I. GENERAL DEFINITION

Let us begin by introducing the natural environment to deal with random phenomena, relying on the assiomatic formulation due to Kolmogorov. The notions introduced in this chapter can be found in many excellent textbooks about basic Probability theory; we will thus limit ourselves to introduce the the fundamental notions needed to start our presentation.

The first ingredient we need is provided by the following

Definizione 1 *A* **probability space** *is a triplet of the form:*

$$(\Omega, \mathcal{F}, P) \tag{1}$$

where:

1. Ω *is a non–empty set, called* **sample space***;*
2. $\mathcal{F}$ *is a* σ*-field of subsets of* Ω*, called the* **set of events***;*
3. P *is a measure* $\mathcal{F}$ *satisfying the condition* $P(\Omega) = 1$*, called* **probability measure**.

The elements of Ω are interpreted as all the possible outcomes of an experiment modelled by the probability space $(\Omega, \mathcal{F}, P)$. Once observed the outcome of the given experiment, we know whether some events have happened or not: the collection of such events is $\mathcal{F}$. The probability measure describes how likely is the outcoming of the events. The mathematical requirement of $\mathcal{F}$ being a σ-field means, by definition, that $\mathcal{F}$ is closed under countable unions and intersections, and under complementations; morevover, Ω itself and the empty set $\emptyset$ belongs to $\mathcal{F}$ by definition. The measure P is σ-additive, i.e., for any countable family of events $\{A_1, \dots, A_n, \dots\} \subset \mathcal{F}$ such that $A_i \cap A_j = \emptyset$ if $i \neq j$:

$$P\left(\bigcup_{j=1}^{+\infty} A_j\right) = \sum_{j=1}^{+\infty} P(A_j) \tag{2}$$

Nota 2 *In general, the sample space can be any non-empty set, finite, infinite countable or uncountable. Whenever* Ω *is finite or countably infinite, the* σ*-field is always taken to be the whole power set of* Ω*,* $\mathcal{F} = \mathcal{P}(\Omega)$*, containing all the subsets of* Ω*. Whenever* Ω *is uncountable, the power set is in general too large, giving rise to pathological situations in which a probability measure cannot be defined (the reader may remember that this is the case in Lebesgue measure theory). In such cases one has to restrict the* σ*-field. Whenever* Ω *is a topological space, we will always choose the* **Borel** σ**-field**, $\mathcal{B}(\Omega)$*, which is the smallest* σ*-field containing all the open subsets of* Ω*.*

The second basic ingredient is the definition of random variable.

Definizione 3 *Let $(E, \mathcal{E})$ be a measurable space, that is a non-empty set E together with a σ-field of subsets $\mathcal{E}$. A* **random variable** *is a function $X : \Omega \to E$ which is* **measurable***, that is:*

$$\forall B \in \mathcal{E}, \quad \{\omega \in \Omega \,:\, X(\omega) \in B\} \in \mathcal{F} \tag{3}$$

When we want to make explicit the σ-fields, we will use the transparent notation:

$$X : (\Omega, \mathcal{F}, P) \to (E, \mathcal{E}) \tag{4}$$

Nota 4 *We will use simple notation of the form $\{X \in B\}$ instead of $\{\omega \in \Omega \,:\, X(\omega) \in B\}$, or, in case of real valued random variables, $\{X \leq x\}$ instead of $\{\omega \in \Omega \,:\, X(\omega) \leq x\}$ and so on.*

Random variables are thus the function recognizing the structure of measurable spaces, just like continous functions between topological spaces and linear applications between vector spaces.

The map:

$$\mathcal{E} \ni B \rightsquigarrow \mu(B) \stackrel{def}{=} P(X \in B) \tag{5}$$

is a probability measure on $\mathcal{E}$ and is called the **law** or **distribution** of the random variable X. The notation $X \sim \mu$ is commonly use to denote the fact that μ is the law of X. The measurable space $(E, \mathcal{E})$, equipped with the measure μ, becomes a probability space $(E, \mathcal{E}, \mu)$.

Nota 5 *We note that the identity map on $(\Omega, \mathcal{F})$:*

$$\Omega \ni \omega \rightsquigarrow id_\Omega(\omega) = \omega \tag{6}$$

is naturally a random variable and its law is precisely P. This simple observation allows to conclude that, whenever a probability measure P is defined on a measurable set, there always exist a random variable taking values in the given set whose law is P. This could appear trivial at first sight, but it is useful: we will always work directly with laws, forgetting the explicit definition of the random variables.

We stress that the notion of **measurability** is very important: from the point of view of the interpretation, the fact that a random variable X is measurable means that, once observed the outcome of an experiment modelled by $(\Omega, \mathcal{F}, P)$, the value of X is known. This will turn out to be a key point in future chapters, when our knowledge will depend on time.

Besides measurability, an extremely important notion is that of **independence**, translating in mathematical language the intuitive idea the term suggests.

Definizione 6 *A collection $\{\mathcal{F}_j\}_{j \in \mathcal{J}}$ (not necessarily finite) of sub-σ-fields of $\mathcal{F}$ are said to be* **independent** *if, for any finite subset $\mathcal{I} \subset \mathcal{J}$ the following equality holds:*

$$P\left(\bigcap_{i \in \mathcal{I}} A_i\right) = \prod_{i \in \mathcal{I}} P(A_i) \tag{7}$$

for any choice of events $A_i \in \mathcal{F}_i$*.*

A collection $\{A_j\}_{j\in\mathcal{J}}$ *of events belonging to* $\mathcal{F}$ *are said to be* **independent** *if the sub-*σ*-fields* $\left\{\mathcal{F}_{A_j}\right\}_{j\in\mathcal{J}}$*,* $\mathcal{F}_{A_j} = \{\emptyset, A_j, A_j^C, \Omega\}$*, are independent.*

A collection $\{X_j\}_{j\in\mathcal{J}}$ *of random variables,* $X_j : (\Omega, \mathcal{F}, P) \to (E_j, \mathcal{E}_j)$*, are said to be* **independent** *if the* **generated** *sub-*σ*-fields* $\{\sigma(X_j)\}_{j\in\mathcal{J}}$*,* $\sigma(X_j)$ *being, by definition, the smallest* σ*-field containing all the events* $\{X_j \in B\}$ *for all* $B \in \mathcal{E}_j$*, are independent.*

II. FIRST EXAMPLES: BINOMIAL LAW, POISSON LAW, AND GEOMETRIC LAW

Now that we have introduced the most important ingredients, we are ready to build up our first examples probability spaces and random variables. We consider an experiment consisting in tossing a (non necessarily balanced) coin n times and counting the number of heads obtained. How can we describe such situation in the language of probability theory? It is quite natural to build up a probability space $(\Omega, \mathcal{F}, P)$ in the following way; let's choose:

$$\Omega = \{\{0,1\}^n\} \tag{8}$$

This means that the possible outcomes have the form $\Omega \ni \omega = (\omega_1, \dots, \omega_n)$ where, using a simple convention, $\omega_i = 0$ if at the i-th toss we get tail and $\omega_i = 1$ if we get head. We may consider as σ-field the whole power set of Ω, $\mathcal{F} = \mathcal{P}(\Omega)$, containing all the subsets of Ω.

The definition of the probability measure requires some more work. We introduce a parameter $p \in [0,1]$, which would be equal to $1/2$ if the coin were perfectly balanced, with the interpretation of how likely is the outcome of head in a single toss. Rigorously, this means we are defining:

$$P(A_i) = p, \quad \forall i = 1, \dots, n, \quad A_i = \{\omega = (\omega_1, \dots, \omega_n) \in \Omega |\ \omega_i = 1\} \tag{9}$$

assuming that all tosses are *equivalent*, i.e. $P(A_i)$ does not depend on i. We observe now that the event "the first x tosses ($x = 0, \dots, n$) give head and the other $(n - x)$ tail", contains only the element of Ω:

$$\omega = (1, 1, \dots, 1, 0, \dots, 0) = A_1 \cap \dots \cap A_x \cap A_{x+1}^C \cap \dots \cap A_n^C \tag{10}$$

If we assume that the tosses are **independent** we have necessarily:

$$P\left(\{\omega = (1, 1, \dots, 1, 0, \dots, 0)\}\right) = p^x (1-p)^{n-x} \tag{11}$$

Moreover any element of ω in which the value 1 appears x times and the value 0 $(n - x)$ times has the same probability by construction. We have thus defined $P(\{\omega\})$ for all $\omega \in \Omega$ and thus:

$$P(A) = \sum_{\omega\in A} P(\{\omega\}) \tag{12}$$

for any event $A \in \mathcal{F} = \mathcal{P}(\Omega)$.

The definition of the probability space $(\Omega, \mathcal{F}, P)$ is now completed.

Let's define now the random variable $X : \Omega \to \{0, 1, \dots, n\}$:

$$\omega = (\omega_1, \dots, \omega_n) \rightsquigarrow X(\omega) = \sum_{i=1}^{n} \omega_i \tag{13}$$

where the set $\{0, 1, \dots, n\}$ is trivially measurable once endowed with its power set σ-field $\mathcal{P}(\{0, 1, \dots, n\})$. The random variable X simply counts the number of heads in n independent coin tosses. The law of X can be obtained very simply starting from the probabilities:

$$P(X = x), \quad x \in \{0, 1, \dots, n\} \tag{14}$$

that can be obtained counting the number of different $\omega \in \Omega$ in which the value 1 appears x times and the value 0 $(n - x)$ times: the event

$$\{X = x\} = \{\omega \in \Omega \,|\, \sum_{i=1}^{n} \omega_i = x\} \tag{15}$$

contains indeed all such elements. The result is:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, \dots, n \tag{16}$$

From the knowledge of $P(X = x)$ we immediately obtain the law of X:

$$\mathcal{P}(\{0, 1, \dots, n\}) \ni B \rightsquigarrow \mu(B) = \sum_{x \in B} \binom{n}{x} p^x (1 - p)^x \tag{17}$$

This law is very famous and it is called **binomial law with parameters** (n, p): we will write $X \sim B(n, p)$. In the particular case $n = 1$, $B(1, p)$ is called **Bernoulli law with parameter** p.

In order to make the notations more compact, it is useful to define a function $p : \mathbb{R} \to \mathbb{R}$ as follows:

$$\mathbb{R} \ni x \rightsquigarrow p(x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & if\ x = 0, \dots, n \\ 0 & otherwise \end{cases} \tag{18}$$

Such function is called **discrete density** of X, it is different from zero only in a countable subset of $\mathbb{R}$ and satisfies:

$$p(x) \geq 0, \quad \sum_{x \in \mathbb{R}} p(x) = 1 \tag{19}$$

where the sum is meant in the language of infinite summations theory. The law can be extended naturally to the measurable set made by real numbers equipped with the Borel σ-field $\mathcal{B}(\mathbb{R})$, the smallest σ-field containing the open subset of $\mathbb{R}$, $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

$$\mathcal{B}(\mathbb{R}) \ni B \rightsquigarrow \mu(B) = \sum_{x \in B} p(x) \tag{20}$$

Nota 7 *In this example we have built up explicitely a probability space* $(\Omega, \mathcal{F}, P)$ *and a* **discrete** *random variable* X *(i.e. it assumes only a countable set of values). The law of such random variable turned out to be completely determined by the discrete density* $p(x)$*. The reader will imediately realize that the precise details of the definition of the space* $(\Omega, \mathcal{F}, P)$ *and of* X *can be completely forgotten once the law of* X *is known: they actually have no impact on the probabilistic description of the experiment.*

Starting from the binomial law, it is possible to build up other very important laws. We consider a law $B(n, \frac{\lambda}{n})$, where $\lambda > 0$ is a fixed parameter, and we investigate the asymptotic behavior as $n \to +\infty$:

$$P(X = x) = \binom{n}{x} \left(\tfrac{\lambda}{n}\right)^x (1 - \tfrac{\lambda}{n})^{n-x} = \tfrac{n!}{x!(n-x)!} \tfrac{\lambda^x}{n^x} (1 - \tfrac{\lambda}{n})^{n-x} = \tag{21}$$
$$= \tfrac{\lambda^x}{x!}(1 - \tfrac{\lambda}{n})^n \tfrac{n(n-1)\dots(n-x+1)}{n^x}(1 - \tfrac{\lambda}{n})^{-x} \overset{n \to +\infty}{\to} \tfrac{\lambda^x}{x!} e^{-\lambda}$$

The last expression provides the definition of the **Poisson law with parameter** λ, related to the discrete density:

$$\mathbb{R} \ni x \rightsquigarrow p(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & x = 0, 1, 2, \dots \\ 0 & otherwise \end{cases} \tag{22}$$

It is simple to check that the above function actually defines a discrete density:

$$\sum_{x \in \mathbb{R}} p(x) = \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!} e^{-\lambda} = e^{\lambda} e^{-\lambda} = 1 \tag{23}$$

A Poisson law is often used to model experiments in which a system of many objects is observed (for example a collection of nuclei), each having a very low probability to undergo a certain phenomenon (for example radioactive decay).

Another interesting question is the following: what is the probability that the first head appears precisely at the x-th toss? Let T be the random variable providing the toss in which we obtain the first head. To evaluate $P(T = x)$. we can use the following simple identity:

$$\{T = x\} \cup \{T > x\} = \{T > x - 1\} \tag{24}$$

implying, since naturally $\{T = x\} \cap \{T > x\} = \emptyset$:

$$P(T = x) + P(T > x) = P(T > x - 1) \tag{25}$$

The key point is that the event $\{T > x\}$ corresponds to no heads in the first x tosses, so that the probability is:

$$P(T > x) = \binom{x}{0} p^0 (1 - p)^{x-0} = (1 - p)^x \tag{26}$$

It follows that:

$$P(T = x) = P(T > x-1) - P(T > x) = (1-p)^{x-1} - (1-p)^x = p(1-p)^{x-1} \tag{27}$$

We call **geometric law of parameter** $p \in [0, 1]$ the law associated with the discrete density:

$$\mathbb{R} \ni x \rightsquigarrow p(x) = \begin{cases} p\,(1-p)^x & x = 0, 1, 2, \dots \\ 0 & otherwise \end{cases} \tag{28}$$

In our example, $T - 1$ has a geometric law.

III. ABSOLUTELY CONTINUOUS RANDOM VARIABLES

In the preceding examples we have presented our first examples of random variables. All such examples involved real valued **discrete** random variables, taking values in a countable subset of $\mathbb{R}$; their law is univocally determined by the discrete density $p(x)$, non-zero only inside a countable set, non-negative and normalized to one. The law has the form:

$$\mathcal{B}(\mathbb{R}) \ni B \rightsquigarrow \mu(B) = \sum_{x \in B} p(x) \tag{29}$$

The generalization to multidimensional **discrete** random variables is straightforward: one simply defines discrete densities $p(\underline{x})$ on $\mathbb{R}^d$, related to the laws of random variables of the form $X = (X_1, \dots, X_d)$ where, naturally, the X_i are real valued discrete random variables.

In general, we will very often meet random variables which are not discrete. The simplest example is provided by the **uniform law** in $(0, 1)$: we will say that a random variable is uniform in $(0, 1)$ if its law has the form:

$$\mu(B) = \int_B dx\, p(x) \tag{30}$$

where:

$$p(x) = \begin{cases} 1 & x \in (0, 1) \\ 0 & x \notin (0, 1) \end{cases} \tag{31}$$

The values of X cover uniformly the interval $(0, 1)$. Random variables uniform in $(0, 1)$ will be very important in the definition of sampling techniques in the following chapters: the key point is that, with a computer, we can *generate* the values of X, in a sense which will be later clarified.

It is evident that, in (31), the discrete density has been replaced by a continuous density and the summation has been replaced by an integral. The random variable X belongs to a very important class of random variables, defined in the following definition.

Definizione 8 *We say that a random variables taking values in* $\mathbb{R}^d$, $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *is* **absolutely continuous** *if the law* μ *of* X *is absolutely continuous with respect to the Lebesgue measure, that is* $\mu(B) = 0$ *whenever* B *has Lebesgue measure equal to zero.*

If X is absolutely continuous, the Radon–Nicodym theorem, from measure theory, ensures the existence of the **density of** X, i.e. a function $p : \mathbb{R}^d \to \mathbb{R}$ non-negative, Borel-measurable, Lebesgue-integrable with $\int_{\mathbb{R}^d} d\underline{x}\, p(\underline{x}) = 1$, and such that:

$$\mu(B) = \int_B d\underline{x}\, p(\underline{x}) = \int_{\mathbb{R}^d} d\underline{x}\, 1_B(\underline{x})\, p(\underline{x}) \quad \forall B \in \mathcal{B}(\mathbb{R}^d) \tag{32}$$

The density of a random variable is unique almost everywhere with respect to Lebesgue measure: if p e p' are two densities of a random variable X, then necessarily they coincide everywhere but inside a set of zero Lebesgue measure.

Nota 9 *We invite the reader to observe the similarity between the discrete and the absolutely continuous case. If a random variable has discrete density* $p_d(\underline{x})$*, then:*

$$\mu(B) = \sum_{\underline{x} \in B} p_d(\underline{x}) \tag{33}$$

while, if it is absolutely continuous, there exist a density $p_c(\underline{x})$ *such that:*

$$\mu(B) = \int_B d\underline{x}\, p_c(\underline{x}) \tag{34}$$

Several authors unify the two cases using the integral notation defining $p_d(\underline{x})$ *as a sum of Dirac's deltas. We prefer not to use such a notation.*

We observe, on the other hand, that there exist random variables which are neither discrete nor absolutely continuous.

In the case of real valued random variables, it is always possible to define the **cumulative distribution function**, $F : \mathbb{R} \to \mathbb{R}$ as follows:

$$\mathbb{R} \ni x \rightsquigarrow F(x) \stackrel{def}{=} \mu\left((-\infty, x]\right) = P(X \le x) \tag{35}$$

By construction, F is increasing, right-continuous, and satisfies:

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to +\infty} F(x) = 1 \tag{36}$$

It is possible to show that there is a one to one correspondence between the cumulative distribution function and the law of a random variable. Moreover, for any function F possessing the above mentioned properties, there exists a random variable having F as the cumulative distribution function.

In particular useful relations are:

$$\mu\left((x, y]\right) = F(y) - F(x), \quad \mu\left((x, y)\right) = F(y^-) - F(x^-), \dots \tag{37}$$

and:

$$\mu(\{x\}) = F(x) - F(x^-) \tag{38}$$

where we use the notation $F(x^-) = \lim_{t \to x^-} F(t)$.

If X is absolutely continuous we have:

$$F(x) = \int_{-\infty}^{x} dy\, p(y) \tag{39}$$

Moreover, if the density is continuous on $\mathbb{R}$, the cumulative distribution function is differentiable on $\mathbb{R}$ and we have:

$$p(x) = \frac{dF(x)}{dx} \tag{40}$$

In general (40) is not true for any absolutely continuous random variable, since the density can happen not to be continuous; however, it can be shown that the cumulative distribution function is always differentiable almost everywhere with respect to the Lebesgue measure, and it is always possible to modify the density in such a way that it coincides with the derivative of F in all points where such derivative exists.

Esempio 1 *Our first examples of absolutely continuous random variables are the following:*

1. *If the density is:*

$$p(x) = \begin{cases} 1, & x \in (0,1) \\ 0, & x \notin (0,1) \end{cases} \tag{41}$$

we will say that X is **Uniform** *in* $(0,1)$.

2. *If the density is:*

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{42}$$

we will say that X is **Standard Normal** *and we will write $X \sim N(0,1)$.*

IV. INTEGRATION OF RANDOM VARIABLES

For real valued random variables, it is possible to introduce the notion of abstract integral with respect to the probability measure. As usual in abstract integration theories, one works with extended functions, that is random variables $X : \Omega \to \overline{R} = \mathbb{R} \cup \{\pm\infty\}$, endowing $\overline{R}$ with the Borel σ-field. Such extension is completely innocuous, and simply meant to work with limits, superiors or inferiors extrema. The construction is very simple, and will be sketched here starting from the class of random variables introduced in the following definition.

Definizione 10 *We say that X is* **simple** *if it can be written as:*

$$X(\omega) = \sum_{i=1}^{n} a_i \, 1_{A_i}(\omega) \tag{43}$$

where n is an integer number, $a_i \in \mathbb{R}$, $A_i \in \mathcal{F}$, $i = 1, \dots, n$.

The integral of simple random variables is defined as follows:

Definizione 11 *If X is simple we define* **expectation** *or* **abstract integral** *of X with respect to the probability measure P, denoted $\int_\Omega X(\omega)P(d\omega)$:*

$$\int_\Omega X(\omega)P(d\omega) \stackrel{def}{=} \sum_{i=1}^{n} a_i P(A_i) \tag{44}$$

If X is a non-negative random variable, we define **expectation** *or* **abstract integral** *of X with respect to the probability measure P, denoted $\int_\Omega X(\omega)P(d\omega)$ the extended real number:*

$$\int_\Omega X(\omega)P(d\omega) \stackrel{def}{=} \sup\left\{\int_\Omega Y(\omega)P(d\omega) \,:\, Y \text{ simple},\ 0 \le Y \le X\right\} \tag{45}$$

which can be equal to $+\infty$.

In the most general case, we let $X^+ = max(X,0)$ e $X^- = -min(X,0)$ and introduce the following definition:

Definizione 12 *We say that $X : \Omega \to \overline{R}$ is* **integrable** *if $\int_\Omega X^+(\omega)P(d\omega) < +\infty$ and $\int_\Omega X^-(\omega)P(d\omega) < +\infty$. In such case we define* **expectation** *or* **abstract integral** *of X with respect to the probability measure P, and denote $\int_\Omega X(\omega)P(d\omega)$, the real number:*

$$\int_\Omega X(\omega)P(d\omega) \stackrel{def}{=} \int_\Omega X^+(\omega)P(d\omega) - \int_\Omega X^-(\omega)P(d\omega) \tag{46}$$

We stress the important identity:

$$\int_\Omega 1_A(\omega)P(d\omega) = P(A), \quad \forall A \in \mathcal{F} \tag{47}$$

Definizione 13 *The set of* **integrable** *random variables $X : \Omega \to \overline{R} = \mathbb{R} \cup \{\pm\infty\}$ will be denoted $\mathcal{L}(\Omega, \mathcal{F}, P)$*

Readers familiar with Lebesgue theory of integration will certainly not be surprised by the following elementary properties of the abstract integral, which will be stated without proof:

1. If X, Y are integrable random variables, $\alpha X + \beta Y$ is integrable for all $\alpha, \beta \in \mathbb{R}$ and:

$$\int_\Omega (\alpha X + \beta Y)(\omega)P(d\omega) = \alpha \int_\Omega X(\omega)P(d\omega) + \beta \int_\Omega Y(\omega)P(d\omega) \tag{48}$$

2. If $X \ge 0$, then $\int_\Omega X(\omega)P(d\omega) \ge 0$. If moreover $Y \ge 0$ is integrable and $0 \le X \le Y$, then X is integrable and $\int_\Omega X(\omega)P(d\omega) \le \int_\Omega Y(\omega)P(d\omega)$.

3. $X \in \mathcal{L}(\Omega, \mathcal{F}, P)$ if and only if $|X| \in \mathcal{L}(\Omega, \mathcal{F}, P)$, and, in such case $|\int_\Omega X(\omega)P(d\omega)| \leq \int_\Omega |X(\omega)|P(d\omega)$

4. If $X = Y$ **almost surely** (a.s), i.e. if there exists an event N, $P(N) = 0$, such that $X(\omega) = Y(\omega)$, $\forall \omega \in N^c$, then $\int_\Omega X(\omega)P(d\omega) = \int_\Omega Y(\omega)P(d\omega)$

The following properties concern limits and approximations:

Teorema 14 *If X is non-negative, there exists a sequence $\{X_n\}_n$ of simple, non-negative random variables, such that $X_n(\omega) \leq X_{n+1}(\omega)$ for each ω and pointwise converging to X, that is:*

$$\lim_{n\to+\infty} X_n(\omega) = X(\omega), \quad \forall \omega \in \Omega \tag{49}$$

Teorema 15 *(***Monotone convergence theorem***) If a sequence $\{X_n\}_n$ of non-negative random variables, such that $X_n(\omega) \leq X_{n+1}(\omega)$, converges pointwise almost surely to a (non-negative) random variable X, that is:*

$$\lim_{n\to+\infty} X_n(\omega) = X(\omega) \quad a.s. \tag{50}$$

then:

$$\lim_{n\to+\infty} \int_\Omega X_n(\omega)P(d\omega) = \int_\Omega X(\omega)P(d\omega) \tag{51}$$

even if $\int_\Omega X(\omega)P(d\omega) = +\infty$. In particular, if $\lim_{n\to+\infty} \int_\Omega X_n(\omega)P(d\omega) < +\infty$, then $X \in \mathcal{L}(\Omega, \mathcal{F}, P)$.

Teorema 16 *(***Dominated convergence theorem***) If a sequence $\{X_n\}_n$ of random variables converges almost surely to a random variable X:*

$$\lim_{n\to+\infty} X_n(\omega) = X(\omega), \quad a.s. \tag{52}$$

and $|X_n| \leq Y$, for all n, where $Y \in \mathcal{L}(\Omega, \mathcal{F}, P)$, then $X_n \in \mathcal{L}(\Omega, \mathcal{F}, P)$, $X \in \mathcal{L}(\Omega, \mathcal{F}, P)$ and:

$$\lim_{n\to+\infty} \int_\Omega X_n(\omega)P(d\omega) = \int_\Omega X(\omega)P(d\omega) \tag{53}$$

Definizione 17 *We say that two random variables X and Y are equivalent if $X = Y$ almost surely. We denote $L^1(\Omega, \mathcal{F}, P)$ the set made of equivalence classes of integrable random variables:*

$$\int_\Omega |X(\omega)|P(d\omega) < +\infty \tag{54}$$

We denote $L^2(\Omega, \mathcal{F}, P)$ the set made of equivalence classes of square-integrable random variables:

$$\int_\Omega |X(\omega)|^2 P(d\omega) < +\infty \tag{55}$$

Nota 18 *In mathematics textbooks, the difference between a random variable and an equivalence class of random variables is often ignored, when this cannot give rise to confusion.*

We mention the following result:

Teorema 19 *$L^1(\Omega, \mathcal{F}, P)$ and $L^2(\Omega, \mathcal{F}, P)$ are linear vector spaces, with $L^2(\Omega, \mathcal{F}, P) \subset L^1(\Omega, \mathcal{F}, P)$; if $X \in L^2(\Omega, \mathcal{F}, P)$ then:*

$$\left(\int_\Omega X(\omega) P(d\omega) \right)^2 \leq \int_\Omega X(\omega)^2 P(d\omega) \tag{56}$$

*Moreover, if $X, Y \in L^2(\Omega, \mathcal{F}, P)$, their product is integrable $XY \in L^1(\Omega, \mathcal{F}, P)$ and the (**Cauchy-Schwarz inequality**) holds:*

$$\left| \int_\Omega X(\omega) Y(\omega) P(d\omega) \right| \leq \sqrt{\int_\Omega X(\omega)^2 P(d\omega) \int_\Omega Y(\omega)^2 P(d\omega)} \tag{57}$$

Let's turn to a useful consequence of the properties of abstract integrals:

Teorema 20 *(**Chebyshev inequality**) If $X \in L^2(\Omega, \mathcal{F}, P)$, for all $a > 0$ the following inequality holds:*

$$P(|X| \geq a) \leq \frac{\int_\Omega X(\omega)^2 P(d\omega)}{a^2} \tag{58}$$

Dimostrazione 1 *From the obvious inequality $X^2(\omega) \geq a^2 1_{|X| \geq a}(\omega)$, it follows that:*

$$\int_\Omega X(\omega)^2 P(d\omega) \geq \int_\Omega a^2 1_{|X| \geq a}(\omega) P(d\omega) = a^2\, P(|X| \geq a) \tag{59}$$

which is just the statement of the theorem.

We will often use the notation:

$$E[X] \stackrel{def}{=} \int_\Omega X(\omega) P(d\omega) \tag{60}$$

when no confusion can rise about the probability space over which we are integrating.

We stress the identity:

$$E[1_A] = P(A), \quad \forall A \in \mathcal{F} \tag{61}$$

which will be frequently used.

Definizione 21 *If $X \in L^2(\Omega, \mathcal{F}, P)$, we call **Variance** of X and denote $Var(X)$ the non-negative real number:*

$$Var(X) \stackrel{def}{=} E\left[(X - E[X])^2\right] \tag{62}$$

The Chebyshev inequality implies the following:

$$P(|X - E[X]| \geq a) \leq \frac{Var(X)}{a^2} \tag{63}$$

which provides an interpretation of the variance: $Var(X)$ controls the dispersion of the values of X around the expectation $E[X]$.

A. Integration with respect to the law of a random variable

Let $X : (\Omega, \mathcal{F}, P) \to (E, \mathcal{E})$ be a random variable, and μ its law, i.e. $X \sim \mu$. We know that $(E, \mathcal{E}, \mu)$ is a probability space. Let now $h : (E, \mathcal{E}, \mu) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a measurable function. Then $h \circ X$ is a composition of measurable functions and therefore a real random variable. We are going to prove now the following very important theorem:

Teorema 22 *If $h \geq 0$ the following equality holds:*

$$\int_\Omega (h \circ X)(\omega) P(d\omega) = \int_E h(x)\mu(dx) \tag{64}$$

even when both members are equal to $+\infty$. Moreover, for any h, $h \circ X \in L^1(\Omega, \mathcal{F}, P)$ if and only if $h \in L^1(E, \mathcal{E}, \mu)$ and, in such case, the above written equality holds.

Dimostrazione 2 *We preliminarly observe that both the members make sense, being two abstract integrals on two different probability spaces. Now, we fix $B \in \mathcal{E}$ and we remind the reader the definition of the law of X:*

$$\mu(B) = P(X \in B) \tag{65}$$

On the other hand, we have:

$$P(X \in B) = \int_\Omega 1_{X \in B}(\omega) P(d\omega) = \int_\Omega 1_B(X(\omega)) P(d\omega) \tag{66}$$

and:

$$\mu(B) = \int_E 1_B(x)\mu(dx) \tag{67}$$

so that the statement of the theorem is true if $h(x) = 1_B(x)$. By linearity, the equality holds also when h is a simple function.

Let now $h \geq 0$; we know that there exists a sequence $\{h_n\}_n$ of non-negative simple functions, satisfying $h_n(x) \leq h_{n+1}(x)$, and pointwise converging to h. Then, the monotone convergence theorem, applied twice, justifies the following chain of equalities:

$$\begin{aligned} \textstyle\int_E h(x)\mu(dx) = \int_E \lim_{n\to\infty} h_n(x)\mu(dx) = \\ = \lim_{n\to\infty} \textstyle\int_E h_n(x)\mu(dx) = \lim_{n\to\infty} \int_\Omega (h_n \circ X)(\omega)P(d\omega) = \\ = \textstyle\int_\Omega \lim_{n\to\infty} h_n \circ X P(d\omega) = \int_\Omega (h \circ X)(\omega)P(d\omega) \end{aligned} \tag{68}$$

proving the theorem in the case $h \geq 0$. Finally, if we consider $|h|$, we immediately conclude that $h \circ X \in L^1(\Omega, \mathcal{F}, P)$ if and only if $h \in L^1(E, \mathcal{E}, \mu)$, and the equality between the abstract integrals follows writing $h = h^+ + h^-$.

Now, we focus on the particular case $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$; moreover, we assume X absolutely continuous. If $h = 1_B$ we have:

$$P(X \in B) = \int_\Omega 1_B(X(\omega))P(d\omega) = \mu(B) = \int_B d\underline{x}\, p(\underline{x}) = \int_{\mathbb{R}^d} d\underline{x}\, 1_B(\underline{x})\, p(\underline{x}) \tag{69}$$

where the last integrals are ordinary Lebesgue integrals. We can extend the above result to simple h exploiting linearity and to generic h using the monotone convergence theorem, obtaining the useful identity:

$$\int_\Omega (h \circ X)(\omega)P(d\omega) = \int_{\mathbb{R}^d} dx\, h(\underline{x})\, p(\underline{x}) \tag{70}$$

which holds for any $h \geq 0$ and for any h such that the two integrals exist.

If in particular $d = 1$, $h(x) = x$ and $X \in L^1(\Omega, \mathcal{F}, P)$, we have:

$$E[X] = \int_\Omega X(\omega)P(d\omega) = \int_{-\infty}^{+\infty} dx\, x\, p(x) \tag{71}$$

Moreover, if $X \in L^2(\Omega, \mathcal{F}, P)$, we have:

$$Var(X) = \int_{-\infty}^{+\infty} dx\, (x - E[X])^2\, p(x) \tag{72}$$

If the random variable X is **discrete** and $p(\underline{x})$ is its discrete density, we have:

$$P(X \in B) = \int_\Omega 1_B(X(\omega))P(d\omega) = \mu(B) = \sum_{\underline{x} \in B} p(\underline{x}) = \sum_{\underline{x} \in \mathbb{R}^d} 1_B(\underline{x})p(\underline{x}) \tag{73}$$

and thus:

$$\int_\Omega (h \circ X)(\omega)P(d\omega) = \sum_{\underline{x} \in \mathbb{R}^d} h(\underline{x})p(\underline{x}) \tag{74}$$

if $h \geq 0$ or if the abstract integral and the finite or infinite sum are finite. If $X \in L^1(\Omega, \mathcal{F}, P)$, we have:

$$E[X] = \int_\Omega X(\omega)P(d\omega) = \sum_{x \in \mathbb{R}} x\, p(x) \tag{75}$$

and if $X \in L^2(\Omega, \mathcal{F}, P)$, we have:

$$Var(X) = \sum_{\underline{x} \in \mathbb{R}} (x - E[X])^2\, p(x) \tag{76}$$

V. TRANSFORMATIONS BETWEEN RANDOM VARIABLES

Let now X be a real valued random variable defined on a probability space $(\Omega, \mathcal{F}, P)$, absolutely continuous with density $p_X(x)$. Let also $g : \mathbb{R} \to \mathbb{R}$ be a Borel-measurable function. $Y = g(X)$ is naturally a random variable. It would be useful to express its law, or its density (if it exists), in terms of the law of X. We start from the cumulative distribution function:

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \in B_y) = \int_{B_y} dx\, p_X(x) \quad (77)$$

where:

$$B_y = \{x \in \mathbb{R} : \ g(x) \leq y\} \quad (78)$$

Let us examine first the simple case in which $g : \mathbb{R} \to \mathbb{R}$ is bijective, differentiable with continuous derivative with $\frac{dg}{dx} \neq 0$; in particular, we assume g strictly increasing. Then g is invertible on $\mathbb{R}$ with inverse g^{-1} which is differentiable with continuous non-vanishing derivative; thus:

$$B_y = \{x \in \mathbb{R} : \ g(x) \leq y\} = \{x \in \mathbb{R} : \ x \leq g^{-1}(y)\} \quad (79)$$

which implies that:

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) \quad (80)$$

where F_X is the cumulative distribution function of X. Under the ipothesis we have fixed about the function g, if F_X is everywhere differentiable (this happens if $p_X(x)$ is continuous), also F_Y is differentiable, ensuringe the existence of the density of $Y = g(X)$:

$$p_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}(F_X \circ g^{-1})(y) = p_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy} \quad (81)$$

If, on the other hand, g is strictly decreasing, we have:

$$F_Y(y) = P(g(X) \leq y) = P(X > g^{-1}(y)) = 1 - F_X(g^{-1}(y)) \quad (82)$$

and thus:

$$p_Y(y) = \frac{dF_Y(y)}{dy} = -\frac{d}{dy}(F_X \circ g^{-1})(y) = p_X(g^{-1}(y)) \left(-\frac{dg^{-1}(y)}{dy} \right) \quad (83)$$

Combining the above results, we have proved the following:

Teorema 23 *If a real random variable X has continuous density $p_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$ is a bijective function, differentiable with continuous derivate never equal to zero, then the random variable $Y = g(X)$ has density given by:*

$$p_Y(y) = p_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \quad (84)$$

As a simple application we consider affine transformations $g(x) = \sigma x + \mu$, $\sigma, \mu \in \mathbb{R}$, $\sigma \neq 0$. Since $g^{-1}(y) = \frac{y-\mu}{\sigma}$, the theorem above provides the density of $Y = \sigma X + \mu$:

$$p_Y(y) = p_X\left(\frac{y-\mu}{\sigma}\right)\frac{1}{|\sigma|} \tag{85}$$

In particular, if $X \sim N(0,1)$ is a standard normal random variable, $Y = \sigma X + \mu$, with $\mu \in \mathbb{R}$, $\sigma \in (0, +\infty)$, has density:

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \tag{86}$$

We will say that $Y = \sigma X + \mu$ is a **normal with parameters μ and σ^2** and we will write $Y \sim N(\mu, \sigma^2)$.

When the hypotheses of the above proved theorem do not hold, we have to work directly with the equality:

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \in B_y) = \int_{B_y} dx\, p_X(x) \tag{87}$$

where:

$$B_y = \{x \in \mathbb{R} : \; g(x) \leq y\} \tag{88}$$

For example, let $X \sim N(0,1)$ be a standard normal and $g(x) = x^2$; we wish to evaluate the density of $Y = g(X) = X^2$. Naturally $F_Y(y)$ is zero if $y \leq 0$. On the other hand, if $y > 0$, we have:

$$\begin{aligned} F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \\ = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned} \tag{89}$$

with:

$$F_X(\pm\sqrt{y}) = \int_{-\infty}^{\pm\sqrt{y}} dx\, \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \tag{90}$$

We see that $F_Y(y)$ is almost everywhere differentiable (except at the origin), and thus we can obtain the density by differentiation, obtaining:

$$p_Y(y) = \frac{1}{\sqrt{2\pi}}\frac{\exp(-y/2)}{\sqrt{y}}\,1_{(0,+\infty)}(y) \tag{91}$$

We say that $Y = X^2$ is **chi-square with one degree of freedom**, and we write $Y \sim \chi^2(1)$.

VI. MULTI-DIMENSIONAL RANDOM VARIABLES

Now we consider random variables taking values in $\mathbb{R}^d$. For simplicity of exposition, we work with $d = 2$, but the results can be readily generalized to higher dimensions, with less transparent notations.

Let therefore $X = (Y, Z)$ be a two-dimensional random variable, absolutely continuous with density $p(y, z)$. An interesting result is the following, which we just state without proof:

Teorema 24 *Y and Z are real random variables absolutely continuous, with densities $p_Y(y)$ e $p_Z(z)$ given by:*

$$p_Y(y) = \int_{-\infty}^{+\infty} dz\, p(y, z), \quad p_Z(z) = \int_{-\infty}^{+\infty} dy\, p(y, z) \tag{92}$$

Moreover, Y and Z are independent if and only if:

$$p(y, z) = p_Y(y) p_Z(z) \tag{93}$$

almost everywhere with respect to Lebesgue measure.

The densities $p_Y(y)$ and $p_Z(z)$ are called **marginal densities** of p. We observe that, once p is known, the marginal densities can be obtained, but the converse is not true: if we know $p_Y(y)$ and $p_Z(z)$, we can obtain p only if Y and Z are independent.

If $Y, Z \in L^2(\Omega, \mathcal{F}, P)$, then the product YZ is integrable: $YZ \in L^1(\Omega, \mathcal{F}, P)$. We call **covariance** of Y and Z, and we denote $Cov(Y, Z)$, the real number:

$$Cov(Y, Z) = E\left[(Y - E[Y])(Z - E[Z])\right] = E\left[YZ\right] - E[Y]E[Z] \tag{94}$$

From the theorem of integration with respect to the law of a random variable, in the special case $X = (Y, Z)$, $(E, \mathcal{E}) = (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ and $h(y, z) = yz$, we obtain the following identity:

$$E\left[YZ\right] = \int_{\mathbb{R}^2} dydz\, yz\, p(y, z) \tag{95}$$

If Y and Z are independent, then:

$$E\left[YZ\right] = \int_{\mathbb{R}^2} dydz\, yz\, p_Y(y) p_Z(z) = E[Y]E[Z] \tag{96}$$

thanks to Fubini theorem from Lebesgue integral theory, and thus:

$$Cov(Y, Z) = 0 \tag{97}$$

The property of having null covariance is called **non-correlation**. We have proved that two independent random variables are non-correlated; the converse, in general is not true.

We call **correlation coefficient** of Y and Z the real number:

$$\rho_{YZ} = \frac{Cov(YZ)}{\sqrt{Var(Y)}\sqrt{Var(Z)}} \tag{98}$$

This is zero if Y and Z are non-correlated, and, in general, satifies the following property:

$$-1 \leq \rho_{YZ} \leq 1 \tag{99}$$

which is a simple consequence of Cauchy-Schwarz inequality.

A. Evaluation of laws

Let $X = (Y, Z)$, as before, a two–dimensional random variable absolutely continuous, with density $p(y, z)$. We wish to evaluate the law of $Y + Z$, a real random variable.

Let's evaluate the cumulative distribution function:

$$F(u) = P(Y + Z \leq u) = P(X \in A_u) = \int_{A_u} dydz\, p(y, z) \tag{100}$$

where:

$$A_u = \{(y, z) \in \mathbb{R}^2 : y + z \leq u\} \tag{101}$$

Then:

$$F(u) = \int_{-\infty}^{+\infty} dy \int_{-\infty}^{u-y} dz\, p(y, z) \tag{102}$$

With a change of variables $z \rightsquigarrow z' = z + y$ we have:

$$F(u) = \int_{-\infty}^{u} dz' \int_{-\infty}^{+\infty} dy\, p(y, z' - y) \tag{103}$$

which provides the expression for the density of the sum of two random variables:

$$p_{Y+Z}(u) = \int_{-\infty}^{+\infty} dy\ p(y, u - y) \tag{104}$$

The calculation we have made is a special case of a general procedure, that can be described as follows: let X be a d-dimensional random variable absolutely continuous with density $p_X(\underline{x})$; moreover, let $g : \mathbb{R}^d \to \mathbb{R}^k$ be a Borel-measurable function. $W = g(X)$ is a k-dimensional random variable. If $p_W(\underline{y})$ were the density of W, then, for any Borel set $A \subset \mathbb{R}^k$, the following equality would hold:

$$\int_{\mathbb{R}^k} d\underline{y}\, 1_A(\underline{y}) p_W(\underline{y}) = P(W \in A) = P(X \in g^{-1}(A)) = \int_{\mathbb{R}^d} d\underline{x}\, 1_A(g(\underline{x})) p_X(\underline{x}) \tag{105}$$

where $g^{-1}(A) = \{\underline{x} \in \mathbb{R}^d : g(\underline{x}) \in A\}$. The above relation is very general and, in some cases, allows to evaluate the density $p_W(\underline{y})$. Let's consider the case $d = k$; we assume that p_X is null outside an open set $D \subset \mathbb{R}^d$ and that $g : D \to V$ is a diffeomorphism between D and an open set $V \subset \mathbb{R}^d$. Naturally $p_W(\underline{y})$ will be zero outside V. Then, if $A \subset V$, a basic theorem from mathematical analysis guarantees the validity of the following change of variables:

$$\int_D d\underline{x}\, 1_A(g(\underline{x})) p_X(\underline{x}) = \int_V d\underline{y}\, 1_A(\underline{y}) p_X(g^{-1}(\underline{y})) \left|\det(J_{g^{-1}}(\underline{y}))\right| \tag{106}$$

which implies the following expression for the density of $W = g(X)$:

$$p_W(\underline{y}) = p_X(g^{-1}(\underline{y})) \left|\det(J_{g^{-1}}(\underline{y}))\right| \tag{107}$$

where $J_{g^{-1}}(\underline{y})$ is the Jacobian matrix of g^{-1} evaluated in $\underline{y}$.

VII. CHARACTERISTIC FUNCTIONS

Let X be a d-dimensional random variable, $X = (X_1, \ldots, X_d)$, $X \sim \mu$.

Definizione 25 *The function ϕ_X:*

$$\phi_X : \mathbb{R}^d \to \mathbb{C} \tag{108}$$

$$\mathbb{R}^d \ni \theta \rightsquigarrow \phi_X(\underline{\theta}) \stackrel{def}{=} E\left[\exp\left(i\sum_{k=1}^{d}\theta_k X_k\right)\right] \tag{109}$$

is called **characteristic function** *of X. The complex-valued integral* (109) *is defined as:*

$$E\left[\exp\left(i\sum_{k=1}^{d}\theta_k X_k\right)\right] = E\left[\cos\left(i\sum_{k=1}^{d}\theta_k X_k\right)\right] + iE\left[\sin\left(\sum_{k=1}^{d}\theta_k X_k\right)\right] \tag{110}$$

Since trigonometric functions are measurable and limited, the abstract integrals always exist, implying that the characteristic function is well defined for every random variable. Moreover, applying the theorem of integration with respect to the law of X, we obtain:

$$\phi_X(\underline{\theta}) = \int_{\mathbb{R}^d} \exp(i\underline{\theta}\cdot\underline{x})\,\mu(d\underline{x}) \tag{111}$$

which, if X is absolutely continuous, becomes:

$$\phi_X(\underline{\theta}) = \int_{\mathbb{R}^d} d\underline{x}\,\exp(i\underline{\theta}\cdot\underline{x})\,p(\underline{x}) \tag{112}$$

If X is discrete, on the other hand, we have:

$$\phi_X(\theta) = \sum_{\underline{x}\in\mathbb{R}^d} \exp(i\underline{\theta}\cdot\underline{x})\,p(\underline{x}) \tag{113}$$

There is a one to one correspondence between characteristic functions and laws of random variables: $\phi_X(\underline{\theta})$ uniquely determines the law of X.

We present now some properties of characteristic functions. The first trivial observation is the equality:

$$\phi_X(0) = 1 \tag{114}$$

Moreover, extending to complex valued functions an inequality from abstract integration theory, we have:

$$|\phi_X(\theta)| \leq E\left[| \exp\left(i\sum_{k=1}^{d}\theta_k X_k\right) | \right] = 1, \forall \theta \in \mathbb{R}^d \tag{115}$$

that is the characteristic function is limited. Moreover, the constant function 1 (which is integrable!), dominates any sequence of functions $\exp(i\underline{\theta}_n \cdot \underline{x})$, in the sense of dominated convergence theorem; if $\underline{\theta}_n \to \underline{\theta}$ for $n \to +\infty$, then $\exp(i\underline{\theta}_n \cdot \underline{x}) \to \exp(i\underline{\theta} \cdot \underline{x})$ and, by dominated convergence theorem, $\phi_X(\underline{\theta}_n) \to \phi_X(\underline{\theta})$. Thus the characteristic function is continuous over $\mathbb{R}^d$.

Let's turn to smoothness. We state without proof the following theorem:

Teorema 26 *If $E[|X|^m] < +\infty$ for some integer m, then the characteristic function of X has continuous partial derivatives till order m and the following equality holds:*

$$\frac{\partial^m}{\partial\theta_{j_1}\dots\partial\theta_{j_m}}\phi_X(\underline{\theta}) = i^m\, E\left[X_{j_1}\dots X_{j_m}\,\exp\left(i\sum_{k=1}^{d}\theta_k X_k\right)\right] \tag{116}$$

In the case of real random variables it follows that, if $X \in L^1(\Omega, \mathcal{F}, P)$:

$$E[X] = -i\frac{d\phi_X(0)}{d\theta} \tag{117}$$

and, if $X \in L^2(\Omega, \mathcal{F}, P)$, we have also:

$$E[X^2] = -\frac{d^2\phi_X(0)}{d\theta^2} \tag{118}$$

We now present some examples:

Esempio 2 *If X is* **uniform** *in $(0,1)$ we have:*

$$\phi_X(\theta) = \int_0^1 dx\, e^{i\theta x} = \begin{cases} \frac{e^{i\theta}-1}{i\theta}, & \theta \neq 0 \\ 1, & \theta = 0 \end{cases} \tag{119}$$

Esempio 3 *If X is* **standard normal** *we have:*

$$\phi_X(\theta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} dx\, \exp\left(i\theta x - \frac{x^2}{2}\right) \tag{120}$$

Since the density is an even function, we can write:

$$\phi_X(\theta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} dx\, \cos(\theta x)\exp\left(-\frac{x^2}{2}\right) \tag{121}$$

We know that X is integrable and square-integrable, and thus we can apply the above theorem to evaluate the derivative of the characteristic function:

$$\frac{d\phi_X(\theta)}{d\theta} = -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} dx\, x\sin(\theta x)\exp\left(-\frac{x^2}{2}\right) \tag{122}$$

We observe that the identity:

$$-\int_{-\infty}^{+\infty} dx\, x \sin(\theta x) \exp\left(-\frac{x^2}{2}\right) = \int_{-\infty}^{+\infty} dx\, \sin(\theta x) \left(\frac{d}{dx} \exp\left(-\frac{x^2}{2}\right)\right) \tag{123}$$

allows us to perform integration by part, providing the following result:

$$\frac{d\phi_X(\theta)}{d\theta} = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dx\, \theta \cos(\theta x) \exp\left(-\frac{x^2}{2}\right) = -\theta \phi_X(\theta) \tag{124}$$

We have tus obtained an ordinary differential equation which, together with the initial condition $\phi_X(0) = 1$, has the unique solution:

$$\phi_X(\theta) = \exp\left(-\frac{\theta^2}{2}\right) \tag{125}$$

Esempio 4 *If X is binomial, $X \sim B(n,p)$, we have:*

$$\phi_X(\theta) = \sum_{x=0}^{n} e^{i\theta x} \binom{n}{x} p^x (1-p)^{n-x} = \left(pe^{i\theta} + 1 - p\right)^n \tag{126}$$

where Newton's binomial theorem has been employed.

Esempio 5 *If X is Poisson with parameter λ, then:*

$$\phi_X(\theta) = \sum_{x=0}^{+\infty} e^{i\theta x} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda}\, e^{e^{i\theta}\lambda} = \exp(\lambda(e^{i\theta} - 1)) \tag{127}$$

We turn to independence. If $X = (Y, Z)$ is a 2-dimensional random variable absolutely continuous, with Y and Z **independent**, then, writing $\underline{\theta} = (\theta_1, \theta_2)$, we have:

$$\phi_X(\theta_1, \theta_2) = \int_{\mathbb{R}^2} dydz\, e^{i\theta_1 y + \theta_2 z}\, p_Y(y) p_Z(z) = \phi_Y(\theta_1)\phi_Z(\theta_2) \tag{128}$$

Naturally the above equality can be trivially extended to d-dimensional random variables. In particular, if Y and Z are **independent**, we have:

$$\phi_{Y+Z}(\theta) = \phi_Y(\theta)\phi_Z(\theta) \tag{129}$$

It is possible to prove that, if the equality:

$$\phi_X(\theta_1, \theta_2) = \phi_Y(\theta_1)\phi_Z(\theta_2) \tag{130}$$

holds over the whole $\mathbb{R}^2$, then Y and Z are independent.

VIII. NORMAL LAWS

We have already defined the standard normal law $N(0,1)$, related to the density:

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{131}$$

We have also evaluated the characteristic function of a random variable $X \sim N(0,1)$:

$$\phi_X(\theta) = \exp\left(-\frac{\theta^2}{2}\right) \tag{132}$$

Moreover, we have presented the law $N(\mu, \sigma^2)$, related to the density:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{133}$$

We already know that, if $X \sim N(0,1)$, then $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$. The characterictic function of Y can thus be readily evaluated:

$$\phi_Y(\theta) = E\left[e^{i\theta Y}\right] = E\left[e^{i\theta(\sigma X + \mu)}\right] = e^{i\theta\mu}\phi_X(\sigma\theta) = \exp\left(i\theta\mu - \frac{\sigma^2\theta^2}{2}\right) \tag{134}$$

and the expected value and the variance are:

$$E[Y] = \mu, \quad Var(X) = \sigma^2 \tag{135}$$

It is useful to extend the definition of a normal law to the d-dimensional case:

Definizione 27 *We say that a d-dimensional random variable X is* **normal** *if its characteristic function has the form:*

$$\phi_X(\underline{\theta}) = \exp\left(i\underline{\theta}\cdot\underline{\mu} - \frac{1}{2}\underline{\theta}\cdot\mathcal{C}\underline{\theta}\right) \tag{136}$$

where $\underline{\mu} \in \mathbb{R}^d$ and $\mathcal{C}$ is a **symmetric, positive semidefinite** *real $d\times d$-matrix. We will write $X \sim N(\underline{\mu}, \mathcal{C})$.*

From linear algebra we learn that, whenever $\mathcal{C}$ is a a **symmetric, positive semidefinite** real $d \times d$-matrix, there exists a **symmetric** real $d \times d$-matrix $\mathcal{A}$ such that:

$$\mathcal{A}^2 = \mathcal{C} \tag{137}$$

Now, if $Z = (Z_1, \ldots, Z_d)$ is a random variable such that $Z_i \sim N(0,1)$, $i = 1, \ldots, d$ and the Z_i are **indipendent**, we have:

$$\phi_Z(\underline{\theta}) = \prod_{i=1}^{d} \exp\left(-\frac{\theta_i^2}{2}\right) = \exp\left(-\frac{|\underline{\theta}|^2}{2}\right) \tag{138}$$

Let's define:

$$X = \mathcal{A}Z + \underline{\mu} \tag{139}$$

We have:

$$\phi_X(\underline{\theta}) = E[e^{i\underline{\theta}\cdot X}] = e^{i\underline{\theta}\cdot\underline{\mu}}\,\phi_Z(\mathcal{A}^T\underline{\theta}) = \exp\left(i\underline{\theta}\cdot\underline{\mu} - \frac{1}{2}\underline{\theta}\cdot\mathcal{C}\underline{\theta}\right) \tag{140}$$

where we have used the fact that $\mathcal{A}$ is symmetric and that $\mathcal{A}^2 = \mathcal{C}$. We have thus shown that, for any choice of the vector $\underline{\mu} \in \mathbb{R}^d$ and of the real, symmetric and positive semidefinite matrix $\mathcal{C}$, there exists a random variable $X \sim N(\underline{\mu}, \mathcal{C})$. Now, the random variable Z has density:

$$p_Z(\underline{z}) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{|\underline{z}|^2}{2}\right) \tag{141}$$

and it is straightforward to check that:

$$E[Z_i] = 0, \quad Cov(Z_i Z_j) = \delta_{ij} \tag{142}$$

Hence follows that:

$$E[X_i] = \sum_{j=1}^{d} \mathcal{A}_{ij} E[Z_j] + \mu_i = \mu_i \tag{143}$$

e:

$$\begin{gathered} Cov(X_i X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = \\ = E\left[(\textstyle\sum_{k=1}^{d} \mathcal{A}_{ik} Z_k)(\sum_{l=1}^{d} \mathcal{A}_{jl} Z_l)\right] = \sum_{k,l=1}^{d} \mathcal{A}_{ik}\mathcal{A}_{jl}\delta_{kl} = \\ = \mathcal{C}_{ij} \end{gathered} \tag{144}$$

where we have used the symmetry of $\mathcal{A}$.

If $\mathcal{C}$ is **invertible**, and thus **positive definite**, than also $\mathcal{A}$ is positive definite and X has density which is given by:

$$p_X(\underline{x}) = \frac{1}{|\det(\mathcal{A})|} p_Z(\mathcal{A}^{-1}(\underline{x} - \underline{\mu})) \tag{145}$$

that is:

$$p_X(\underline{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(\mathcal{C})}} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{d}(x_i - \mu_i)\mathcal{C}_{ij}^{-1}(x_j - \mu_j)\right) \tag{146}$$

We conclude this paragraph with some important observations about normal laws. The first is that linear-affine transformations map normal random variables into normal random variables. In fact, if $X \sim N(\underline{\mu}, \mathcal{C})$ and $Y = \mathcal{B}X + \underline{d}$ we have:

$$\phi_Y(\underline{\theta}) = E[\exp(i\underline{\theta} \cdot Y)] = \exp(i\theta \cdot \underline{d})\phi_X({}^T\mathcal{B}\underline{\theta}) = \tag{147}$$
$$= \exp(i\theta \cdot \underline{d}) \exp\left(i({}^T\mathcal{B})\underline{\theta} \cdot \underline{\mu} - \tfrac{1}{2}({}^T\mathcal{B})\underline{\theta} \cdot \mathcal{C}({}^T\mathcal{B})\underline{\theta}\right) =$$
$$= \exp(i\theta \cdot (\underline{d} + \mathcal{B}\underline{\mu}) - \tfrac{1}{2}({}^T\mathcal{B})\underline{\theta} \cdot \mathcal{C}({}^T\mathcal{B})\underline{\theta}) =$$
$$= \exp(i\theta \cdot (\underline{d} + \mathcal{B}\underline{\mu}) - \tfrac{1}{2}\underline{\theta} \cdot (\mathcal{B}\,\mathcal{C}\,({}^T\mathcal{B}))\underline{\theta})$$

that is $Y \sim N\left(\underline{d} + \mathcal{B}\underline{\mu}, \mathcal{B}\,\mathcal{C}\,({}^T\mathcal{B})\right)$.

We consider now a real random variable of the form $Y = \underline{a} \cdot X = \sum_{i=1}^{d} a_i X_i$, where $\underline{a} \in \mathbb{R}^d$ is a vector. The following calculation:

$$\phi_Y(\theta) = E[\exp(i\theta Y)] = E[\exp(\theta \underline{a} \cdot Y)] = \phi_X(\theta \underline{a}) = \tag{148}$$
$$= \exp\left(i\theta \underline{a} \cdot \underline{\mu} - \tfrac{\theta^2}{2}\underline{a} \cdot \mathcal{C}\underline{a}\right)$$

show that $Y \sim N\left(\underline{a} \cdot \underline{\mu}, \underline{a} \cdot \mathcal{C}\underline{a}\right)$. In particular the components of a normal are normal.

Another important observation concerns independence and non correlation. If $X_1, \dots, X_n$ are **indipendent** real random variables, $X_i \sim N(\mu_i, \sigma_i^2)$, then $X = (X_1, \dots, X_n)$ is normal, $X \sim N(\underline{\mu}, \mathcal{C})$ with $\mathcal{C}_{ij} = \sigma_i^2 \delta_{ij}$, as one can trivially check writing the characteristic function. On the other hand, if $X = (X_1, \dots, X_n)$ is normal, $X \sim N(\underline{\mu}, \mathcal{C})$, with diagonal covariance matrix $\mathcal{C}_{ij} = \sigma_i^2 \delta_{ij}$, then $X_1, \dots, X_n$ are **indipendent** real random variables, $X_i \sim N(\mu_i, \sigma_i^2)$, since the characteristic function is factorized. Thus, if the joint law is normal, independence and non correlation are equivalent properties.

IX. THE LAW OF LARGE NUMBERS AND THE CENTRAL LIMIT THEOREM

We conclude our review of probability theory with two very important results about convergence and approximation.

Before introducing such theorems, we will sketch basic notions about the convergence of random variables. Let's consider a sequence of d-dimensional random variables $\{X_n\}_{n=0}^{\infty}$ and a d-dimensional "limit" random variable X.

Definizione 28 *1. We say that* $\{X_n\}_{n=0}^{\infty}$ **converges almost surely** *to* X *if there exists an event* $N \in \mathcal{F}$, $P(N) = 0$, *such that:*

$$\lim_{n \to \infty} X_n(\omega) = X(\omega), \quad \forall \omega \in N^C \tag{149}$$

2. We say that $\{X_n\}_{n=0}^{\infty}$ **converges in probability** *to* X *if, for all* $\eta > 0$*:*

$$\lim_{n \to \infty} P\left(|X_n - X| \geq \eta\right) = 0 \tag{150}$$

3. *We say that* $\{X_n\}_{n=0}^{\infty}$ **converges in distribution** *to* X *if:*

$$\lim_{n\to\infty} \phi_{X_n}(\theta) = \phi_X(\theta) \tag{151}$$

where $\phi_{X_n}(\theta)$ *and* $\phi_X(\theta)$ *are the characteristic functions of* X_n *and* X *respectively.*

It can be shown that almost surely convergence implies convergence in probability, and convergence in probability implies convergence in distribution, which is the weakest notion of convergence.

Let now $\{X_k\}_{k\geq 1}$ be a sequence of real valued random variables, **independent** and **identically distributed**: this means that all the X_k have the same law. We also assume that all the X_k are square-integrable, and we introduce the notation:

$$\mu = E[X_k], \quad \sigma^2 = Var(X_k) \tag{152}$$

μ and σ^2 are finite by contruction, and do not depend on k because we have assumed the random variables identically distributed. Let's define the **empirical mean**:

$$S_n = \frac{1}{n}\sum_{k=1}^{n} X_n \tag{153}$$

We perform now some calculations:

$$E[S_n] = \frac{1}{n}\sum_{k=1}^{n} E[X_n] = \mu \quad , \tag{154}$$

$$\begin{aligned} Var(S_n) = E[(S_n-\mu)^2] = E[S_n^2] - \mu^2 = \\ = \tfrac{1}{n^2}\textstyle\sum_{i,j=1}^{n} E[X_iX_j] - \mu^2 = \\ = \tfrac{1}{n^2}\textstyle\sum_{i=1}^{n} E[X_i^2] + \tfrac{1}{n^2}\sum_{i\neq j=1}^{n} E[X_iX_j] - \mu^2 = \\ = \tfrac{1}{n^2}\textstyle\sum_{i=1}^{n} \left(Var(X_i) + E[X_i]^2\right) + \tfrac{1}{n^2}\sum_{i\neq j=1}^{n} E[X_i]E[X_j] - \mu^2 = \\ = \tfrac{\sigma^2}{n} + \tfrac{\mu^2}{n} + \tfrac{n(n-1)}{n^2}\mu^2 - \mu^2 = \tfrac{\sigma^2}{n} \end{aligned} \tag{155}$$

We use now the Chebyshev inequality:

$$P\left(|S_n-\mu| \geq \eta\right) \leq \frac{Var(S_n)}{\eta^2} = \frac{\sigma^2}{n\eta^2} \stackrel{n\to+\infty}{\to} 0 \tag{156}$$

We have proved in this way a very important result:

Teorema 29 *(***Weak law of large numbers***) The sequence of empirical means* $\{S_n\}_{n\geq 1}$ *of independent and indentically distributed real square-integrable random variables* $\{X_k\}_{k\geq 1}$ *with expected value* μ*,* **converges in probability** *to* μ*:*

$$\lim_{n\to+\infty} P\left(|S_n-\mu| \geq \eta\right) = 0, \quad \forall \eta > 0 \tag{157}$$

Nota 30 *It can be shown that this convergence result can be proved also under weakened ipothesis, removing the assumption of finite variance, and with a stronger notion of convergence:*

Teorema 31 *(***Strong law of large numbers***) The sequence of empirical means* $\{S_n\}_{n\geq 1}$ *of independent and indentically distributed real integrable random variables* $\{X_k\}_{k\geq 1}$ *with expected value* μ*,* **converges almost surely** *to* μ*.*

We omit the proof of such result.

Now, let's introduce:

$$S_n^\star = \frac{S_n - \mu}{\sigma/\sqrt{n}} \tag{158}$$

We may write:

$$S_n^\star = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_k, \quad Y_k = \frac{X_k - \mu}{\sigma} \tag{159}$$

where the random variables Y_k are clearly independent and identically distributed, and satisfy:

$$E[Y_k] = 0, \quad Var(Y_k) = 1 \tag{160}$$

Nota 32 *We observe that the expression:*

$$S_n^\star = \frac{Y_1}{\sqrt{n}} + \cdots + \frac{Y_n}{\sqrt{n}} \tag{161}$$

suggests the idea of a sum of many small independent non systematic (zero mean) effects. This could remind the reader the theory of errors which he/she has learned in university courses.

Since the Y_k are identically distributed, they have the same characteristic function, which we will denote simply ϕ. We evaluate now the characteristic function of $S_n^\star$:

$$\phi_{S_n^\star}(\theta) = E[\exp(i\theta S_n^\star)] = E[\exp(i\theta \tfrac{1}{\sqrt{n}} \textstyle\sum_{k=1}^{n} Y_k)] = \tag{162}$$

$$= \left(\phi(\tfrac{\theta}{\sqrt{n}})\right)^n = \exp\left(n \log(\phi(\tfrac{\theta}{\sqrt{n}}))\right) = \exp\left(n \log\left(1 + \phi(\tfrac{\theta}{\sqrt{n}}) - 1\right)\right)$$

Since the characteristic functions are always continuous, we have $\phi(\frac{\theta}{\sqrt{n}}) \to \phi(0) = 1$ if $n \to +\infty$ for any fixed θ; this implies the asymptotic behavior:

$$n \log\left(1 + \phi(\frac{\theta}{\sqrt{n}}) - 1\right) \overset{n\to+\infty}{\sim} n\left(\phi(\frac{\theta}{\sqrt{n}}) - 1\right) \tag{163}$$

The Y_k are by construction square-integrable, and thus:

$$\frac{d\phi(0)}{d\theta} = iE[Y_k] = 0, \quad \frac{d^2\phi(0)}{d\theta^2} = -Var(Y_k) = -1 \tag{164}$$

so that:

$$\phi(\frac{\theta}{\sqrt{n}}) - 1 = -\frac{\theta^2}{2n} + o(\frac{1}{n}) \tag{165}$$

which implies:

$$n \log \left(1 + \phi(\frac{\theta}{\sqrt{n}}) - 1\right) \stackrel{n\to+\infty}{\sim} n \left(\phi(\frac{\theta}{\sqrt{n}}) - 1\right) \stackrel{n\to+\infty}{\sim} -\frac{\theta^2}{2} \tag{166}$$

We have thus found the following very important result:

$$\lim_{n\to+\infty} \phi_{S_n^\star}(\theta) = \exp\left(-\frac{\theta^2}{2}\right) \tag{167}$$

where in the right hand side we have the characteristic function of a standard normal random variable $Z \sim N(0,1)$.

We summarize what we have found in the following:

Teorema 33 *(***Central Limit Theorem***) If $\{X_k\}_{k\geq 1}$ is a sequence of real valued square integrable random variables,* **independent** *and* **identically distributed***, letting $\mu = E[X_k]$ and $\sigma^2 = Var(X_k)$, the sequence:*

$$S_n^\star = \frac{\frac{1}{n}\sum_{k=1}^n X_k - \mu}{\sigma/\sqrt{n}} \tag{168}$$

converges in distribution *to a standard normal random variable $Z \sim N(0,1)$.*

We use the notation:

$$S_n^\star \stackrel{\mathcal{D}}{\to} Z \sim N(0,1), \quad n \to +\infty \tag{169}$$

where the arrow indicate convergence in distribution.