

PROBABILITY: RANDOM VARIABLES

Often in a random experiment we are not so much interested in the details of the outcome, but rather in a numerical value determined by the outcome. These numerical quantities are called **random variables** and we will indicate them with capital letters: **X, Y, ...**

Example 1. A family has three children, therefore, there are the following eight possibilities

$$(m, m, m), (m, m, f), (m, f, m), (f, m, m), \\ (m, f, f), (f, m, f), (f, f, m), (f, f, f)$$

For example, (m, m, f) means that the firstborn is a boy, the second is a boy, and the youngest is a girl.

Let's assume that the **probability is uniform**, so each of the possible eight outcomes has probability **1/8**. We are interested in the number of girls in the family, *i.e.*, in **X** which may take one of the possible four values: 0, 1, 2, 3.

Let's determine the probability that X takes each of the four values:

$$\{\mathbf{X} = 0\} = \{(\mathbf{m}, \mathbf{m}, \mathbf{m})\}, \quad \mathbf{P}(\mathbf{X} = 0) = 1/8;$$

$$\{\mathbf{X} = 1\} = \{(\mathbf{m}, \mathbf{m}, \mathbf{f}), (\mathbf{m}, \mathbf{f}, \mathbf{m}), (\mathbf{f}, \mathbf{m}, \mathbf{m})\}, \quad \mathbf{P}(\mathbf{X} = 1) = 3/8;$$

$$\{\mathbf{X} = 2\} = \{(\mathbf{m}, \mathbf{f}, \mathbf{f}), (\mathbf{f}, \mathbf{f}, \mathbf{m}), (\mathbf{f}, \mathbf{m}, \mathbf{f})\}, \quad \mathbf{P}(\mathbf{X} = 2) = 3/8;$$

$$\{\mathbf{X} = 3\} = \{(\mathbf{f}, \mathbf{f}, \mathbf{f})\}, \quad \mathbf{P}(\mathbf{X} = 3) = 1/8.$$

Notice that $\mathbf{P(X = 0)} + \mathbf{P(X = 1)} + \mathbf{P(X = 2)} + \mathbf{P(X = 3)} = \mathbf{1}$.

If we define $\mathbf{P(x) = P(X = x)}$ we have

$$\sum_{x=0}^3 \mathbf{P(x)} = \mathbf{1}$$

according to a fundamental property of the probability.

We may also write

$$\sum_{i=1}^4 \mathbf{P(x_i)} = \mathbf{1}$$

where we assume that $x_1 = 0$, $x_2 = 1$, $x_3 = 3$ and $x_4 = 4$.

Example 2. A point is uniformly placed on the segment $[0, 2]$ which implies that the probability density $\rho(x)$ is constant in the interval and vanishes elsewhere.

According to fundamental property of probability we must have

$$\int_{-\infty}^{\infty} \rho(x) dx = 1$$

therefore:

$$\rho(x) = \frac{1}{2} \quad \text{if } x \in [0, 2]$$

and

$$\rho(x) = 0 \quad \text{if } x \in (-\infty, 0) \quad \text{or} \quad x \in (2, \infty).$$

The probability that **X** assume a value in the interval $[a, b]$ is

$$P(a \leq X \leq b) = \int_a^b \rho(x) dx$$

where a and b are real numbers (eventually, a may equal $-\infty$ and/or b may equal ∞).

For example, according to our probability density, we have

$$P(-1 \leq X \leq 3.4) = 1 \quad P(1 \leq X \leq 5.2) = 0.5$$

$$P(-1 \leq X \leq 1.5) = 0.75 \quad P(2 \leq X \leq 4) = 0$$

RANDOM VARIABLE: DEFINITIONS

Definizione. A random variable is a numerical mapping that assigns a real number X to each outcome of a random experiment.

In the first example, $X = x$, with $x = 0, 1, 2, 3$, is the outcome and $P(x)$ is its probability.

In the second example, $X \leq x$, for any real x , is the outcome and $\rho(x)$ is its probability density.

A random variable is called **discrete** if it may assume a **finite or countable number of values** (not necessarily integers), it is called **continuous** if its values are **uncountable** (real numbers).

Let us define the distribution function (or cumulative probability) $F(x)$ as $F(x) = P(X \leq x) = P(-\infty \leq X \leq x)$ where x can be any real number or ∞ ($F(\infty)$ always equals 1).

In the first example the variable is discrete:

$$F(2.4) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{7}{8}$$

$$F(1.2) = P(X = 0) + P(X = 1) = \frac{1}{2}$$

In the second example the variable is continuous:

$$F(-0.2) = P(-\infty \leq X \leq -0.2) = 0$$

$$F(0.8) = P(-\infty \leq X \leq 0.8) = 0.4$$

$$F(2.8) = P(-\infty \leq X \leq 2.8) = 1$$

DISCRETE RANDOM VARIABLES

Let us consider first the case of **discrete** random variables and assume that **X** may take the **n** values $x_1, \dots, x_n$ (in ascending order). For the following formulae we assume a finite **n**, but everything remains unchanged if it is infinite.

Definition The function

$$p(x_i) = P(X = x_i), \quad i = 1, 2, \dots, n$$

is called **probability** of the random variable **X**.

Notice that $0 \leq p(x_i) \leq 1$ for $i = 1, 2, \dots, n$ and that $\sum_{i=1}^n p(x_i) = 1$.

In the first example: $X \in \{0, 1, 2, 3\}$,

$$\begin{aligned} p(0) = P(X = 0) &= \frac{1}{8}, & p(1) = P(X = 1) &= \frac{3}{8} \\ p(2) = P(X = 2) &= \frac{3}{8}, & p(3) = P(X = 3) &= \frac{1}{8} \end{aligned}$$

Note that this discrete variable takes integer values, which is not always the case.

Example 3. Two dice are rolled, the variable Z , which is the sum of the outcomes, may assume the integer values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 with probabilities $1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36$. For example, $Z = 4$ corresponds to the 3 outputs (3, 1), (2, 2), (1, 3) and, therefore, the probability is $3/36$ (36 is the number of possible outputs).

DISTRIBUTION FUNCTION FOR DISCRETE VARIABLES

Definition. The function

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}$$

is called **distribution function** of the random variable **X**.

In general, the distribution function can be written as:

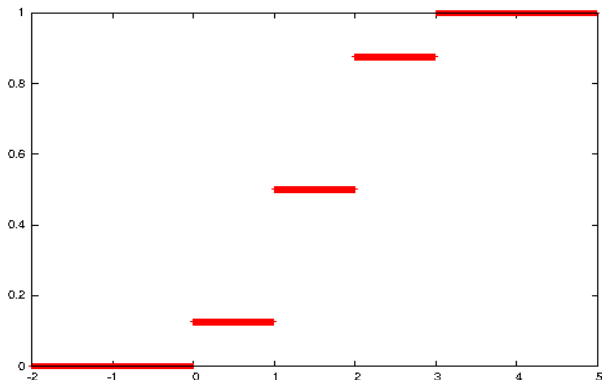
$$F(x) = \sum_{x_i \leq x} p(x_i)$$

where the sum goes over all $x_i \leq x$.

Given that the probabilities x_i aren't negative, **F(x)** is a non-decreasing function of **x**. Moreover:

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Below, the distribution function of the random variable in example 1 (number of daughters).



At the points of discontinuity, corresponding to the values that the discrete random variable may assume, the value of $F(x)$ is the largest one.

EXPECTED VALUE

Let X be a discrete random variable which may assume the n values $x_1, \dots, x_n$. One has $\sum_{i=1}^n P(X = x_i) = 1$. One can also use shorter notations as $p_i = P(x_i) = P(X = x_i)$.

Definition. The **expected value** (or mean or average of the discrete random variable X , indicated by $\mathbb{E}(\mathbf{X})$ is

$$\mathbb{E}(\mathbf{X}) = \sum_{i=1}^n x_i P(\mathbf{X} = x_i) = \sum_{i=1}^n x_i p(x_i)$$

Note that the definitions are the same as those for descriptive statistics for the sample mean, with relative frequencies replaced by probabilities.

In **example 1** the random variable X may assume the values 0, 1, 2, 3 with corresponding probabilities

$$p(0) = \frac{1}{8}, \quad p(1) = \frac{3}{8}, \quad p(2) = \frac{3}{8}, \quad p(3) = \frac{1}{8}$$

. Therefore:

$$\mathbb{E}(X) = \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5$$

.
Example 4: A dice is rolled, the random variable X may assume the values 1, 2, 3, 4, 5, 6 with corresponding probabilities

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}$$

. Therefore:

$$\mathbb{E}(X) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

Let g be a function, then $g(X)$ is also a random variable and his expected value is given by

$$\mathbb{E}(g(\mathbf{X})) = \sum_i g(\mathbf{x}_i) \mathbf{P}(\mathbf{X} = \mathbf{x}_i) = \sum_i g(\mathbf{x}_i) \mathbf{p}(\mathbf{x}_i).$$

Notice that the properties of linearity of the expected value are the same as those of the sample mean: if a and c are constant and one consider the random variable $Y = aX + c$, then

$$\mathbb{E}(a\mathbf{X}+c)=a\mathbb{E}(\mathbf{X})+c,$$

where the result is simply obtained by considering $aX+c=g(X)$.

Example 4: A dice is rolled, and before the roll, a euro is bet. The winnings are equal to one third of the number rolled. In symbols: $g(X) = X/3$. Therefore, the average win is

$$\mathbb{E}(g(\mathbf{X})) = \frac{1}{3}\mathbb{E}(\mathbf{X}) = \frac{3.5}{3} = 1,166666667$$

The game can be profitable!!!! You bet one euro against an expected winning value of one euro and 16 cents.

Consider again the **Example 4**. A dice is rolled and X is the outcome. Assume $g(\mathbf{X}) = \sqrt{\mathbf{X}}$, then

$$\mathbb{E}(g(\mathbf{X})) = \frac{1}{6}(\sqrt{1} + \sqrt{2} + \sqrt{3} + \sqrt{4} + \sqrt{5} + \sqrt{6}) = 1,805$$

THE BINOMIAL DISTRIBUTION

Suppose we have a population of n individuals. Each individual can answer yes or no to a given question. The probability of a positive answer is μ and the probability of a negative answer is $1 - \mu$. The total number X of positive answers can be any integer x such that $0 \leq x \leq n$. It can be shown (this is a possible exercise) that

$$p(x) = P(X=x) = \binom{n}{x} \mu^x (1 - \mu)^{n-x}.$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Indeed,

$$\mu^x(1 - \mu)^{n-x}$$

is the probability of a given sequence with x positive answers (and $n - x$ negative ones), For example, the sequence where the first x answers are positive and the remaining $n - x$ sequences are negative has this probability, but also the sequence where the first $n - x$ answer are negative and the remaining x are positive.

Moreover, the binomial coefficient

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

is the number of sequences of n answers with a number x of positive ones.

The number of sequences $\binom{n}{x}$ with x positive outcomes can be understood by imagining to draw from a box containing n balls where x are white (yes) and $n - x$ are black (no). You take the balls from the box one after the other and you align on a table. You can chose among n balls at the first move, among $n - 1$ at the second and so on.

The number of possible choice sequences is therefore $n!$. Nevertheless, all the $x!$ possible permutation of the x white balls and all the $(n - x)!$ possible permutation of the $n - x$ black balls leave the sequence unchanged.

This is enough to explain the binomial coefficient $\frac{n!}{x!(n-x)!}$.

A MANDATORY PROPERTY OF A PROBABILITY

An important use of the binomial coefficient is the following Newton formula:

$$(a + b)^n = \sum_{x=0}^n \binom{n}{x} a^x b^{n-x}.$$

If $a = \mu$ and $b = 1 - \mu$ the above formula becomes

$$\sum_{x=0}^n p(x) = \sum_{x=0}^n \binom{n}{x} \mu^x (1 - \mu)^{n-x} = 1$$

which is a mandatory property of a probability.

EXPECTED VALUE (OR AVERAGE OR MEAN)

The expected value of X can be easily computed. Given that

$$\begin{aligned}\sum_{x=0}^n x \binom{n}{x} a^x b^{n-x} &= a \frac{\partial}{\partial a} \sum_{x=0}^n \binom{n}{x} a^x b^{n-x} = \\ &= a \frac{\partial}{\partial a} (a + b)^n = na(a + b)^{n-1}\end{aligned}$$

by choosing $a = \mu$ and $b = 1 - \mu$ one obtains the expected value

$$\mathbb{E}(X) = \sum_{x=0}^n x p(x) = \sum_{x=0}^n x \binom{n}{x} \mu^x (1 - \mu)^{n-x} = n\mu.$$

A very important property of the expected value is the following: if X and Y are two random variables and Z is their sum, then

$$\mathbb{E}(Z) = \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Concerning the binomial choice, the single k individual may answer yes ($X_k = 1$) with probability μ or no ($X_k = 0$) with probability $1 - \mu$, which implies $\mathbb{E}(X_k) = \mu$ for all k . Moreover, the X variable of previous binomial model (total number of positive answers) can be written as $X = \sum_{k=1}^n X_k$, which finally implies

$$\mathbb{E}(X) = \sum_{k=1}^n \mathbb{E}(X_k) = n\mu.$$

This is an easier way to compute the average number of positive answers or, if you like, the expected outcome of a referendum.