[STAT-06] Expectation and variance in probability distributions

Miguel-Angel Canela
Associate Professor, IESE Business School

Expectation in discrete distributions

Let me consider a collection of n independent observations of a discrete variable X, assuming, to simplify the notation, that the range is finite, with k possible values. Every value x_i occurs n_i times, with a proportion $p_i = n_i/n$. Grouping repeated values, the mean is

$$\bar{x} = \frac{n_1\,x_1 + n_2\,x_2 + \ldots + n_k\,x_k}{n} = p_1\,x_1 + p_2\,x_2 + \ldots + p_k\,x_k.$$

So, the mean is the average of the values of X, weighted by their respective proportions. If the probability is understood as the limit of the proportion when the number of observations tends to infinity, it is natural to use this formula as a definition of the mean of a discrete distribution, with probabilities replacing proportions. More specifically, the expectation (or mean) of X is defined as

$$\mathrm{E}[X] = \sum_x x\,p\big[X = x\big].$$

Although we usually refer to the expectation of a variable, it would be more precise to refer to the expectation of a distribution, since it is the distribution that determines the expectation. The Greek letter μ is typically used to denote the mean of a distribution. Subscripts, as in μ_1, or μ_X, can be used to avoid confusion. The properties of the expectation are the same as in Descriptive Statistics. I don't repeat them here.

Variance

The **variance** of X, i.e. of the distribution of X, is defined as

$$\mathrm{var}[X] = \mathrm{E}\big[\big(X - \mathrm{E}[X]\big)^2\big].$$

It is easily seen that $\mathrm{var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$, which a faster formula for manual calculations. The **standard deviation** is the square root of the variance, $\mathrm{sd}[X] = \mathrm{var}[X]^{1/2}$. We use the Greek σ to denote the standard deviation of a distribution.

The properties of the variance of a discrete random variable are the same as in Descriptive Statistics. Note that, for distributions, the denominator is n, not $n-1$. This will be explained later in this course.

Example 1. Let X be the outcome of a regular die. The expected value is

$$\mathrm{E}[X] = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}\,.$$

To get the variance, I first calculate the expectation of X^2,

$$\mathrm{E}[X^2] = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}\,,$$

and, then,

$$\mathrm{var}[X] = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{175}{12} \,. \;\square$$

It is customary, sweeping away confusion, to distinguish through careful notation the mean of a probability distribution, the **population mean**, from the **sample mean**, which is the average of a set of observations. The same for the variance. The general rule is to use Greeks for the parameters of probability models and Latin characters for sample statistics. Thus, μ and $\bar{x}$ denote means (population and sample, respectively), σ^2 and s^2 variances, ρ and r correlations, etc.

Independent variables

When X and Y are statistically independent, we have

(i) $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$.

(ii) $\mathrm{var}\big[X + Y\big] = \mathrm{var}[X] + \mathrm{var}[Y]$.

These properties, which are easy to prove, will be discussed later, when we introduce the correlation of random variables.

Homework

A. Suppose that a certain gambler is equally like to win or to lose and that, when he/she wins, his/her fortune is doubled, but, when he/she loses, is cut in half. If the gambler begins playing with a fortune c, what is the expected value of his fortune after n independent plays of the game?