# [STAT-03] Discrete probability distributions

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

### Random variables and probability distributions

Suppose that a collection of events and a probability are given on a sample space $S$, satisfying axioms of the preceding lecture. A **random variable** is a function $X : S \to \mathbb{R}$ such that, for every interval $I$ of the real line, the set of all sample units $s \in S$ such that $X(s) \in I$ is an event. The probability of this event is denoted by $\mathrm{p}[X \in I]$. If $I = (a, b)$, by $\mathrm{p}[a < X < b]$, if $I = (-\infty, a]$, by $\mathrm{p}[X \leq a]$. Et cetera.

To get the intuition of what this definition means, suppose that $S$ is a population of executives. Consider the following two examples:

- Gender. We define $X$ as 1, for female executives, and 0 for male executives. So, $\mathrm{p}[X = 1]$ would be the probability that an executive is female. This is a **discrete variable**.
- Income. We define $X$ as the income, in thousand USD per year. Here, $\mathrm{p}[500 < X < 1000]$ would be the probability of having an income between 500,000 and one million. This is a **continuous variable**.

¶ In the real world, everything is discrete. For instance, if the income is given in euros, it will be rounded to, at most, two decimal places. But, even if this is discrete, it is practical to take it as it were continuous.

**Example 1.** Let $X$ be the outcome of a regular die, with values 1, 2, 3, 4, 5, and 6. Some probabilities associated to $X$ are

$$\mathrm{p}\big[X = 2\big] = \frac{1}{6}\,, \qquad \mathrm{p}\big[1 < X < 5\big] = \frac{1}{2}\,, \qquad \mathrm{p}\big[X > 4\big] = \frac{1}{3}\,. \,\square$$

In Statistics textbooks, variable is synonym of random variable. Note that random variables, as defined here, take numeric values. Then, the so called **categorical variables**, whose values are taken in a finite set of categories, are not proper random variables. For instance, GENDER, with values MALE and FEMALE, is a categorical variable, which creates a partition of the sample space (in this case a human population) in two complementary events. Coding genders, e.g. as $X = 1$ for male and $X = 0$ for female, we get a proper random variable. These 0/1 variables, called **dummy variables**, or just dummies, are used in statistical analysis to include categorical variables in regression equations.

Roughly speaking, the **probability distribution** of $X$ is the specification of the probabilities of the events associated to $X$. How this is managed in practice depends on the nature of the variable. This lecture deals with the simplest case, that of a discrete variable. Continuous variables will be discussed in the next lecture.

¶ Following a textbook convention, I use upper case ($X$, $Y$, etc) for random variables, and low case ($x$, $y$, etc) for their values. So, expressions like $\mathrm{p}[X = x]$ make sense.

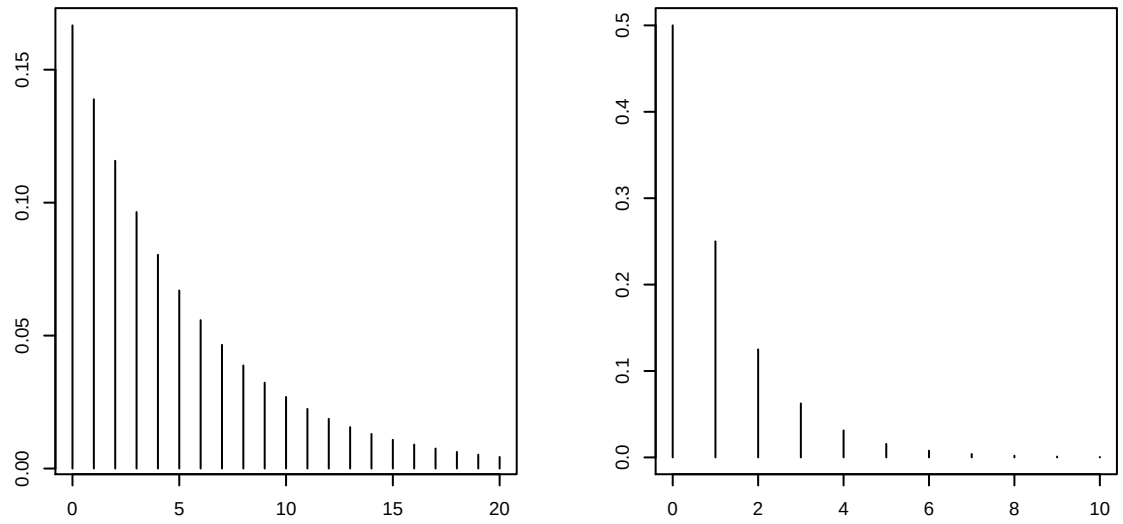


**Figure 1. Geometric distributions**

## Discrete probability distributions

The range of a **discrete variable** is a (finite or infinite) sequence of values $x_1$, $x_2$, .... The probability distribution is the sequence of probabilities $\pi_1 = \mathrm{p}\big[X = x_1\big]$, $\pi_2 = \mathrm{p}\big[X = x_2\big]$, .... In Example 1, all these probabilities are equal to 1/6. This is a **uniform discrete distribution**.

**Example 2.** The range of a discrete distribution can be infinite. An example is the **geometric distribution**. Suppose that we toss a die until we get six. Let $X$ be the number of tosses previous to the six. Here, $X = k$ occurs when we get a sequence of $k$ non-sixes followed by one six. So,

$$\mathrm{p}\big[X = k\big] = \left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^k, \quad k = 0, 1, 2, \ldots$$

Suppose now that we toss a coin until we get head. We get another geometric distribution, with

$$\mathrm{p}\big[X = k\big] = \left(\frac{1}{2}\right)^{k+1}, \quad k = 0, 1, 2, \ldots$$

Figure 1 is a graphical representation of these two geometric distributions. □

Note that, though we frequently confound them, a variable is not the same as its distribution, since different variables can have the same distribution. For instance, with a fair coin, coding head as 0 and tail as 1, we get a random variable and, coding them the other way around, another variable. Both variables have the same distribution.

## Expectation

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 (this is the frequentist approach), 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 frequently mention 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 $\mu$ is typically used to denote the mean of a distribution. Subscripts, as in $\mu_1$, or $\mu_X$, can be used to avoid confusion. The properties of the expectation are the same as in Descriptive Statistics (lecture STAT-01). 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 is a faster formula in manual calculations. The **standard deviation** is the square root of the variance, $\mathrm{sd}[X] = \mathrm{var}[X]^{1/2}$. We use the Greek $\sigma$ 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 (continuation).** Let $X$ be the outcome of a regular die. The expected value of $X$ 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,

$$\mathrm{var}[X] = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}\,. \;\square$$

It is customary, to sweep 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, $\mu$ and $\bar{x}$ denote means (population and sample, respectively), $\sigma^2$ and $s^2$ variances, $\rho$ and $r$ correlations, etc.

## Joint and marginal distributions

Let $X$ and $Y$ be discrete variables. The **joint probability distribution** specifies the joint probabilities $\mathrm{p}\big[X = x, Y = y\big]$. Here, the comma means "and", that is, intersection. The probability of an event associated to these variables is obtained summing the probabilities of the pairs $(x, y)$ included in this event.

The joint distribution of two discrete random variables is a **bivariate distribution**. The individual (univariate) distributions of $X$ and $Y$ are then called **marginal distributions**. The marginal probabilities can be derived from the joint probabilities by summing across the values of the other variable,

$$\mathrm{p}\big[X = x\big] = \sum_y \mathrm{p}\big[X = x, Y = y\big].$$

The opposite is not true, since there could be different joint distributions with the same marginals. To get the joint distribution, we need to specify, in addition to the marginals, the dependence structure. The extension of the definition of joint bivariate distribution to an arbitrary number of variables leads, in a natural way, to the concept of **multivariate distribution**.

**Example 3.** Let $D_1$ and $D_2$ be the outcomes of two dice, $X = D_1 + D_2$ and $Y = |D_1 - D_2|$. The joint probability distribution is given in Table 1, with $X$ in the rows and $Y$ in the columns. The blank cells correspond to null probabilities. The marginal probabilities are the row and column totals, placed on the right and bottom margins.

**TABLE 1. Joint probability distribution (Example 3)**

| | 0 | 1 | 2 | 3 | 4 | 5 | Total |
|---|---|---|---|---|---|---|---|
| 2 | 1/36 | | | | | | 1/36 |
| 3 | | 1/18 | | | | | 1/18 |
| 4 | 1/36 | | 1/18 | | | | 1/12 |
| 5 | | 1/18 | | 1/18 | | | 1/9 |
| 6 | 1/36 | | 1/18 | | 1/18 | | 5/36 |
| 7 | | 1/18 | | 1/18 | | 1/18 | 1/6 |
| 8 | 1/36 | | 1/18 | | 1/18 | | 5/36 |
| 9 | | 1/18 | | 1/18 | | | 1/9 |
| 10 | 1/36 | | 1/18 | | | | 1/12 |
| 11 | | 1/18 | | | | | 1/18 |
| 12 | 1/36 | | | | | | 1/36 |
| Total | 1/6 | 5/18 | 2/9 | 1/6 | 1/9 | 1/18 | 1 |

## Conditional distributions and statistical independence

Let $X$ and $Y$ be discrete random variables. The **conditional probability distribution** of $Y$, given $X = x$, is defined by the conditional probabilities

$$\mathrm{p}\big[Y = y \,|\, X = x\big] = \frac{\mathrm{p}\big[X = x, Y = y\big]}{\mathrm{p}\big[X = x\big]}\,.$$

We say that $X$ and $Y$ are **statistically independent** when every event related to $X$ is statistically independent of every event related to $Y$. This is the same as the joint distribution being the product of the marginal distributions, that is, as the product formula

$$\mathrm{p}\big[Y = y, X = x\big] = \mathrm{p}\big[X = x\big]\,\mathrm{p}\big[Y = Y\big].$$

We can extend the definition of independence to a set of more than two discrete variables, as we did with events. We have independence when the product formula is valid for any subset. We can build an example showing that three variables can be *pairwise* independent but not independent, using the same idea as in the case of three events (see lecture STAT-02).

When $X$ and $Y$ are statistically independent, the following formulas are valid:

(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?

**B.** In a lottery, $k$ numbers are selected from the $N$ numbers 1, 2, ..., $N$. Find the expected value of the sum $S_k$ of these numbers.