Chapter 2

Random Variables

This chapter is dedicated to the presentation of one-dimensional random variables. After an example illustrating this concept, we will consider the axiomatic definition: a random variable is a function that assigns a real number to each element ω of a fundamental set. We then study discrete and continuous random variables. The first thing to clarify is that throughout the rest of the document, we will use the abbreviation "r.v." for "random variable," both singular and plural.

2.1 Introductory Example

Consider the random experiment of rolling a fair die twice. The fundamental set is $\Omega = \{(i, j), i, j = 1, 2, ..., 6\}$, and the 36 pairs are equally probable. We associate with each pair the sum of the two numbers, i.e., $(i, j) \mapsto i+j$. We thus define a function from the fundamental set Ω to the set of numbers $\{2, 3, 4, ..., 12\}$. In this case, we say that the function X: "the sum of the two numbers rolled" is a random variable defined on the fundamental set Ω and taking values in the set $E = \{2, 3, 4, ..., 12\}$.

2.2 Definitions, Vocabulary

Let us begin with the following fundamental definition:

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A **random variable** defined on Ω and taking values in $(E, \mathcal{B})$ is any function X from Ω to E such that:

$$\forall B \in \mathcal{B}, \quad X^{-1}(B) = \{\omega \in \Omega, X(\omega) \in B\} \in \mathcal{F}$$

Such a function is said to be measurable. In practice, we rarely verify the measurability. When $E = \mathbb{R}$, we obtain the following definition:

Definition 1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A **real random variable** defined on Ω is any mapping X from Ω to $\mathbb{R}$ satisfying:*

$$\forall x \in \mathbb{R}, \quad X^{-1}\left(]-\infty, x]\right) = \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}.$$

* Due to the structure of the Borel σ-algebra $\mathbb{B}(\mathbb{R})$ (the smallest σ-algebra containing all intervals), the definition implies that a random variable X always satisfies: $\forall B \in \mathbb{B}(\mathbb{R})$, $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{F}$.

* The following notation is commonly used: if $B \in \mathbb{B}(\mathbb{R})$, then the set $X^{-1}(B)$ is abbreviated as $\{X \in B\}$, i.e.,

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} = \{X \in B\}.$$

Similarly, the set $X^{-1}\left(]-\infty, x]\right)$ is denoted as $\{X \leq x\}$, i.e.,

$$X^{-1}\left(]-\infty, x]\right) = \{\omega \in \Omega \mid X(\omega) \leq x\} = \{X \leq x\}.$$

2.2.1 Law of a Random Variable

In the introductory example, if we want to evaluate the probability associated with each number in the set E, we need to evaluate the probability associated with each event $X^{-1}(\{i\})$, $i = 2, 3, \ldots, 12$, which should be stated as: the probability that the outcome of the random variable X is i. Thus, to each possible outcome of the mapping X, we associate a probability, and all these probabilities together form the probability distribution of the random variable X. We now introduce the probability distribution of a random variable in general.

Theorem 1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let X be a random variable defined on Ω with values in $(E, \mathcal{B})$, where $\mathcal{B}$ is a σ-algebra on E. The mapping P_X defined by:*

$$\begin{aligned} P_X \quad &: \quad \mathcal{B} \longrightarrow [0,1] \\ &: \quad B \mapsto P_X(B) = \mathbb{P}(X^{-1}(B)) \end{aligned}$$

is a probability measure on $(E, \mathcal{B})$.

Proof. * First, we have $X^{-1}(E) = \Omega$, and thus: $P_X(E) = \mathbb{P}(X^{-1}(E)) = \mathbb{P}(\Omega) = 1$.

* If $(B_n)_{n\in\mathbb{N}}$ is a sequence of pairwise disjoint elements of $\mathcal{B}$, then $(X^{-1}(B_n))_{n\in\mathbb{N}}$ is a sequence of pairwise disjoint elements of $\mathcal{F}$. Since $X^{-1}\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \bigcup_{n\in\mathbb{N}} X^{-1}(B_n)$, and $\mathbb{P}$ is a probability measure, we have:

$$P_X\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{n\in\mathbb{N}} B_n\right)\right) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} X^{-1}(B_n)\right) = \sum_{n\in\mathbb{N}} \mathbb{P}(X^{-1}(B_n)) = \sum_{n\in\mathbb{N}} P_X(B_n).$$

This shows that P_X is a probability measure on $(E, \mathcal{B})$. □

Definition 2. *The probability measure P_X given by the above theorem is called the probability distribution of the random variable X.*

Now, the question arises: how to describe the distribution of a random variable? This depends directly on the space $(E, \mathcal{B})$, and in this case, there are two important spaces: the case where $E = \mathbb{R}$, in which case the random variable X is called **continuous**, and the case where E is at most countable (finite or countable), leading to the following definition:

Definition 3. *A real random variable X is called **discrete** if the set $X(\Omega)$ is at most countable (finite or countable): there exists a sequence of distinct real numbers $(x_i)_{i\in\mathbb{N}}$ such that $X(\Omega) = \{x_i \mid i \in \mathbb{N}\}$.*

Probability Distribution of a Continuous Random Variable

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let X be a random variable defined on Ω with values in $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$, and let P_X be its probability distribution. As we have seen, determining the distribution P_X is nothing other than determining a probability measure on $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$, and according to what we saw in the previous chapter, every probability measure on $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ is characterized by its cumulative distribution function F. This justifies the following definition:

Definition 4. *Let X be a real random variable. The cumulative distribution function of X, denoted F (or F_X if necessary), is the following function:*

$$\begin{aligned} F \quad &: \quad \mathbb{R} \longrightarrow [0,1] \\ &: \quad x \mapsto F(x) = P_X\left(]-\infty, x]\right) = \mathbb{P}(X \leq x). \end{aligned}$$

In this case, the probability distribution P_X of the random variable X is entirely determined by its cumulative distribution function F.

The terminology "cumulative distribution function of X" is somewhat improper because, in fact, F depends on the probability measure P_X and not directly on X. One of the interests of the cumulative distribution function is that it allows expressing $\mathbb{P}(X \in I)$ for any real interval $I \subseteq \mathbb{R}$. This is the subject of the following result:

Proposition 2. *Let X be a real random variable, and let F be its cumulative distribution function. Then:*

1. *F is increasing.*
2. $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to+\infty} F(x) = 1$.
3. *F is right-continuous at every point x, i.e.,* $\lim_{h\searrow 0} F(x+h) = F(x)$.
4. *If a and b are two real numbers such that $a < b$, we have (denoting $F(x-0) = \lim_{h\nearrow 0} F(x-h)$):*

(a). $\mathbb{P}(X = b) = P_X(\{b\}) = F(b) - F(b-0)$.
(b). $\mathbb{P}(a < X \leq b) = P_X(]a,b]) = F(b) - F(a)$.
(c). $\mathbb{P}(a \leq X \leq b) = P_X([a,b]) = F(b) - F(a-0)$.
(d). $\mathbb{P}(a < X < b) = P_X(]a,b[) = F(b-0) - F(a)$.
(e). $\mathbb{P}(a \leq X < b) = P_X([a,b[) = F(b-0) - F(a-0)$.
(f). $\mathbb{P}(X < b) = P_X(]-\infty,b[) = F(b-0)$.
(g). $\mathbb{P}(X \geq b) = P_X([b,+\infty[) = 1 - F(b-0)$.
(h). $\mathbb{P}(X > b) = P_X(]b,+\infty[) = 1 - F(b)$.

A consequence of this result is that F is continuous at a point x if and only if $P_X(\{x\}) = 0$. Note also that since F is increasing, its points of discontinuity form an at most countable set. Also, if F is continuous, we obtain $(b) = (c) = (d) = (e) = F(b) - F(a)$, and $(g) = (h) = 1 - F(b)$.

Example 1. *Let X be a real random variable with cumulative distribution function F given by (where $[x]$ denotes the integer part of x):*

$$F(x) = \begin{cases} \frac{1}{5}e^{x-1}, & \text{if } x < 1, \\ \frac{1}{4}[x], & \text{if } 1 \leq x < 3, \\ 1 - \frac{1}{2x}, & \text{otherwise.} \end{cases}$$

We easily obtain the following probabilities: $P(X = 0) = 0$, $P(X = 2) = F(2) - F(2-0) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$, and, noting that $F(3) = \frac{5}{6}$, $F(1) = \frac{1}{4}$, and $F(1-0) = \frac{1}{5}$: $P(1 < X \leq 3) = F(3) - F(1) = \frac{7}{12}$, $P(1 \leq X \leq 3) = F(3) - F(1-0) = \frac{19}{30}$.

We now come to a very profound result. It shows, in fact, that the distribution of a real random variable is entirely characterized by its cumulative distribution function.

Theorem 3. *The distribution P_X of a real random variable X is entirely determined by its cumulative distribution function F.*

More precisely, if $F : \mathbb{R} \longrightarrow [0,1]$ is an increasing, right-continuous function satisfying $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to+\infty} F(x) = 1$, *there exists a unique probability measure P on $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ such that:*

$$\forall x \in \mathbb{R}, \quad F(x) = P\left(]-\infty, x]\right).$$

We will now focus on two particular types of real random variables: **discrete** random variables on one hand, and **absolutely continuous** random variables on the other.

2.3 Discrete Random Variables

As we have seen in the definition, a real random variable X is called **discrete** if the set $X(\Omega)$ is at most countable, i.e., there exists a sequence of distinct real numbers $(x_i)_{i\in\mathbb{N}}$ such that $X(\Omega) = \{x_i \mid i \in \mathbb{N}\}$. If we denote for $i \in \mathbb{N}$, $p_i = \mathbb{P}(X = x_i)$, we must have $\sum_{i\in\mathbb{N}} p_i = 1$. Some of the p_i may be zero.

Example 2. *Imagine that we roll a die several times in a row:*

** If X represents the result of the first roll, then: $X(\Omega) = \{1,2,3,4,5,6\}$ and $p_1 = p_2 = \cdots = p_6 = \frac{1}{6}$.*

** If X now represents the first roll for which we obtain 1, then: $X(\Omega) = \mathbb{N}^*$ and for all $k \in \mathbb{N}^*$: $p_k = \mathbb{P}(X = k) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}$.*

We will see that the distribution of a discrete real random variable is relatively simple to characterize. First, for a real number x, determine the set $\{X \leq x\}$. Since $X(\Omega) = \{x_i \mid i \in \mathbb{N}\}$, we have $\{X \leq x\} = \bigcup_{i\in\mathbb{N}, x_i\leq x} \{X = x_i\}$.

From this set equality, we deduce two things: first, X is measurable if and only if the sets $\{X = x_i\}$ belong to $\mathcal{F}$ for all $i \in \mathbb{N}$; and second, using the σ-additivity of the probability measure $\mathbb{P}$, we obtain:

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}\left(\bigcup_{i\in\mathbb{N}, x_i \leq x} \{X = x_i\}\right) = \sum_{i\in\mathbb{N}, x_i \leq x} \mathbb{P}(\{X = x_i\}) = \sum_{i\in\mathbb{N}, x_i \leq x} p_i.$$

We see then that the cumulative distribution function of X—and consequently its distribution—is entirely determined by the real numbers $p_i = \mathbb{P}(\{X = x_i\})$. We thus obtain the following result:

Theorem 4. *Let X be a discrete real random variable taking the values $(x_i)_{i\in\mathbb{N}}$. The distribution of X is completely determined by the values*

$$p_i = \mathbb{P}(\{X = x_i\}) = P_X(\{x_i\}).$$

More precisely, for any Borel set B,

$$\mathbb{P}(\{X \in B\}) = \sum_{i\in\mathbb{N}, x_i \in B} p_i.$$

For a discrete variable, what does the graph of the function F look like? We have already said, with the notations of the previous theorem, that $F(x) = \sum_{i\in\mathbb{N}, x_i \leq x} p_i$. If we assume that the x_i are ordered in increasing order, we have for $x < x_0$, $F(x) = 0$, then for $x \in [x_0, x_1[$, $F(x) = p_0$, then on $[x_1, x_2[$, $F(x) = p_0 + p_1$, and so on. F is thus a piecewise constant function.

Example 3. *Let X be a random variable taking the values $-2, 1, 2$ with respective probabilities $\mathbb{P}(\{X = -2\}) = \frac{1}{3}$, $\mathbb{P}(\{X = 1\}) = \frac{1}{2}$, $\mathbb{P}(\{X = 2\}) = \frac{1}{6}$. The cumulative distribution function of X is given by:*

$$F(x) = \begin{cases} 0, & \text{if } x < -2, \\ \frac{1}{3}, & \text{if } x \in [-2, 1[, \\ \frac{5}{6}, & \text{if } x \in [1, 2[, \\ 1, & \text{otherwise.} \end{cases}$$

2.4 Absolutely Continuous Random Variables

In this section, we consider the case of absolutely continuous real random variables; let's start with a definition:

Definition 5. *Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ be a function. f is a probability density function if:*
1. *f is positive, i.e., $\forall x \in \mathbb{R} : f(x) \geq 0$.*
2. *$\int_{\mathbb{R}} f(x)\, dx = 1$.*

Example 4. *The function $f(x) = \frac{1}{4} e^{-\frac{|x|}{2}}$ is a probability density function.*

Definition 6. *Let X be a real random variable with cumulative distribution function F. We say that X is absolutely continuous if there exists a probability density function f such that:*

$$\forall x \in \mathbb{R}, \quad F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\, dt.$$

In this case, we say that X has density f, or that f is the probability density function of the random variable X.

If X is absolutely continuous with density f, then for any interval Any union of intervals (and even any Borel set) B,

$$\mathbb{P}(X \in B) = \int_B f(t)\, dt$$

The question that now arises is: how can we recognize whether a real random variable X has a density? Here is a result that clarifies the relationship between the cumulative distribution function and the probability density function.

Let X be a random variable with cumulative distribution function F.
1. If X has a density, then F is continuous, and thus $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$.
2. If F is differentiable (of class C^1), then X has a density f defined by $f(x) = F'(x)$.

2.5 Characteristics of a Random Variable

A random variable is entirely determined either by its cumulative distribution function, its density, or its probability function. In statistics, we generally focus on certain characteristic values that do not fully describe the random variable but are often very important. The most important characteristic values are the mathematical expectation and the variance.

2.5.1 Mathematical Expectation

The mathematical expectation, denoted $E(X)$, of a random variable X, also called the mean, if it exists.

1. In the case of a discrete random variable X, whose possible values are $x_i, i \in \mathbb{N}$, the expectation of X is defined by the expression

$$E(X) = \sum_{i\in\mathbb{N}} x_i \mathbb{P}(X = x_i) = \sum_{i\in\mathbb{N}} x_i p_i, \text{ if it exists.}$$

In concrete terms, the expectation of X is the weighted average of the values that X can take, with the weights being the probabilities of these values.

2. In the case of an absolutely continuous random variable X, whose probability density is f, the expectation of X is defined by:

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx, \text{ if it exists.}$$

Example 5. *We roll a fair die once, and let X be the random variable representing the number obtained. Then X takes the values $1, 2, 3, 4, 5, 6$ with equal probability $p_i = \frac{1}{6}$ for $i = 1, 2, 3, 4, 5, 6$. Thus, $E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5$.*

Example 6. *Let X be a real random variable with probability density f defined by*

$$f(x) = \begin{cases} e^{-x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}.$$

Then:

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx = \int_{-\infty}^{0} 0\, dx + \int_{0}^{+\infty} x e^{-x} dx = 1$$

Remark 1. *– The expectation is not always defined. Consider the case where X is a random variable that can take an infinite number of distinct values, $E(X) = \sum_{i=1}^{+\infty} x_i p_i$. When the sum makes sense and does not depend on the order of the terms, and only in this case, we consider that X has a mathematical expectation.*

– The expectation of a random variable does not necessarily belong to the possible values of X. Thus, the mathematical expectation of the values taken by a fair die is 3.5, while the die can only show integer values. $E(X)$ is a number around which the possible values of the random variable X are distributed.

A random variable X is said to be centered if $E(X) = 0$.

2.5.2 Expectation of a Function of a Random Variable

In general, we can define the expectation of a function of a random variable X, leading to the following definition:

Let X be a real random variable defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and Φ: $\mathbb{R} \to \mathbb{R}$ a function. We define the random variable $Y = \Phi(X)$ by:

$$Y : \Omega \to \mathbb{R}, \quad \omega \mapsto \Phi(X(\omega))$$

And if Φ is integrable, we have:

1. In the case of a discrete random variable X, whose possible values are $x_i, i \in \mathbb{N}$, the expectation of Y is defined by:

$$E(Y) = E(\Phi(X)) = \sum_{i\in\mathbb{N}} \Phi(x_i)\,\mathbb{P}(X = x_i) = \sum_{i\in\mathbb{N}} \Phi(x_i)\,p_i, \text{ if it exists.}$$

2. In the case of an absolutely continuous random variable X, whose probability density is f, the expectation of Y is defined by:

$$E(Y) = E(\Phi(X)) = \int_{-\infty}^{+\infty} \Phi(x)\,f(x)\,dx, \text{ if it exists.}$$

2.5.3 Properties of Mathematical Expectation

1. The mathematical expectation of a constant is the constant itself. That is, if $X = c$ ($c \in \mathbb{R}$), then $E(X) = c$.
2. If X is a random variable and $\alpha \in \mathbb{R}$, then: $E(\alpha X) = \alpha E(X)$.
3. If X and Y are two random variables, then: $E(X + Y) = E(X) + E(Y)$.
4. Using points 2 and 3, we obtain: If X and Y are two random variables, and $\alpha, \beta \in \mathbb{R}$, then $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$. That is, the mathematical expectation is a linear operator.
5. More generally, if $(X_i)_{i=1,2,\ldots,n}$ is a sequence of random variables and $(\alpha_i)_{i=1,2,\ldots,n}$ is a sequence of real numbers, then $E\left(\sum_{i=1}^{n} \alpha_i X_i\right) = \sum_{i=1}^{n} \alpha_i E(X_i)$.
6. If X is a random variable with expectation $E(X)$, then $Y = X - E(X)$ is a centered random variable ($E(Y) = 0$), meaning any random variable can be centered.
7. If X is a positive random variable, i.e., $P(X \geq 0) = 1$, then $E(X) \geq 0$.

2.5.4 Moments of a Random Variable

The moment of order r of the random variable X is the expectation $E(X^r)$, defined by:

1. In the case of a discrete random variable X, whose possible values are $x_i, i \in \mathbb{N}$:

$$E(X^r) = \sum_{i\in\mathbb{N}} x_i^r \mathbb{P}(X = x_i) = \sum_{i\in\mathbb{N}} x_i^r p_i, \text{ if it exists.}$$

2. In the case of an absolutely continuous random variable X, whose probability density is f:

$$E(X^r) = \int_{-\infty}^{+\infty} x^r f(x)\,dx, \text{ if it exists.}$$

We note that:

* The expectation $E(X)$ of a random variable X is nothing other than its moment of order 1.

* For $r = 0$, we obtain $E\left(X^0\right) = E(1) = 1$.

The moment of order r centered around the value $a \in \mathbb{R}$ of the random variable X is the expectation $E((X-a)^r)$, defined by:

1. In the case of a discrete random variable X, whose possible values are $x_i, i \in \mathbb{N}$:

$$E((X-a)^r) = \sum_{i\in\mathbb{N}} (x_i - a)^r \,\mathbb{P}(X = x_i) = \sum_{i\in\mathbb{N}} (x_i - a)^r p_i, \text{ if it exists.}$$

2. In the case of an absolutely continuous random variable X, whose probability density is f:

$$E((X-a)^r) = \int_{-\infty}^{+\infty} (x-a)^r f(x)\,dx, \text{ if it exists.}$$

2.5.5 Variance and Standard Deviation of a Random Variable

If X is a real random variable, the first information we seek is the mean value, $E(X)$. Next, we are interested in the dispersion of X around this mean value: this is the notion of variance.

Let X be a real random variable such that $E\left(X^2\right)$ exists. Then the variance of X, denoted $V(X)$ or $Var(X)$, is the moment of X of order 2 centered around $E(X)$. That is,

$$V(X) = E\left((X - E(X))^2\right) = E\left(X^2\right) - (E(X))^2$$

The standard deviation of the random variable X, denoted σ or σ_X, is the positive square root of its variance. That is:

$$\sigma_X = \sqrt{V(X)}$$

Example 7. *We roll a fair die once, and let X be the random variable representing the number obtained. Then X takes the values $1, 2, 3, 4, 5, 6$ with equal probability $p_i = \frac{1}{6}$ for $i = 1, 2, 3, 4, 5, 6$. Thus, $E(X) = 3.5$. To calculate the variance, we first calculate $E\left(X^2\right)$:*

$$E\left(X^2\right) = \sum_{i=1}^{6} x_i^2 p_i = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \cdots + 6^2 \cdot \frac{1}{6} = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6} = 15.17$$

$$V(X) = E\left(X^2\right) - (E(X))^2 = 15.17 - (3.5)^2 = 15.17 - 12.25 = 2.92$$

and $\sigma_X = \sqrt{V(X)} = \sqrt{2.92} = 1.7$.

2.5.6 Properties of the Variance of a Random Variable

1. The variance and standard deviation are positive values, i.e., $V(X) \geq 0$ and $\sigma_X \geq 0$.
2. The variance is zero if and only if the random variable is constant. That is, $X = c$ $(c \in \mathbb{R}) \Leftrightarrow V(X) = 0$.
3. If X is a random variable and $\alpha, \beta \in \mathbb{R}$, then: $V(\alpha X + \beta) = \alpha^2 V(X)$, and $\sigma_{\alpha X + \beta} = |\alpha| \sigma_X$.

Let X be a real random variable. X is said to be **standardized** if $V(X) = 1$.

Let X be a real random variable with expectation $E(X)$ and standard deviation σ_X. The **centered standardized** random variable associated with X, denoted X^*, is defined by:

$$X^* = \frac{X - E(X)}{\sigma_X}$$

It can be easily verified that $E(X^*) = 0$ and $V(X^*) = 1$.