Basic elements of probability theory

Stefano Lovo*

April 22, 2019

Abstract

These note includes some basic formal tools in probability theory.

1 Random variable

A random variable is a variable whose realization depends on some random outcome. For example the number resulting when throwing a dice, is an integer number between 1 and 6, that is not known before the dice is thrown. The price of one share of General Electric in one year from now, could be any positive amount, and it is unknown today. Also, the number of rainy days in Paris during the next 6 months could be any natural number between 0 and 66. In economics, random variable are also used to describe variables that are known by some economic agents but not by others. Consider for example and individual Alan participating to an auction for some antique. Alan knows the maximum price P_A that he is willing to bid for the antique, however another bidder Bob does not know Alan's taste and hence does not know P_A. Thus P_A is a number for Alan (say 500 Euros), but it can be seen as a random variable for Bob.

In what follows, to denote a random variable we will use the notation $\tilde{x}$.

The following elements formally describe a random variable:

*HEC,Paris and GREGHEC, 78351 Jouy-en-Josas, France. lovo@hec.fr.

- The **domain** D indicating all the possible values that the random variable can take. For example the domain of the random variable describing the outcome of the dice is $D = \{1, 2, 3, 4, 5, 6\}$, whereas the domain of the random variable describing the price of one share of General Electric in one year from now is the the set of all non-negative real numbers, $\mathbf{R}^+$.

- An **algebra** defined over D. Loosely speaking, an algebra is the collections of all possible subset of the domain. For example the set $\{1, 3\}$ is a subset of the domain D of the the random variable describing the outcome of the dice, so it is an element of the algebra associated to that random variable.

- A probability $\Pr(\cdot)$ that associated to each element X of the algebra a number between 0 and 1. For any given subset X of the domain D, we denote with $\Pr(\tilde{x} \in X)$ the probability of X. It indicates how likely it is that the realization of the random variable $\tilde{x}$ is a number belonging to the set X. The probability of X is larger the more it is likely that $\tilde{x}$ realization will actually be in X. The probaility satisfy the following properties

  - The probability of any event X is a included between 0 and 1, formally $\Pr(\tilde{x} \in X) \in [0, 1]$,
  - If an event X is impossible, then $\Pr(\tilde{x} \in X) = 0$.
  - If an event X is certain, then $\Pr(\tilde{x} \in X) = 1$.
  - If two events X and Y are mutually exclusive, then the probability that either one of the event realizes is equal to the sum of the probability of X and the probability of Y. Formally, if

    $$X \cap Y = \emptyset \Rightarrow \Pr(\tilde{x} \in X \cup Y) = \Pr(\tilde{x} \in X) + \Pr(\tilde{x} \in Y).$$

  - The domain D includes all possible realizations of the random variable. Hence

    $$\Pr(\tilde{X} \in D) = 1.$$

  Take the dice example. Then $\Pr(\tilde{x} \in \{2\}) = 1/6$, because there are one chances out of six that the dice outcome is equal to 2, but $\Pr(\tilde{x} \in \{8\}) = 0$

because a dice only has 6 faces. Also because the dice outcome cannot be both 1 and 3, that is $\{1\} \cap \{3\} = \emptyset$, one has that the probability that the dice outcome is either 1 or 3 is

$$\Pr(\tilde{x} \in \{1,3\}) = \Pr(\tilde{x} \in \{1\}) + \Pr(\tilde{x} \in \{3\}) = 1/6 + 1/6 = 1/3.$$

1.1 Cumulative distribution function of a random variable

From the probability of a random variable one can deduce the **cumulative distribution function** (c.d.f. henceforth) of the random variable. Take any real number x, then the cumulative distribution function of the random variable answers the following question: what is the probability that the realization of $\tilde{x}$ is not larger than the number x? Thus, the cumulative distribution function of a random variable $\tilde{x}$ is a function F mapping any real number into a probability, namely

$$F(x) = \Pr(\tilde{x} \leq x)$$

Three remarks. First, because F is a probability, no matter what x is, $F(x)$ cannot be negative and cannot be larger than 1. Second for any two number x and y with x smaller then y, if the realization of $\tilde{x}$ is not larger than x, then it cannot be larger than y. Hence $\Pr(\tilde{x} \leq y) \geq \Pr(\tilde{x} \leq x)$, that is the c.d.f. F is a non-decreasing function. Second, the c.d.f. is defined on the set of real numbers $\mathbf{R}$, not only on the domain D of the random variable. Take the dice example. Even if the domain is $D = \{1, 2, 3, 4, 5, 6\}$ one can have $F(4.253)$, that is the probability that the outcome of the dice is not larger than 4.253. This is equal to to the probability that the dice outcome is either 1,2,3 or 4. So $F(4.253) = 2/3$.

When the c.d.f F is differentiable, then we say that the random variable $\tilde{x}$ is **continuously distributed** and denote with f the derivative of F, that is

$$f(x) := F'(x) \geq 0$$

where the inequality follows from the fact that F is non-decreasing. We call the function f the **density function** of the random variable $\tilde{x}$. Observe that

$$F(x) = \int_{-\infty}^{x} f(x)dx$$

The simplest example of continuous distribution is the **uniform distribution**. Loosely speaking, if $\tilde{x}$ is uniformly distributed on $[0,1]$ then its realzation must be a number included between 0 and 1, but no number in this interval is more likely to be realized than any other number in this interval. In this case we have

$$F(x) = 0 \text{ for } x \leq 0,$$

$$F(x) = x \text{ for } x \in [0,1],$$

$$F(x) = 1 \text{ for } x \geq 1.$$

1.2 Expectation, variance and standard deviation of a random variable

The **expectation** of a random variable $\tilde{x}$, that we will denote $E[\tilde{x}]$, can be interpreted as the weighted average of all its possible realizations, using as weight for each possible realization the probability with which the random variable equals that realization.

Take the dice example, then we have

$$E[\tilde{x}] = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = 3.5$$

If $\tilde{x}$ is continuously distributed with density f, then the expectation of $\tilde{x}$ is

$$E[\tilde{x}] = \int_{-\infty}^{+\infty} xf(x)dx$$

Interpretation: the expectation of a random variable x can be seen as the average of the realizations of $\tilde{x}$ over many independent random drawn according to the c.d.f. F. For example if you were to throw the dice n times and then take the average of the n outcomes, then this average is more and more likely to approach $E[\tilde{x}] = 3.5$ as n becomes larger and larger.

The **variance** of a random variable $\tilde{x}$ is the expectation of the random variable $\tilde{y}$ obtained by taking the square of the difference between the realization of $\tilde{x}$ and the expectation of $\tilde{x}$. Formally

$$Var(\tilde{x}) = E[(\tilde{x} - E[\tilde{x}])^2]$$

For a continuously distributed random variables we have that

$$Var(\tilde{x}) = \int_{-\infty}^{+\infty} (x - E[\tilde{x}])^2 f(x) dx$$

The **standard deviation** of a random variable $\tilde{x}$, that we will denote $\sigma(\tilde{x})$ is the square-root of its variance:

$$\sigma(\tilde{x}) = \sqrt{Var(\tilde{x})}$$

Interpretation: both the variance and the standard deviation are a measure of the dispersion of the realization of $\tilde{x}$ around its expectation $E[\tilde{x}]$. The larger these variables the bigger the average distance between the realization of n draws and the expectation of $\tilde{x}$. Thus the larger is the variance and the standard deviation of a random variable, less certain we are about the actual realization of the random variable. For example if you were to draw the same random variable n times and then take the average distance between each outcome of the draw and the expectation of the random variable, then this average distance is more and more likely to approach $\sigma[\tilde{x}]$ as n becomes larger and larger.

1.3 Conditional probability

Consider two events A and B. The probability of event A conditional on B, written $\Pr(A|B)$ is the probability that the event A will happen given that we are certain that event B happens. Take the dice example. Let A be the event 'the outcome of the dice is 2'. Let B be the event ' the outcome of the dice is not larger than 3' and let C be the event, 'the autcome of the dice is either 1 or 6'. Suppose someone throws the dice in the room next to yours. From your perspective $\Pr(A) = \Pr(\tilde{x} \in \{2\}) = 1/6$. However if after throwing the dice, the person in the other rooms tells you that the outcome is not larger then 3, then what is the probability that the outcome of the dice is precisely 2. That is, what is $\Pr(A|B)$?

By the Bayes' theorem we know that

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)\Pr(B|A)}{\Pr(B)} \tag{1}$$

In our example $\Pr(B) = \Pr(\tilde{x} \in \{1, 2, 3\}) = 1/6 + 1/6 + 1/6 = 1/2$. Because event A (the outcome is 2) implies event B (the outcome not larger than 3) we have that $\Pr(A \cap B) = \Pr(A)$ and hence $\Pr(A|B) = \frac{1/6}{1/2} = 1/3$. In fact if you know that the outcome of the dice was not larger then 3 there is one chance out of three that it is 2.

One can use Bayes' theorem to compute conditional cumulative distribution function. In particular we will denote with $F(x|B)$ the probability that the outcome of random variably $\tilde{x}$ is not larger than x, given that event B has realized.

For example consider a random variable $\tilde{x}$ continuously distributed with c.d.f. F. What $F(x|\tilde{x} < y)$? Applying Bayes' rule we have that

$$F(x|\tilde{x} < y) = \int_{-\infty}^{\min\{x,y\}} \frac{f(z)}{F(y)} dz = \min\left\{1, \frac{F(x)}{F(y)}\right\} \tag{2}$$

1.4 Independent random variables

We say that two random variables $\tilde{x}$ and $\tilde{y}$ are **independently distributed** if the realization of one does not affect the c.d.f. of the other. For example, the random variables representing the outcome of a dice and the one representing the temperature in Rome next January 1st are clearly independent. Knowing the the outcome of a dice is 2, tell us nothing about the probability that the temperature in Rome next January 1st will be above or below 10° C. Similarly knowing the temperature in Rome does not help in predicting the outcome of a dice.

More formally, take two random variables $\tilde{x}$ and $\tilde{y}$ with c.d.f. are F and G, respectively. The two random variable are independent if and only if for any two real numbers x and y one has

$$\Pr(\tilde{x} < x \text{ and } \tilde{y} < y) = F(x)F(y).$$

Also, when $\tilde{x}$ and $\tilde{y}$ are independent, then it must be that for any two real numbers x and y, one has

$$F(x|y) = F(x)$$

and

$$G(x|y) = G(x)$$

In other words, information about the realization of one of the random variable does not help to better predict the other random variable

In many situations it make sense to assume that two random variable are independently distribute, but they are distributed according to the same c.d.f. F. When this happens we say that the two random variable are **identically and independently distributed** (**i.i.d.** henceforth). An example of two i.i.d. random variables is the outcome of a first draw of a dice, and the outcome of a second draw of the same dice. The fact that the first outcome of the draw is 6, does not change the shape of the dice, and hence does not change the probability distribution of the outcome of the second draw. That is, given any number x the probability that the outcome of the first draw is less than x is identical to the probability that the outcome of the second draw is less than x.

1.5 Correlated random variables

Generally speaking if two random variables $\tilde{x}$ and $\tilde{y}$ are not independent, then the realization of one random variables affects the probability distribution of the other random variables. In particular for some y one and x one has that

$$Pr(\tilde{x} < x|\tilde{y} = y) = F(x|\tilde{y} = y) \neq F(x) = Pr(\tilde{x} < x).$$

There are different measure of correlations. One of which is the **correlation coefficient** defined as

$$\rho_{\tilde{x}\tilde{y}} = \frac{E[(\tilde{x} - E[\tilde{x}])(\tilde{y} - E[\tilde{y}])}{\sigma(\tilde{x})\sigma(\tilde{y})]}$$

The correlation is always between -1 and 1, it is $+1$ in the case of a perfect direct (increasing) linear relationship (correlation), -1 in the case of a perfect decreasing (inverse) linear relationship. In all other cases, $\rho_{\tilde{x}\tilde{y}}$ takes some value in the open interval $(-1, 1)$ indicating the degree of linear dependence between the variables.

Observe that correlation does not imply causality. For example, let $\tilde{x}$ be the random variable providing the size of the human population of a country, say Portugal, in 5 years from now. Let $\tilde{y}$ be the random variable providing the number of newborn babies in Portugal in 5 years from now, and let $\tilde{z}$ denote the number of people leaving in Portugal having taken the picture of a stork that year. Clearly there is a positive correlation between $\tilde{x}$ and $\tilde{y}$, but there also is a positive correlation between $\tilde{y}$ and $\tilde{z}$, but whereas $\tilde{x}$ causes $\tilde{y}$, the same cannot be said for $\tilde{z}$ and $\tilde{y}$.