# [STAT-08] Expectation in continuous distributions

**Miguel-Angel Canela**
**Associate Professor, IESE Business School**

### Expectation and variance

The definition of the expectation of a continuous distribution is the same as in the discrete case, but the sum is replaced by the integral,

$$\mathrm{E}[X] = \int_{-\infty}^{+\infty} x\, f(x)\, dx.$$

This definition is sound as far as the integral makes sense. As you may remember, improper integrals are sometimes divergent. Although you will never find them in your research, there are continuous distributions without mean. The variance and standard deviation are defined as in the discrete case. The properties are the same. The expectation is denoted by $\mu$ and the variance by $\sigma^2$. The **standardization** of a continuous variable,

$$Z = \frac{X - \mu}{\sigma}$$

produces a zero mean, unit variance variable, as in Descriptive Statistics.

**Example 1.** For the uniform distribution in the unit interval

$$\mu = \int_0^1 x\, dx = \frac{1}{2}\,,$$

$$\sigma^2 = \mathrm{E}\big[X^2\big] - \mu^2 = \int_0^1 x^2\, dx - \frac{1}{4} = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}\,.$$

For the general uniform distribution $\mathcal{U}(a, b)$, we get, using the properties of the expectation operator,

$$\mu = \frac{a + b}{2}\,, \qquad \sigma^2 = \frac{(b - a)^2}{12}\,. \ \square$$

**Example 2.** For the exponential distribution, integrating by parts and using L'H'ôpital rule,

$$\mu = \int_0^{+\infty} x\, e^{-x}\, dx = \Big[ - x\, e^{-x} - e^{-x} \Big]_{x=0}^{x=+\infty} = 1.$$

A similar, but longer, calculation gives

$$\sigma^2 = \mathrm{E}\big[X^2\big] - \mu^2 = \int_0^{+\infty} x^2\, e^{-x}\, dx - 1 = 1. \ \square$$

For the general exponential distribution $\mathcal{E}(\lambda)$, we get, because of the properties of the expectation,

$$\mu = \frac{1}{\lambda}\,, \qquad \sigma^2 = \frac{1}{\lambda^2}\,. \ \square$$

The median is the center of a probability distribution, in the sense that one half of the occurrences are expected to fall on each side of the median. Nevertheless, we favour means as central values, because of their better mathematical properties. In the uniform distribution, the mean and the median are equal, due to the symmetry of the PDF with respect to $x = 1/2$. This is no longer true for asymmetric distributions and, in extreme situations, makes the mean unreliable as a central value. We will meet this issue in many applications.

### Covariance and correlation

For two continuous variables, the **covariance** is defined as

$$\operatorname{cov}\big[X, Y\big] = \operatorname{E}\big[(X - \operatorname{E}[X])(Y - \operatorname{E}[Y])\big].$$

An alternative expression is $\operatorname{cov}\big[X, Y\big] = \operatorname{E}\big[XY\big] - \operatorname{E}[X]\operatorname{E}[Y]$. It is easy to check that the two formulas are the same thing. The correlation is defined as in Descriptive Statistics, as

$$\operatorname{cor}\big[X, Y\big] = \frac{\operatorname{cov}\big[X, Y\big]}{\operatorname{sd}[X]\operatorname{sd}[Y]}\,.$$

Covariance and correlation matrices are also defined as in Descriptive Statistics. They are positive definite. Using Greeks for the parameters of the distributions, the covariance would be $\sigma_{XY}$ and the correlation $\rho$.

*Theorem.* Let $X$ and $Y$ be independent variables. Then:

(i) $\operatorname{E}\big[XY\big] = \operatorname{E}[X]\operatorname{E}[Y]$.

(ii) $\operatorname{cor}\big[X, Y\big] = 0$.

This theorem is capital in Statistics. The second assertion is a direct consequence of the first one, which is easy to prove, in both the discrete and the continuous case. Why is this important? Because, though independence is relatively easy to examine and test for discrete variables, it is not so in the continuous case. On the other hand, uncorrelatedness is routinely tested in regression analyses.

So, two independent continuous variables are uncorrelated, and we test independence through correlation. Nevertheless, though confusion is common, independence and uncorrelatedness are not the same thing. The following example illustrates this.

**Example 3.** Let $X \sim \mathcal{U}(-1, 1)$ and $Y = X^2$. It is obvious that $X$ and $Y$ are not independent. Nevertheless,

$$\operatorname{E}[X] = \int_{-1}^{1} x\,dx = 0, \qquad \operatorname{E}[Y] = \int_{-1}^{1} x^2\,dx = \frac{2}{3}, \qquad \operatorname{E}\big[XY\big] = \int_{-1}^{1} x^3\,dx = 0. \;\square$$

### Skewness and kurtosis

The expectation of a power of a random variable is called a **moment**. Thus, the moment of order $k$ of $X$ is $\operatorname{E}[X^k]$. In particular, the moment of first order is the mean. Expectations of powers of $X - \mu$ are called **central moments**. The central moment of second order is the variance. Also of interest are the central moments of order 3 and 4, coming next.

We measure the lack of symmetry of a distribution through the third moment. The typical measure is the **skewness**, given by the standardized third moment. For $Z = (X - \mu)/\sigma$,

$$\operatorname{Sk}[X] = \operatorname{E}[Z^3] = \frac{\operatorname{E}[(X - \mu)^3]}{\sigma^3}\,.$$

Note that $\mathrm{Sk}[X] = 0$ when there is symmetry with respect to the mean. The **kurtosis** is

$$\mathrm{K}[X] = \mathrm{E}[Z^4] - 3 = \frac{\mathrm{E}[(X-\mu)^4]}{\sigma^4} - 3.$$

Replacing $\mu$, $\sigma$ and the expectation operator by their sample versions, we obtain sample versions of skewness and kurtosis. Since skewness and kurtosis are zero for a normal distribution, as we will see in the next lecture, sample skewness and kurtosis (those found in the data) are taken as measures of departure from normality.

¶ Some define the kurtosis without subtracting 3. Then, they call excess kurtosis what I call here kurtosis. Be careful with this, since the explanation about this detail is frequently missing.

### Homework

**A.** Calculate the skewness and the kurtosis of the uniform distribution.

**B.** The same for the exponential distribution.