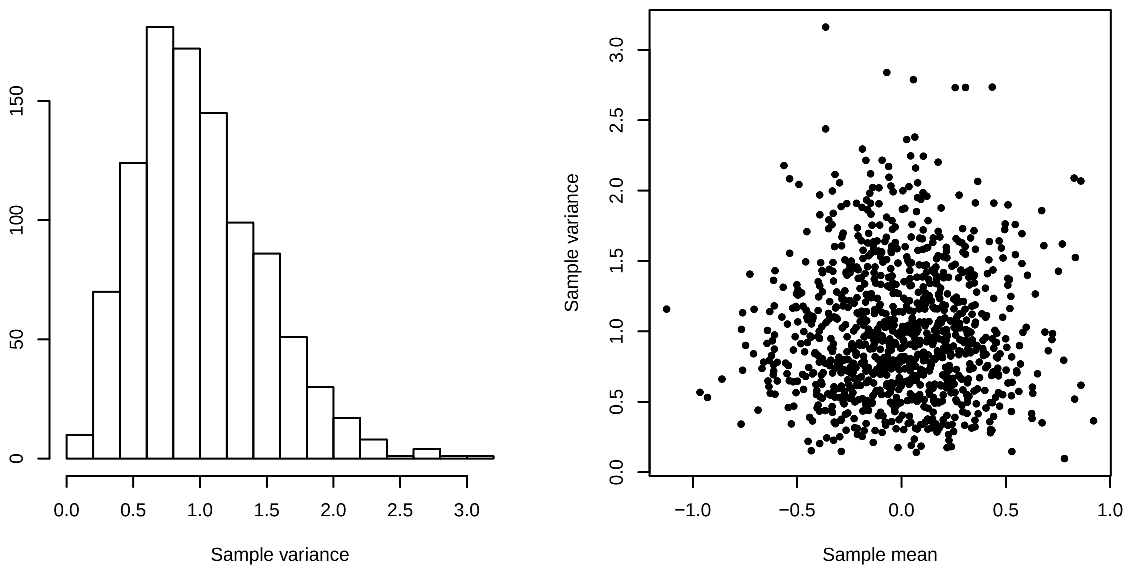

Figure 3. Distribution of the sample variance

Under a normality assumption, the mean and sample variance are independent. I omit the details, but you can find support for this assertion in the simulation below.

Example 1 (continuation). The left panel of Figure 3 is a histogram of the sample variance, consistent with the expected χ^2 profile. The mean is 1.0021 and the standard deviation 0.4732. Note that, according to the theory, the sample variance is distributed as one ninth of a $\chi^2(9)$ variable. The correlation between the sample mean and variance is 0.0167, also in agreement with the theory. This is illustrated by the right panel of the figure.

Sample covariance and correlation

Statistical analysis also involve the **sample covariance**,

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}\big(X_i - \bar{X}\big)\big(Y_i - \bar{Y}\big),$$

and the **sample correlation**

$$R = \frac{S_{XY}}{S_X\,S_Y} = \frac{\sum\big(X_i - \bar{X}\big)\big(Y_i - \bar{Y}\big)}{\Big[\sum\big(X_i - \bar{X}\big)^2\sum\big(Y_i - \bar{Y}\big)^2\Big]^{1/2}}\,.$$

The covariance is rarely tested, but the correlation is a key statistic. The sampling distribution of the correlation is a bit difficult. For the moment being, we will be satisfied with a simulation (exercise A).

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

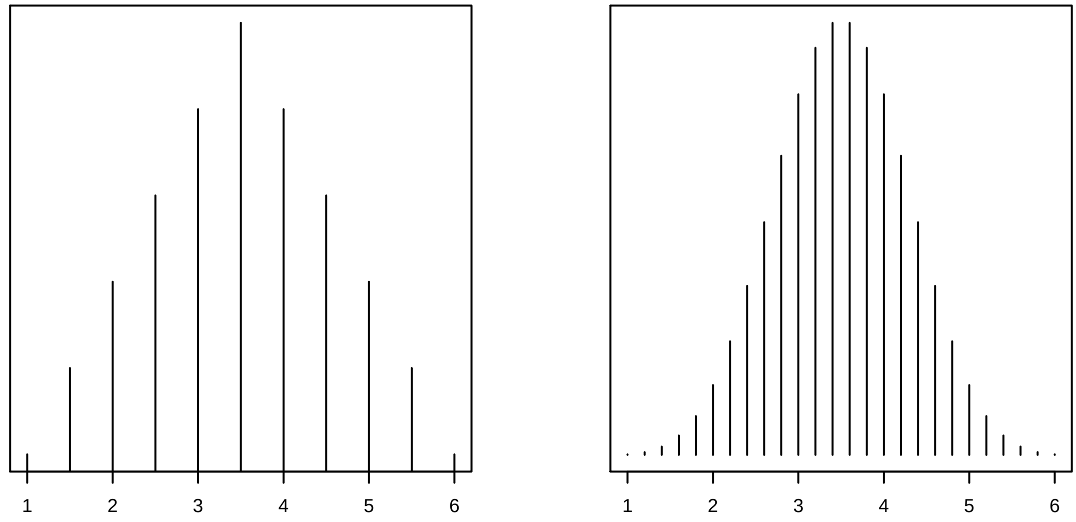

Figure 4. Distribution of the mean of two (left) and five (right) throws of a die

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 these notes 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 $X_1, X_2, \dots, X_n, \dots$ 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 when calculating 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 2. One of the great things of the central limit theorem is that there is no restriction on the type of distribution sampled, 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 4, which shows the distribution the sum of the outcomes of 2 and 5 dice. 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).

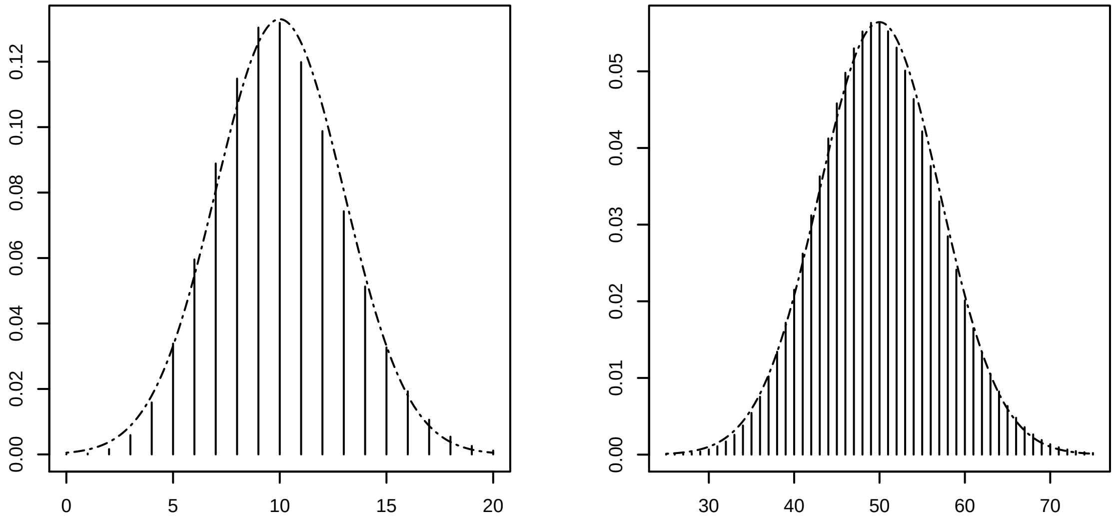

Figure 5. Normal approximation for the $\mathcal{B}(100, 0.1)$ (left) and $\mathcal{P}(50)$ (right)

Approximation by the normal

Example 3. 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 5 (left) illustrates this approximation. For $\pi = 0.1$ and $n = 100$, the mean is $n\pi = 10$ 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 4. 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 the right part of Figure 5, which shows the $\mathcal{P}(50)$ probabilities together with the $\mathcal{N}(50, 50)$ approximation. $\square$

Example 5. 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,

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

This is illustrated in Figure 6, for 5 and 25 degrees of freedom. The dashed line is the density curve of the normal approximation.

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

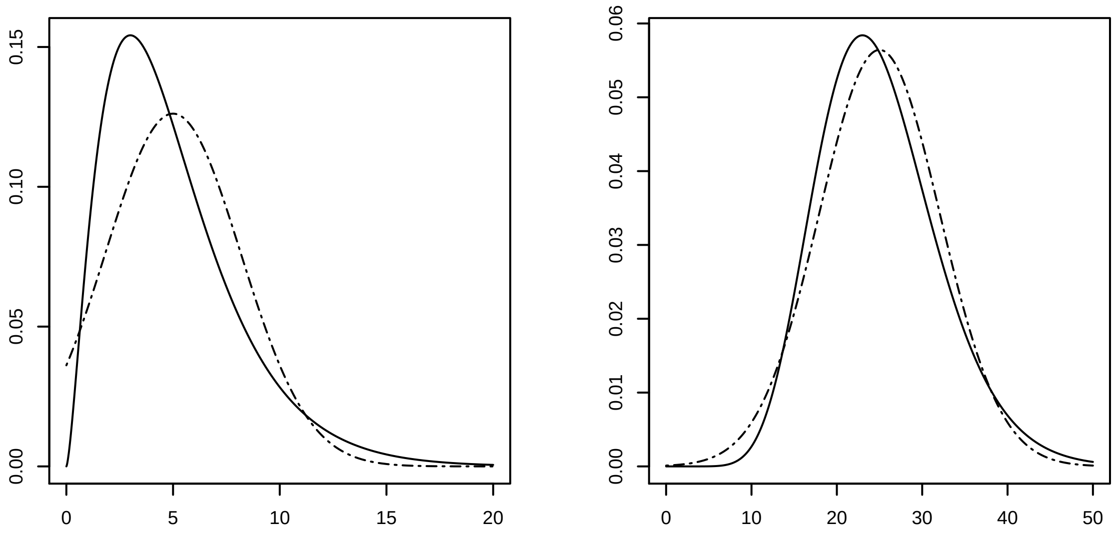

Figure 6. Normal approximation for the $\chi^2(5)$ and $\chi^2(25)$ distributions

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 6. The `amzn` data set contains the daily returns of Amazon stock price in 2013. The skewness is Sk $= 0.549$ and kurtosis 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 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 correlation for two independent exponential distributions with $\lambda = 1$.

B. 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.