# Lecture 4
# STOCHASTIC PROCESSES

Stéphane ATTAL

**Abstract** This lecture contains the basics of Stochastic Process Theory. It starts with a quick review of the language of Probability Theory, of random variables, their laws and their convergences, of conditional laws and conditional expectations. We then explore stochastic processes, their laws, existence theorems, path regularity. We construct and study the two cornerstones of Stochastic Process Theory and Stochastic Calculus: the Brownian motion and the Poisson processes. We end up this lecture with the very probabilistic notions of filtrations and of stopping times.



Stéphane ATTAL
Institut Camille Jordan, Université Lyon 1, France
e-mail: attal@math.univ-lyon1.fr

We assume that the reader is familiar with the general Measure and Integration Theory. We start this lecture with those notions which are really specific to Probability Theory.

## 4.1 Monotone Class Theorems

Before entering into the basic concepts of Probability Theory, we want to put the emphasis on a particular set of results: the *Monotone Class Theorems*. These theorems could be classified as being part of the general Measure and Integration Theory, but they are used so often in Probability Theory that it seemed important to state them properly. It is also our experience, when teaching students or discussing with colleagues, that these theorems are not so much well-known outside of the probabilist community.

### *4.1.1 Set Version*

**Definition 4.1.** Let $\Omega$ be a set, we denote by $\mathcal{P}(\Omega)$ the set of all subsets of $\Omega$. A collection $\mathcal{S}$ of subsets of $\Omega$ is a *$\delta$-system* (on $\Omega$) if

i) $\Omega \in \mathcal{S}$,

ii) if $A, B$ belong to $\mathcal{S}$ with $A \subset B$ then $B \setminus A$ belongs to $\mathcal{S}$,

iii) if $(A_n)$ is an increasing sequence in $\mathcal{S}$ then $\cup_n A_n$ belongs to $\mathcal{S}$.

A collection $\mathcal{C}$ of subsets of $\Omega$ is a *$\pi$-system* (on $\Omega$) if it is stable under finite intersections.

The following properties are immediate consequences of the definitions, we leave the proofs to the reader.

**Proposition 4.2.**

*1) $\delta$-systems are stable under passage to the complementary set.*

*2) The intersection of any family of $\delta$-systems on $\Omega$ is a $\delta$-system on $\Omega$.*

*3) A collection of subsets is a $\sigma$-field if and only if it is a $\delta$-system and a $\pi$-system.*

**Definition 4.3.** A consequence of Property 2) above is that, given any collection $\mathcal{C}$ of subsets of $\Omega$, there exists a smallest $\delta$-system $\mathcal{S}$ on $\Omega$ which contains $\mathcal{C}$. This $\delta$-system is simply the intersection of all the $\delta$-systems containing $\mathcal{C}$ (the intersection is non-empty for $\mathcal{P}(\Omega)$ is always a $\delta$-system containing $\mathcal{C}$ ). We call it the *$\delta$-system generated* by $\mathcal{C}$.

**Theorem 4.4 (Monotone Class Theorem, Set Version).** *Let $\Omega$ be a set and $\mathcal{S}$ be a $\delta$-system on $\Omega$. If $\mathcal{S}$ contains a $\pi$-system $\mathcal{C}$, then $\mathcal{S}$ contains the $\sigma$-field $\sigma(\mathcal{C})$ generated by $\mathcal{C}$. More precisely, the $\delta$-system generated by $\mathcal{C}$ coincides with $\sigma(\mathcal{C})$.*

*Proof.* Let $\delta(\mathcal{C})$ be the $\delta$-system generated by $\mathcal{C}$. We have $\mathcal{C} \subset \delta(\mathcal{C}) \subset \mathcal{S}$. For any $A \in \delta(\mathcal{C})$ define $\delta(\mathcal{C})_A = \{B \in \delta(\mathcal{C})\,;\ A \cap B \in \delta(\mathcal{C})\}$ . It is easy to check that $\delta(\mathcal{C})_A$ is a $\delta$-system (left to the reader), contained in $\delta(\mathcal{C})$.

If $A$ belongs to $\mathcal{C}$ we obviously have $\mathcal{C} \subset \delta(\mathcal{C})_A$, for $\mathcal{C}$ is a $\pi$-system. By minimality of $\delta(\mathcal{C})$ we must have $\delta(\mathcal{C}) \subset \delta(\mathcal{C})_A$. Hence $\delta(\mathcal{C})_A = \delta(\mathcal{C})$. This is to say that $A \cap B$ belongs to $\delta(\mathcal{C})$ for all $A \in \mathcal{C}$ and all $B \in \delta(\mathcal{C})$.

Now, take $B \in \delta(\mathcal{C})$ and consider the collection $\delta(\mathcal{C})_B$ . We have just proved that $\delta(\mathcal{C})_B$ contains $\mathcal{C}$. Hence it contains $\delta(\mathcal{C})$, hence it is equal to $\delta(\mathcal{C})$. This is to say that $A \cap B$ belongs to $\delta(\mathcal{C})$ for all $A \in \delta(\mathcal{C})$, all $B \in \delta(\mathcal{C})$. In other words, $\delta(\mathcal{C})$ is a $\pi$-system.

By Property 3) in Proposition 4.2 the collection $\delta(\mathcal{C})$ is a $\sigma$-field. As $\delta(\mathcal{C})$ contains $\mathcal{C}$, it contains $\sigma(\mathcal{C})$. This proves the first part of the theorem.

A $\sigma$-field is a $\delta$-system, hence $\sigma(\mathcal{C})$ is a $\delta$-system containing $\mathcal{C}$. As a consequence $\sigma(\mathcal{C})$ contains the minimal one $\delta(\mathcal{C})$ and finally $\sigma(\mathcal{C}) = \delta(\mathcal{C})$. □

Let us mention here an important consequence of Theorem 4.4.

**Corollary 4.5.** *Let $\mathcal{C}$ be a $\pi$-system of subsets of $\Omega$. Then any probability measure $\mathbb{P}$ on $\sigma(\mathcal{C})$ is determined by its values on $\mathcal{C}$ only.*

*Proof.* Let $\mathbb{P}$ and $\mathbb{P}'$ be two probability measures on $\sigma(\mathcal{C})$ such that $\mathbb{P}(A) = \mathbb{P}'(A)$ for all $A \in \mathcal{C}$. The set $\mathcal{S}$ of all $A \in \sigma(\mathcal{C})$ satisfying $\mathbb{P}(A) = \mathbb{P}'(A)$ is a $\delta$-system (proof left to the reader). It contains $\mathcal{C}$, hence by Theorem 4.4 it contains $\sigma(\mathcal{C})$. □

Note that the above corollary ensures that a measure on $\sigma(\mathcal{C})$ is determined by its value on $\mathcal{C}$. The important point here is that *one already knows that there exists a measure on $\sigma(\mathcal{C})$ which extends $\mathbb{P}$ on $\mathcal{C}$*. This is the place here to recall the Carathéodory Extension Theorem which gives a condition for the *existence of such an extension*. We do not prove here this very classical result of Measure Theory.

**Definition 4.6.** A collection $\mathcal{A}$ of subsets of $\Omega$ is called a *field* if

i) $\Omega \in \mathcal{A}$

ii) $\mathcal{A}$ is stable under passage to the complementary set

iii) $\mathcal{A}$ is stable under finite unions.

A *measure* on a field $\mathcal{A}$ is a map $\mu : \mathcal{A} \mapsto \mathbb{R}^+ \cup \{+\infty\}$ such that

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n)$$

for any sequence $(A_n)$ of two-by-two disjoint elements of $\mathcal{A}$, such that $\cup_n A_n$ belongs to $\mathcal{A}$. The measure $\mu$ is called *$\sigma$-finite* if $\Omega$ can be written as the union of a sequence of sets $(E_n)$, each of which belonging to $\mathcal{A}$ and satisfying $\mu(E_n) < \infty$.

**Theorem 4.7 (Carathéodory's Extension Theorem).** *Let $\mathcal{A}$ be a field on a set $\Omega$. If $\mu$ is a $\sigma$-finite measure on $\mathcal{A}$, then there exists a unique extension of $\mu$ as a $\sigma$-finite measure on $\sigma(\mathcal{A})$.*

### *4.1.2 Functional Versions*

There are very useful forms of the Monotone Class Theorem which deal with bounded functions instead of sets.

**Definition 4.8.** A *monotone vector space on a set* $\Omega$ is a family $\mathcal{H}$ of bounded real-valued functions on $\Omega$ such that

i) $\mathcal{H}$ is a (real) vector space,

ii) $\mathcal{H}$ contains the constant functions,

iii) if $(f_n)$ is any increasing sequence in $\mathcal{H}$ of positive functions, converging to a bounded function $f = \sup_n f_n$, then $f$ belongs to $\mathcal{H}$.

The first functional version of the Monotone Class Theorem is an easy application of Theorem 4.4.

**Theorem 4.9 (Monotone Class Theorem, Functional Version 1).** *Let $\mathcal{H}$ be a monotone vector space on $\Omega$ and let $\mathcal{C}$ be a $\pi$-system on $\Omega$ such that $\mathbb{1}_A$ belongs to $\mathcal{H}$ for all $A \in \mathcal{C}$. Then $\mathcal{H}$ contains all the real-valued, $\sigma(\mathcal{C})$-measurable, bounded functions on $\Omega$.*

*Proof.* Let $\mathcal{S} = \{A \subset \Omega\,;\ \mathbb{1}_A \in \mathcal{H}\}$. Then, by the properties of $\mathcal{H}$, it is easy to check that $\mathcal{S}$ is a $\delta$-system. As $\mathcal{S}$ contains $\mathcal{C}$ by hypothesis, it contains $\sigma(\mathcal{C})$ by Theorem 4.4.

If $f$ is a real-valued, bounded, $\sigma(\mathcal{C})$-measurable function on $\Omega$, then it can be decomposed as $f = f^+ - f^-$, for some positive functions $f^+$ and $f^-$ which are also bounded and $\sigma(\mathcal{C})$-measurable. In particular these two functions are limit of increasing sequences of positive simple bounded functions of the form $\sum_{i=1}^{k} a_i \mathbb{1}_{A_i}$, with $A_i \in \sigma(\mathcal{C})$ for all $i$. This implies that $f^+$ and $f^-$ belong to $\mathcal{H}$ and hence $f$ belongs to $\mathcal{H}$. □

There exists a stronger version of this functional form of the Monotone Class Theorem. Before proving it we need a technical lemma.

**Lemma 4.10.** *Every monotone vector space is stable under uniform limit.*

*Proof.* Let $\mathcal{H}$ be a monotone vector space and let $(f_n)$ be a sequence in $\mathcal{H}$ converging uniformly to $f_\infty$. Up to considering the sequence $(f_n - f_0)$ and, up to extracting a subsequence, we can assume that $f_0 = 0$ and $\sum_{k\in\mathbb{N}} \|f_{k+1} - f_k\|_\infty < \infty$.

For all $n \in \mathbb{N}$, put $a_n = \sum_{k\geq n} \|f_{k+1} - f_k\|_\infty$ and $g_n = f_n - a_n\mathbb{1}$. The function $g_n$ belongs to $\mathcal{H}$, for $\mathcal{H}$ is a monotone vector space. Furthermore, we have

$$g_{n+1} - g_n = f_{n+1} - f_n + \|f_{n+1} - f_n\|_\infty \mathbb{1}\,.$$

As a consequence, the function $g_{n+1} - g_n$ is positive and belongs to $\mathcal{H}$. This is to say that the sequence $(g_n)$ is increasing, it converges to $f_\infty$, which is a bounded function for it is dominated by $a_0$. Hence $f_\infty$ belongs to $\mathcal{H}$. □

**Theorem 4.11 (Monotone Class Theorem, Functional Version 2).** *Let $\mathcal{H}$ be a monotone vector space on $\Omega$. If $\mathcal{C} \subset \mathcal{H}$ is stable under pointwise multiplication then $\mathcal{H}$ contains all the real-valued, $\sigma(\mathcal{C})$-measurable, bounded functions on $\Omega$.*

*Proof.* Let $f_1, \ldots, f_n \in \mathcal{C}$ and let $\Phi : \mathbb{C}^n \mapsto \mathbb{C}$ be a continuous function. We claim that $\Phi(f_1, \ldots, f_n)$ belongs to $\mathcal{H}$. Indeed, by the Stone-Weierstrass Theorem, on any compact set the function $\Phi$ is uniform limit of a sequence $(R_m)$ of polynomial functions in $n$ variables. In particular, if we put $M = \max_{1\leq i\leq n} \|f_i\|_\infty$, then $\Phi$ is the uniform limit of $(R_m)$ on $\{\|x\| \leq M\}$. By hypothesis on $\mathcal{C}$ and $\mathcal{H}$, all the functions $R_m(f_1, \ldots, f_n)$ belong to $\mathcal{H}$ and, by Lemma 4.10, the function $\Phi(f_1, \ldots, f_n)$ belongs to $\mathcal{H}$. Take

$$\Phi(x_1, \ldots, x_n) = \prod_{1\leq i\leq n} \min\left(1\,,\, r\,(x_i - a_i)^+\right)$$

for $r > 0$ and $a_1, \ldots, a_n \in \mathbb{R}$. The (increasing) limit of $\Phi(x_1, \ldots, x_n)$ when $r$ tends to $+\infty$ is $\prod_{1\leq i\leq n} \mathbb{1}_{(x_i > a_i)}$. Hence, if $f_1, \ldots, f_n$ belong to $\mathcal{C}$ the function $\prod_{1\leq i\leq n} \mathbb{1}_{(f_i > a_i)}$ belongs to $\mathcal{H}$, for all $a_1, \ldots, a_n \in \mathbb{R}$.

The collection of sets of the form $\cap_{1\leq i\leq n}(f_i > a_i)$ constitutes a $\pi$-system $\mathcal{D}$ on $\Omega$, whose indicator functions all belong to $\mathcal{H}$. By Theorem 4.9, the space $\mathcal{H}$ contains all the $\sigma(\mathcal{D})$-measurable bounded functions. But clearly $\sigma(\mathcal{D})$ is equal to $\sigma(\mathcal{C})$, this gives the result. □

Let us end up this section with a direct application of Theorem 4.11.

**Corollary 4.12.** *Let $E$ be a metric space together with its Borel $\sigma$-field* $\mathrm{Bor}(E)$. *If $\mathbb{P}$ and $\mathbb{P}'$ are two probability measures on $(E, \mathrm{Bor}(E))$ such that*

$$\int_E f \,\mathrm{d}\mathbb{P} = \int_E f \,\mathrm{d}\mathbb{P}'$$

*for all continuous bounded function $f$ from $E$ to $\mathbb{R}$, then $\mathbb{P} = \mathbb{P}'$.*

*Proof.* Put $\mathcal{H}$ to be the space of bounded Borel functions $f$ from $E$ to $\mathbb{R}$ such that $\int_E f \,\mathrm{d}\mathbb{P} = \int_E f \,\mathrm{d}\mathbb{P}'$. Then clearly $\mathcal{H}$ is a monotone vector space. Put $\mathcal{C}$ to be the algebra of bounded continuous functions from $E$ to $\mathbb{R}$. Our hypothesis implies that $\mathcal{C}$ is included in $\mathcal{H}$, and clearly $\mathcal{C}$ is stable under pointwise product. By Theorem 4.11 the space $\mathcal{H}$ contains all the bounded $\sigma(\mathcal{C})$-measurable functions. But it is easy to check that $\sigma(\mathcal{C})$ coincides with $\mathrm{Bor}(E)$, hence the conclusion. □

Note that the above corollary remains true if we replace "continuous bounded functions" by any set $\mathcal{C}$ of Borel bounded functions which is stable under pointwise product and which satisfies $\sigma(\mathcal{C}) = \mathrm{Bor}(E)$. For example this is true for bounded $C^\infty$-functions, for compactly supported $C^\infty$-functions, etc.

It is also easy to check that this corollary extends to the case where $\mathbb{P}$ and $\mathbb{P}'$ are locally bounded measures on $E$. We leave this extension as an exercise for the reader.

## 4.2 Random Variables

In this section we give a very quick overview of the main definitions and concepts attached to the notion of random variable: their law, their characteristic function, the notion of independence etc.

### *4.2.1 Completion of Probability Spaces*

**Definition 4.13.** A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega$ is a set, $\mathcal{F}$ is a $\sigma$-field of subsets of $\Omega$ and $\mathbb{P}$ is a probability measure on $\mathcal{F}$.

A subset $A$ of $\Omega$ is *negligible* if there exists $B \in \mathcal{F}$ such that $A \subset B$ and $\mathbb{P}(B) = 0$. Note that this notion is relative to $\mathcal{F}$ and $\mathbb{P}$ but we shall not mention them. The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *complete* if $\mathcal{F}$ contains all the negligible sets (relative to $\mathbb{P}$).

An incomplete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ can be easily completed in the following way. Let $\mathcal{N}$ denote the set of all negligible sets of $(\Omega, \mathcal{F}, \mathbb{P})$. Put $\overline{\mathcal{F}}$ to be the $\sigma$-field generated by $\mathcal{F}$ and $\mathcal{N}$. We leave to the reader to prove the following results, they are all very easy except the last property which requires a little more carefulness.

**Proposition 4.14.**

*1) The $\sigma$-field $\overline{\mathcal{F}}$ coincides with the set of subsets of $\Omega$ which are of the form $B \cup N$ for some $B \in \mathcal{F}$ and some $N \in \mathcal{N}$.*

*2) The probability measure $\mathbb{P}$ extends to a probability measure $\overline{\mathbb{P}}$ on $\overline{\mathcal{F}}$ by putting $\overline{\mathbb{P}}(B \cup N) = \mathbb{P}(B)$, for all $B \in \mathcal{F}$, all $N \in \mathcal{N}$.*

*3) The probability space $(\Omega, \overline{\mathcal{F}}, \overline{\mathbb{P}})$ is complete.*

*4) A subset $A \subset \Omega$ belongs to $\overline{\mathcal{F}}$ if and only if there exist $B_1, B_2 \in \mathcal{F}$ such that $B_1 \subset A \subset B_2$ and $\mathbb{P}(B_2 \setminus B_1) = 0$.*

*5) A function $f$ from $\Omega$ to $\mathbb{R}$ is measurable for $\overline{\mathcal{F}}$ if and only if there exist two $\mathcal{F}$-measurable functions $g$ and $h$ such that $g \leq f \leq h$ and $\int_\Omega (h - g)\, \mathrm{d}\mathbb{P} = 0$.*

**Definition 4.15.** The complete probability space $(\Omega, \overline{\mathcal{F}}, \overline{\mathbb{P}})$ is called the *completion* of $(\Omega, \mathcal{F}, \mathbb{P})$.

From now on, all our probability spaces are supposed to be complete, we drop the notations $\overline{\mathcal{F}}$ and $\overline{\mathbb{P}}$ and write $\mathcal{F}$ and $\mathbb{P}$ instead.

### *4.2.2 Laws of Random Variables*

**Definition 4.16.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *random variable* on $\Omega$ is a measurable map $X : \Omega \mapsto \mathbb{R}$, where $\mathbb{R}$ is equipped with its Borel $\sigma$-field $\mathrm{Bor}(\mathbb{R})$.

The *law*, or *distribution*, of a random variable $X$ is the measure $\mu_X = X \circ \mathbb{P}$, image of $\mathbb{P}$ under $X$, that is,

$$\mu_X(A) = \mathbb{P}\left(X^{-1}(A)\right)$$

for all $A \in \mathrm{Bor}(\mathbb{R})$. It is a probability measure on $\mathbb{R}$.

In Probability Theory one commonly uses notations such as $\mathbb{P}(X \in A)$ instead of $\mathbb{P}\left(X^{-1}(A)\right)$, or $\mathbb{P}(X \leq t)$ instead of $\mathbb{P}\left(X^{-1}(]-\infty, t])\right)$, etc.

We recall here the Transfer Theorem, which is very useful when dealing with image measures. This is a very classical theorem of Measure and Integration Theory, we do not prove it here.

**Theorem 4.17 (Transfer Theorem).** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(E, \mathcal{E})$ be a measurable space. Let $X$ be a measurable map from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$ and let $\mu_X$ be the measure on $(E, \mathcal{E})$ which is the image of $\mathbb{P}$ under $X$. Then a measurable function $f : E \to \mathbb{R}$ is $\mu_X$-integrable if and only if the (measurable) function $f \circ X : \Omega \to \mathbb{R}$ is $\mathbb{P}$-integrable.*

*In that case we have*

$$\boxed{\int_\Omega f \circ X(\omega)\, \mathrm{d}\mathbb{P}(\omega) = \int_E f(x)\, \mathrm{d}\mu_X(x)\,.} \tag{4.1}$$

**Definition 4.18.** The *expectation* of an integrable random variable $X$ is the integral

$$\mathbb{E}[X] = \int_\Omega X(\omega)\, \mathrm{d}\mathbb{P}(\omega) = \int_{\mathbb{R}} x\, \mathrm{d}\mu_X(x)\,.$$

The *variance* of a square integrable random variable $X$ is the quantity

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2\,.$$

The quantity

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

is the *standard deviation* of $X$.

More generally, when then are defined, the quantities $\mathbb{E}[X^k]$, $k \in \mathbb{N}$, are called the *moments* of $X$.

**Definition 4.19.** We have defined random variables as being real-valued only, but in Quantum Probability Theory one often considers complex-valued functions of random variables. If $X$ is a real-valued random variable and if $f = g + ih$ is a measurable function from $\mathbb{R}$ to $\mathbb{C}$ (where $g$ and $h$ are the real and the imaginary parts of $f$ respectively), then one extends the expectation linearly by

$$\mathbb{E}[f(X)] = \mathbb{E}[g(X)] + i\,\mathbb{E}[h(X)]\,,$$

when this is well-defined.

**Definition 4.20.** The *characteristic function* of a random variable $X$ is the Fourier transform of its law, that is, the function $\widehat{\mu}_X$ on $\mathbb{R}$ defined by

$$\widehat{\mu}_X(t) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}tX}\right] = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx}\, \mathrm{d}\mu_X(x)\,,$$

for all $t \in \mathbb{R}$.

The well-known invertibility of the Fourier transform has the following important consequence.

**Proposition 4.21.** *The characteristic function of a random variable completely determines its law.*

### *4.2.3 Vector-Valued Random Variables*

**Definition 4.22.** Let $n > 1$ be an integer. A *$n$-tuple* of random variables (one says *pair*, *triple* when $n = 2, 3$ respectively) is a family $(X_1, \dots, X_n)$ of $n$ random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. One often speaks of a *vector-valued random variable* by considering $X = (X_1, \dots, X_n)$ as a random variable with values in $\mathbb{R}^n$.

The *law*, or *distribution*, of a $n$-tuple $(X_1, \dots, X_n)$ is the probability measure $\mu_{(X_1,\dots,X_n)}$ on $\mathbb{R}^n$, image of $\mathbb{P}$ under the mapping $(X_1, \dots, X_n)$ from $\Omega$ to $\mathbb{R}^n$. By Corollary 4.5, this measure is characterized by

$$\mu_{(X_1,\dots,X_n)}(A_1 \times \cdots \times A_n) = \mathbb{P}\left(\bigcap_{i=1}^{n} X_i^{-1}(A_i)\right)$$

for all $A_1, \dots, A_n$ in $\mathrm{Bor}(\mathbb{R})$.

**Definition 4.23.** The Transfer Theorem 4.17 applies to the case of vector-valued random variables, by taking $E = \mathbb{R}^n$. Hence, the Transfer Formula (4.1) becomes

$$\boxed{\begin{aligned}\mathbb{E}\left[f(X_1, \dots, X_n)\right] &= \int_\Omega f\left(X_1(\omega), \dots, X_n(\omega)\right) \mathrm{d}\mathbb{P}(\omega) \\ &= \int_{\mathbb{R}^n} f(x_1, \dots, x_n)\, \mathrm{d}\mu_{(X_1,\dots,X_n)}(x_1, \dots, x_n)\,.\end{aligned}} \tag{4.2}$$

For a $n$-tuple $X = (X_1, \dots, X_n)$ of random variables (or a $n$-dimensional random variable), we let $\mathbb{E}[X]$ be the $n$-dimensional vector $(\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$, if well-defined.

We also denote by $\mathrm{Cov}(X)$ the *covariance matrix* of $X$, that is, the matrix

$$\Big(\mathbb{E}\big[\,(X_i - \mathbb{E}[X_i])\;(X_j - \mathbb{E}[X_j])\,\big]\Big)_{i,j=1}^{n} = \Big(\mathbb{E}[X_i\,X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j]\Big)_{i,j=1}^{n},$$

if it is well-defined (i.e. all the random variables $X_i$ are square integrable).

The matrix $\mathrm{Cov}(X)$ is always real (symmetric) positive, as can be easily checked by the reader.

**Definition 4.24.** The *characteristic function* of a $n$-tuple $(X_1, \dots, X_n)$ is the Fourier transform of its law, that is, the function $\widehat{\mu}_{(X_1,\dots,X_n)}$ on $\mathbb{R}^n$ defined by

$$\widehat{\mu}_{(X_1,\dots,X_n)}(t_1, \dots, t_n) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}(t_1X_1 + \cdots + t_nX_n)}\right].$$

In the same way as for Proposition 4.21 we have a uniqueness property.

**Proposition 4.25.** *The characteristic function of a $n$-tuple completely determines its law.* □

**Definition 4.26.** For a $n$-tuple $(X_1, \dots, X_n)$, we call *marginal laws*, or simply *marginals*, the laws of the (strict) sub-families $(X_{i_1}, \dots, X_{i_p})$ of $(X_1, \dots, X_n)$.

**Proposition 4.27.** *The law of a $n$-tuple $(X_1, \dots, X_n)$ completely determines its marginals.*

*Proof.* Consider a sub-family $(X_{i_1}, \dots, X_{i_p})$ taken from $(X_1, \dots, X_n)$. For any family of Borel sets $A_{i_1}, \dots, A_{i_p}$, define the sets $B_1, \dots, B_n$ by $B_j = A_{i_k}$ if $j = i_k$ for some $k$ and $B_j = \mathbb{R}$ otherwise. We then have

$$\mu_{(X_{i_1},\dots,X_{i_p})}(A_{i_1} \times \cdots \times A_{i_p}) = \mu_{(X_1,\dots,X_n)}(B_1 \times \cdots \times B_n) \,,$$

that is, the claim. □

Note that the converse of Proposition 4.27 is not true: the marginal laws do not determine the law of the whole $n$-tuple, in general.

### *4.2.4 Independence*

**Definition 4.28.** If one knows that an event $A \in \mathcal{F}$ (with $\mathbb{P}(A) \neq 0$) is already realized, then the probability to obtain the realization of an event $B$ is not $\mathbb{P}(B)$ anymore, it is given by the quantity

$$\mathbb{P}(B \,|\, A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \,, \tag{4.3}$$

that is, the (normalized) restriction of the measure $\mathbb{P}$ to the set $A$. The above quantity is called *conditional probability of $B$ knowing $A$*.

**Definition 4.29.** We say that two events $A, B \in \mathcal{F}$ are *independent* if

$$\mathbb{P}(B \,|\, A) = \mathbb{P}(B) \,,$$

that is, if knowing that one of the events is realized has no impact on the probability of the other one to occur. Clearly, by (4.3) this property is equivalent to the relation

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$$

and it is then equivalent to $\mathbb{P}(A \,|\, B) = \mathbb{P}(A)$.

A family of events $\{A_1, \dots, A_n\}$ is said to be an *independent family of events* if

$$\mathbb{P}\left(\bigcap_{i=1}^{k} A_{n_i}\right) = \prod_{i=1}^{k} \mathbb{P}(A_{n_i})$$

for any subset $\{n_1, \dots, n_k\}$ of $\{1, \dots, n\}$.

We say that $n$ random variables $X_1, \ldots, X_n$ are *independent* of each other if for any $B_1, \ldots, B_n \in \mathrm{Bor}(\mathbb{R})$, the events $\left\{X_1^{-1}(B_1), \ldots, X_n^{-1}(B_n)\right\}$ form an independent family of events.

**Proposition 4.30.** *The random variables $X_1, \ldots, X_n$ are independent if and only if the law $\mu_{(X_1,\ldots,X_n)}$ of the $n$-tuple $(X_1, \ldots, X_n)$ is the product measure $\mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$.*

*Proof.* If $\mu_{(X_1,\ldots,X_n)} = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$, then clearly

$$\mu_{(X_{n_1},\ldots,X_{n_k})} = \mu_{X_{n_1}} \otimes \cdots \otimes \mu_{X_{n_k}}$$

for any subset $\{n_1, \ldots, n_k\}$ of $\{1, \ldots, n\}$. In particular

$$\begin{aligned}\mathbb{P}\left(X_{n_1}^{-1}(B_{n_1}) \cap \ldots \cap X_{n_k}^{-1}(B_{n_k})\right) &= \mu_{(X_{n_1},\ldots,X_{n_k})}(B_{n_1} \times \ldots \times B_{n_k}) \\ &= \mu_{X_{n_1}}(B_{n_1}) \ldots \mu_{X_{n_k}}(B_{n_k}) \\ &= \mathbb{P}\left(X_{n_1}^{-1}(B_{n_1})\right) \ldots \mathbb{P}\left(X_{n_k}^{-1}(B_{n_k})\right).\end{aligned}$$

This gives the independence.

Conversely, if the random variables $X_1, \ldots, X_n$ are independent, then we get

$$\mu_{(X_1,\ldots,X_n)}(B_1 \times \ldots \times B_n) = \mu_{X_1}(B_1) \ldots \mu_{X_n}(B_n)$$

with the same kind of computation as above. Hence the probability measures $\mu_{(X_1,\ldots,X_n)}$ and $\mu_{X_1} \otimes \cdots \otimes \mu_{X_n}$ coincide on the product sets $B_1 \times \ldots \times B_n$. By Corollary 4.5 they are equal. □

This proposition shows that *in the case of independence* the individual laws of the $X_i$'s determine the law of the $n$-tuple $(X_1, \ldots, X_n)$. This property is in general false without the independence hypothesis.

There exists a very useful functional characterization of independence.

**Proposition 4.31.** *Let $X_1, \ldots, X_n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Then they are independent if and only if*

$$\mathbb{E}\left[f_1(X_1) \ldots f_n(X_n)\right] = \mathbb{E}\left[f_1(X_1)\right] \ldots \mathbb{E}\left[f_n(X_n)\right] \tag{4.4}$$

*for all bounded Borel functions $f_1, \ldots, f_n$ from $\mathbb{R}$ to $\mathbb{R}$.*

*Proof.* If (4.4) holds true for all bounded Borel functions $f_1, \ldots, f_n$ on $\mathbb{R}$ then taking $f_i = \mathbb{1}_{A_i}$, for all $i$, gives

$$\begin{aligned}\mu_{(X_1,\ldots,X_n)}(A_1 \times \ldots \times A_n) &= \mathbb{P}\left(X_1^{-1}(A_1) \cap \ldots \cap X_n^{-1}(A_n)\right) \\ &= \mathbb{E}\left[\mathbb{1}_{A_1}(X_1) \ldots \mathbb{1}_{A_n}(X_n)\right] \\ &= \mathbb{E}\left[\mathbb{1}_{A_1}(X_1)\right] \ldots \mathbb{E}\left[\mathbb{1}_{A_n}(X_n)\right] \\ &= \mathbb{P}\left(X_1^{-1}(A_1)\right) \ldots \mathbb{P}\left(X_n^{-1}(A_n)\right) \\ &= \mu_{X_1}(A_1) \ldots \mu_{X_n}(A_n).\end{aligned}$$

This proves the "if" part.

Conversely, if $X_1, \dots, X_n$ are independent then, by the same computation as above, we get that (4.4) holds true for $f_1, \dots, f_n$ being indicator functions. By linearity this remains true for $f_1, \dots, f_n$ being simple functions. By Lebesgue's Theorem, Equation (4.4) holds true for any bounded Borel functions $f_1, \dots, f_n$ , approximating them by simple functions. □

**Definition 4.32.** We are given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A family $(\mathcal{F}_n)$ of sub-$\sigma$-fields of $\mathcal{F}$ is *independent* if

$$\mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_n}\right) = \prod_{k=1}^{n} \mathbb{P}(A_{i_k})$$

for all $\{i_1, \dots, i_n\} \subset \mathbb{N}$ and all $A_{i_k} \in \mathcal{F}_{i_k}$, respectively.

**Lemma 4.33.** *Let $(\mathcal{A}_n)$ be $\pi$-systems which generate the $\sigma$-fields $(\mathcal{F}_n)$ respectively. Then the family $(\mathcal{F}_n)$ is independent if and only if*

$$\mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_n}\right) = \prod_{k=1}^{n} \mathbb{P}(A_{i_k}) \tag{4.5}$$

*for all $\{i_1, \dots, i_n\} \subset \mathbb{N}$ and all $A_{i_k} \in \mathcal{A}_{i_k}$, respectively.*

*Proof.* There is only one direction to prove. Assume that (4.5) holds on the $\mathcal{A}_n$'s then consider the following measures on $\mathcal{F}_{i_n}$:

$$A \mapsto \mathbb{P}\left(A_{i_1} \cap \ldots \cap A_{i_{n-1}} \cap A\right)$$

and

$$A \mapsto \prod_{k=1}^{n-1} \mathbb{P}(A_{i_k}) \times \mathbb{P}(A)\,.$$

They coincide on $\mathcal{A}_{i_n}$, hence they coincide on $\mathcal{F}_{i_n}$ by Corollary 4.5. Repeating this procedure while exchanging the roles of $i_1, \dots, i_n$ gives (4.5) easily. □

**Definition 4.34.** In the following, if $X_1, \dots, X_n$ are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, we denote by $\sigma(X_1, \dots, X_n)$ the $\sigma$-field generated by $X_1, \dots, X_n$.

**Proposition 4.35.** *If $(X_1, \dots, X_n)$ are independent random variables then, for any $k \in \{1, \dots, n-1\}$, the $\sigma$-field $\sigma(X_1, \dots, X_k)$ is independent of $\sigma(X_{k+1}, \dots, X_n)$.*

*Proof.* By hypothesis we have, for all $A_1, \dots, A_n \in \mathrm{Bor}(\mathbb{R})$

$$\begin{aligned}\mathbb{P}\left(X_1 \in A_1, \dots, X_n \in A_n\right) &= \\ &= \mathbb{P}(X_1 \in A_1, \dots, X_k \in A_k)\,\mathbb{P}\left(X_{k+1} \in A_{k+1}, \dots, X_n \in A_n\right)\,.\end{aligned}$$

The set of events of the form $X_1^{-1}(A_1) \cap \ldots \cap X_k^{-1}(A_k)$ is a $\pi$-system which generates $\sigma(X_1, \ldots, X_k)$. The set of events $X_{k+1}^{-1}(A_{k+1}) \cap \ldots \cap X_n^{-1}(A_n)$ is a $\pi$-system which generates $\sigma(X_{k+1}, \ldots, X_n)$. We can apply Lemma 4.33 to these two $\sigma$-fields. $\square$

Be aware that if a random variable $Z$ is independent of a random variable $X$ and also of a random variable $Y$, then $\sigma(Z)$ is not independent of $\sigma(X, Y)$ in general. The proposition above needs the triple $(X, Y, Z)$ to be independent, which is a stronger condition. As a counter-example (and as an exercise) take $X$ and $Y$ to be two independent Bernoulli random variables (cf next section) with parameter $p = 1/2$ and take $Z = XY$.

### *4.2.5 Different Types of Convergence*

In Probability Theory we make use of many modes of convergence for random variables. Let us list them here.

**Definition 4.36.** Let $(X_n)$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $X$ be a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

We say that $(X_n)$ *converges almost surely* to $X$ if $(X_n(\omega))$ converges to $X(\omega)$ for almost all $\omega \in \Omega$.

For $1 \leq p < \infty$, we say that $(X_n)$ *converges in* $L^p$ to $X$ if the random variables $|X_n|^p$ and $|X|^p$ are all integrable and if $\mathbb{E}[\,|X_n - X|^p\,]$ converges to 0.

We say that $(X_n)$ converges to $X$ in *probability* if, for all $\varepsilon > 0$, the quantity $\mathbb{P}(|X_n - X| > \varepsilon)$ tends to 0 when $n$ tends to $+\infty$.

We say that $(X_n)$ converges to $X$ in *law* if, for all $A \in \mathrm{Bor}(\mathbb{R})$, the quantity $\mathbb{P}(X_n \in A)$ converges to $\mathbb{P}(X \in A)$.

Here is a list of results to be known about these convergences. We do not prove them and we convey the interested reader to reference books for the proofs (see a list in the Notes).

**Theorem 4.37.**

*1) The almost sure convergence implies the convergence in probability.*

*2) The convergence in probability implies the existence of a subsequence which converges almost surely.*

*3) The $L^p$-convergence implies the convergence in probability.*

*4) The convergence in probability implies the convergence in law.*

*5) The convergence in law is equivalent to the pointwise convergence of the characteristic function.*

### *4.2.6 Uniform Integrability*

We concentrate here on a notion which is very important, for it has many applications in more advanced problems concerning stochastic processes. It is a notion which relates the almost sure convergence and the $L^1$-convergence for sequences of random variables.

**Definition 4.38.** Let $\mathcal{U} = \{X_i\,;\ i \in I\}$ be any family of integrable random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, one can think of $\mathcal{U}$ as a subset of $L^1(\Omega, \mathcal{F}, \mathbb{P})$. The family $\mathcal{U}$ is *uniformly integrable* if

$$\lim_{a\to+\infty} \sup_{X\in\mathcal{U}} \mathbb{E}\left[|X|\ \mathbb{1}_{|X|\geq a}\right] = 0\,. \tag{4.6}$$

**Proposition 4.39.** *Let $\mathcal{U}$ be a subset of $L^1(\Omega, \mathcal{F}, \mathbb{P})$. The following assertions are equivalent.*

*i) The family $\mathcal{U}$ is uniformly integrable.*

*ii) One has*

*a)* $\sup_{X\in\mathcal{U}} \mathbb{E}\left[|X|\right] < \infty$

*and*

*b) for all $\varepsilon > 0$, there exists $\delta > 0$ such that $A \in \mathcal{F}$ and $\mathbb{P}(A) \leq \delta$ imply $\mathbb{E}\left[|X|\ \mathbb{1}_A\right] \leq \varepsilon$, for all $X \in \mathcal{U}$.*

*Proof.* If $\mathcal{U}$ is uniformly integrable and $A \in \mathcal{F}$ then

$$\begin{aligned}\sup_{X\in\mathcal{U}} \mathbb{E}\left[|X|\ \mathbb{1}_A\right] &= \sup_{X\in\mathcal{U}} \mathbb{E}\left[|X|\ \mathbb{1}_{|X|\geq a}\mathbb{1}_A\right] + \sup_{X\in\mathcal{U}} \mathbb{E}\left[|X|\ \mathbb{1}_{|X|<a}\mathbb{1}_A\right]\\ &\leq \sup_{X\in\mathcal{U}} \mathbb{E}\left[|X|\ \mathbb{1}_{|X|\geq a}\right] + a \sup_{X\in\mathcal{U}} \mathbb{E}\left[\mathbb{1}_{|X|<a}\mathbb{1}_A\right]\,.\end{aligned}$$

Let $c$ be such that the first term above is smaller than $\varepsilon/2$ for all $a \geq c$. We then have $\sup_{X\in\mathcal{U}} \mathbb{E}\left[|X|\ \mathbb{1}_A\right] \leq \varepsilon/2 + c\,\mathbb{P}(A)$. Taking $A = \Omega$ gives *a)*. Taking $\delta = \varepsilon/2c$ gives b). We have proved that *i)* implies *ii)*.

Conversely, assume that *ii)* is satisfied. Let $\varepsilon > 0$ and let $\delta > 0$ be as described in b). Let $c = \sup_{X\in\mathcal{U}} \mathbb{E}[|X|]/\delta < \infty$. Let $A$ be the event $(|X| \geq c)$. We have $\mathbb{P}(A) \leq \mathbb{E}[|X|]/c \leq \delta$. Thus $\mathbb{E}\left[|X|\ \mathbb{1}_A\right] \leq \varepsilon$ for all $X \in \mathcal{U}$. □

The main use of the notion of uniform integrability is the following result.

**Theorem 4.40.** *Let $(X_n)$ be a sequence of random variables belonging to $L^1(\Omega, \mathcal{F}, P)$. Suppose that $(X_n)$ converges almost surely to a random variable $X_\infty \in L^1(\Omega,\ \mathcal{F},\ P)$. Then $(X_n)$ converges to $X_\infty$ in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ if and only if the family $(X_n)$ is uniformly integrable.*

*Proof.* If $(X_n)$ converges to $X_\infty$ in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ we then have

$$\sup_n \mathbb{E}[|X_n|] \leq \sup_n \mathbb{E}[|X_n - X_\infty|] + \mathbb{E}[|X_\infty|] < +\infty$$

and

$$\mathbb{E}[|X_n|\,\mathbb{1}_A] \leq \mathbb{E}[|X_\infty|\,\mathbb{1}_A] + \mathbb{E}[|X_n - X_\infty|\,\mathbb{1}_A]\ .$$

The second term of the right hand side is dominated by $\varepsilon$ for $n$ large enough (independently of $A$); as any finite sequence of random variables is always uniformly integrable, the conclusion follows by Proposition 4.39.

Conversely, if $(X_n)$ is uniformly integrable, let $\varepsilon > 0$ and $\delta$ be such that if $a \geq \delta$ then $\sup_n \mathbb{E}[|X_n|\ \mathbb{1}_{|X_n|\geq a}] \leq \varepsilon/3$. Consider the function

$$\phi_a(x) = \begin{cases} 0\,, & \text{if } x \leq -a-\varepsilon\,, \\ -\frac{a}{\varepsilon}(x+a+\varepsilon)\,, & \text{if } x \in [-a-\varepsilon, -a]\,, \\ x\,, & \text{if } x \in [-a, a]\,, \\ -\frac{a}{\varepsilon}(x-a-\varepsilon)\,, & \text{if } x \in [a, a+\varepsilon]\,, \\ 0\,, & \text{if } x \geq a+\varepsilon\,. \end{cases}$$

By Fatou's lemma we have $\mathbb{E}[|X_\infty|\,\mathbb{1}_{|X_\infty|\geq a}] \leq \varepsilon/3$ and

$$\begin{aligned} \mathbb{E}\left[\,|X_n - X_\infty|\,\right] &\leq \mathbb{E}\left[\,|X_n - \phi_a(X_n)|\,\right] + \mathbb{E}\left[\,|\phi_a(X_n) - \phi_a(X_\infty)|\,\right] \\ &\qquad + \mathbb{E}\left[\,|\phi_a(X_\infty) - X_\infty|\,\right] \\ &\leq \mathbb{E}\left[\,|X_n|\,\mathbb{1}_{|X_n|\geq a}\,\right] + \mathbb{E}\left[\,|\phi_a(X_n) - \phi_a(X_\infty)|\,\right] \\ &\qquad + \mathbb{E}\left[\,|X_\infty|\,\mathbb{1}_{|X_\infty|\geq a}\,\right]\ . \end{aligned}$$

As $\mathbb{E}[|\phi_a(X_n) - \phi_a(X_\infty)|]$ tends to 0 by Lebesgue's theorem, this gives the result. □

## 4.3 Basic Laws

In this section we describe some basic examples of laws of random variables. They are the most common and useful probability distributions: Bernoulli, binomial, Poisson, exponential and Gaussian. They are the typical distributions that we meet in Quantum Mechanics, in Quantum Probability.

### *4.3.1 Bernoulli, Binomial*

**Definition 4.41.** The very first non-trivial probability laws are the *Bernoulli laws*. This term is characteristic of any probability law $\mu$ on $\mathbb{R}$ which is supported by the set $\{0, 1\}$. That is, probability measures $\mu$ such that $\mu(\{1\}) = p$ and $\mu(\{0\}) = q = 1 - p$ for some $p \in [0, 1]$. The measure $\mu$ is called the *Bernoulli law with parameter p*.

One often also calls "Bernoulli" any probability measure on $\mathbb{R}$ supported by two points. But generally, when these two points are not $\{0,1\}$, one should precise "Bernoulli random variable on $\{a,b\}$" etc.

**Definition 4.42.** If $X_1,\dots,X_n$ are $n$ independent Bernoulli random variables, with parameter $p$, and if we put $S_n = X_1 + \dots + X_n$ then we easily get the following probabilities

$$\mathbb{P}(S_n = i\,) = C_n^i\, p^i\, q^{n-i}\,,$$

for $i = 0,\dots,n$. The law of $S_n$ is called the *binomial distribution* with parameter $n$ and $p$, denoted $\mathcal{B}(n,p)$.

### *4.3.2 Exponential, Poisson*

Another fundamental law in Probability Theory is the *exponential law* and its naturally associated *Poisson law*. The exponential law is the "memory-lack law". Consider a positive random variable $T$ which is a waiting time, say. Assume that, knowing that a period of duration $s$ has passed without the expected event occurring, does not change the law of the remaining time to wait. In other words, the probability that $T > s+h$, knowing that $T > s$ already, should be the same as the probability that $T > h$. The first probability is

$$\mathbb{P}(T > s+h \,|\, T > s) = \frac{\mathbb{P}(T > s+h)}{\mathbb{P}(T > s)}$$

and thus our assumption is

$$\mathbb{P}(T > s+h) = \mathbb{P}(T > s)\,\mathbb{P}(T > h)$$

for all $s,h \geq 0$. The function $f(s) = \ln \mathbb{P}(T > s)$ satisfies $f(s+h) = f(s)+f(h)$ for all $s,h \in \mathbb{R}^+$ As it is right-continuous, increasing and $f(0) = 0$, it must be a linear function with positive slope. Thus the distribution of our waiting time, if non-trivial, shall be of the form

$$\mathbb{P}(T > t) = \mathbb{1}_{\mathbb{R}^+}(t)\,\lambda\,\mathrm{e}^{-\lambda t}$$

for some $\lambda > 0$. This is to say that the law of $T$ is the measure

$$\mathbb{1}_{\mathbb{R}^+}(t)\,\lambda\,\mathrm{e}^{-\lambda t}\,\mathrm{d}t\,.$$

**Definition 4.43.** This is the *exponential law* with parameter $\lambda$. We shall denote it by $\mathcal{E}(\lambda)$.

Now, consider a sequence of independent random variables $X_1, X_2, \dots$, each of which follows the exponential law with parameter $\lambda$. We think of this

sequence as successive intervals of time for the occurrences of some events (clients arriving, phone calls etc ...). Define the random variable

$$N = \sup\{n\,;\ X_1 + X_2 + \ldots + X_n \leq 1\}\,,$$

that is, the number of arrivals during the time interval $[0, 1]$.

**Proposition 4.44.** *For every $k \in \mathbb{N}$, we have*

$$\mathbb{P}(N = k) = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\,.$$

*Proof.* The probability $\mathbb{P}(N = k)$ is equal to

$$\mathbb{P}(X_1 + \ldots + X_k \leq 1 < X_1 + \ldots X_{k+1})\,.$$

By the independence of the $X_i$'s the law of $(X_1, \ldots, X_n)$ is the product measure of the laws of the $X_i$'s (Proposition 4.30). By the Transfer Theorem 4.17, the probability $\mathbb{P}(N = k)$ is equal to

$$\int_{x_1+\ldots+x_k\leq 1<x_1+\ldots+x_{k+1}} \mathbb{1}_{\mathbb{R}^+}(x_1, \ldots, x_n)\,\lambda^{k+1}\,\mathrm{e}^{-\lambda x_1}\ldots\mathrm{e}^{-\lambda x_{k+1}}\,\mathrm{d}x_1\,\ldots\,\mathrm{d}x_{k+1}\,.$$

With the change of variables $s_1 = x_1$, $s_2 = x_1 + x_2$, $\ldots$, $s_{k+1} = x_1 + \ldots + x_{k+1}$ this gives

$$\lambda^k \int_{0\leq s_k\leq 1<s_{k+1}} \lambda\,\mathrm{e}^{-\lambda s_{k+1}}\,\mathrm{d}s_1\,\ldots\,\mathrm{d}s_{k+1} = \frac{\lambda^k}{k!}\,\mathrm{e}^{-\lambda}\,. \qquad \square$$

**Definition 4.45.** The law we have obtained for $N$ above is called the *Poisson law* with parameter $\lambda$. We shall denote it by $\mathcal{P}(\lambda)$.

A similar computation shows that the random variable $N_t$ representing the number of arrivals during the interval $[0, t]$ follows a Poisson law with parameter $\lambda t$.

### *4.3.3 Gaussian laws*

**Definition 4.46.** The last fundamental law we shall describe here is the *Normal law*, or *Gaussian law*. A random variable is said to follow the normal law $\mathcal{N}(m, \sigma)$ if its probability law is of the form

$$\frac{1}{\sigma\sqrt{2\pi}}\,\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)\,\mathrm{d}x$$

for some $m \in \mathbb{R}$ and $\sigma > 0$. The expectation of such a random variable is $m$ and its variance is $\sigma^2$, as can be checked easily.

It is sometimes useful to consider the Dirac masses $\delta_m$ as being special cases of Gaussian laws, they correspond to the limit case $\sigma = 0$.

Note that, very often in the literature, the same law is denoted by $\mathcal{N}(m, \sigma^2)$, that is, the second parameter is the variance instead of the standard deviation. Here we shall follow our notation $\mathcal{N}(m, \sigma)$.

Here follow some useful results concerning Gaussian laws. We do not develop their proofs, for they are rather easy and standard.

**Proposition 4.47.**
*1) A Gaussian random variable $X$ with expectation $m$ and variance $\sigma^2$ has the following characteristic function*

$$\widehat{\mu}_X(t) = \mathrm{e}^{\mathrm{i}tm - \sigma^2 t^2/2}\,.$$

*2) The sum of independent Gaussian random variables is again a Gaussian random variable.*

*3) If a sequence of Gaussian random variables $(X_n)$ converges in $L^2$ to a random variable $X$, then $X$ is also Gaussian. The expectation and the variance of $X$ are the limits of the expectation and variance of $X_n$.*

A simple computation gives the moments of the Gaussian law.

**Proposition 4.48.** *If $X$ is a random variable with law $\mathcal{N}(0, \sigma)$, then*

$$\mathbb{E}[X^{2p}] = \frac{(2p)!}{2^p\, p!}\, \sigma^{2p}\,.$$

*The odd moments $\mathbb{E}[X^{2p+1}]$ all vanish.*

*Proof.* The odd moments vanish by a parity argument. For the even moments, this is a simple induction. Put

$$\alpha_{2p} = \mathbb{E}[X^{2p}] = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} x^{2p}\, \mathrm{e}^{-\frac{x^2}{2\sigma^2}}\, \mathrm{d}x\,.$$

By a simple integration by part we get

$$\alpha_{2p} = (2p-1)\sigma^2\, \alpha_{2p-2}\,.$$

The result is then easily deduced by induction. □

### *4.3.4 Gaussian Vectors*

The Normal distributions are central in Probability Theory for they have many interesting properties and they are the laws to which one converges by the Central Limit Theorem. We do not wish (and do not need) to develop all these properties here; the only point we want to develop now is the multidimensional extension of the notion of Gaussian random variable.

There are many different ways of defining Gaussian vectors, they lead to many equivalent characterizations. In fact, the most practical approach is to start with a rather abstract definition. We shall come to more explicit formulas later on in this subsection.

**Definition 4.49.** A family $(X_1, \dots, X_n)$ of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ is a *Gaussian family*, or a *Gaussian vector*, if every real linear combination $\sum_{i=1}^n \alpha_i X_i$ is a Gaussian random variable in $\mathbb{R}$.

In particular this definition implies that each random variable $X_i$ has to be a Gaussian random variable, but this is not enough in general. Though, note that, by Proposition 4.47, any family $(X_1 \dots, X_n)$ of *independent* Gaussian random variables is a Gaussian family. This is far not the only case of Gaussian vectors, as we shall see along this subsection.

**Proposition 4.50.** *Let $X = (X_1, \dots, X_n)$ be a Gaussian vector in $\mathbb{R}^n$, with expectation vector $\mathbf{m}$ and covariance matrix $\mathsf{D}$. Then its characteristic function is*

$$\widehat{\mu}_X(\mathbf{t}) = \exp\left(\mathrm{i}\,\langle \mathbf{t}\,,\,\mathbf{m}\rangle - \langle \mathbf{t}\,,\,\mathsf{D}\,\mathbf{t}\rangle/2\right), \tag{4.7}$$

*for all $\mathbf{t} \in \mathbb{R}^n$.*

*In particular the law of a Gaussian vector is completely determined by its expectation vector and its covariance matrix.*

*Proof.* Let $\langle \mathbf{t}\,,\,X\rangle = \sum_{k=1}^n t_k X_k$ be any (real) linear combination of the $X_i$'s. By hypothesis it is a Gaussian random variable. Its expectation is $\sum_{k=1}^n t_k \mathbb{E}[X_k] = \langle \mathbf{t}\,,\,\mathbf{m}\rangle$ . Let us compute its variance. We have

$$\mathbb{E}\left[\langle \mathbf{t}\,,\,X\rangle^2\right] = \sum_{i,j=1}^n t_i t_j\, \mathbb{E}[X_i X_j]$$

and

$$\langle \mathbf{t}\,,\,\mathbf{m}\rangle^2 = \sum_{i,j=1}^n t_i t_j\, \mathbb{E}[X_i]\, \mathbb{E}[X_j]\,.$$

This gives

$$\begin{aligned}\mathrm{Var}\left(\langle \mathbf{t}\,,\,X\rangle\right) &= \mathbb{E}\left[\langle \mathbf{t}\,,\,X\rangle^2\right] - \langle \mathbf{t}\,,\,\mathbf{m}\rangle^2 \\ &= \sum_{i,j=1}^{n} t_i t_j \left(\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j]\right) \\ &= \langle \mathbf{t}\,,\,\mathsf{D}\,\mathbf{t}\rangle\ .\end{aligned}$$

By Proposition 4.47 the characteristic function of $\langle \mathbf{t}\,,\,X\rangle$ is

$$\phi(x) = \exp\left(ix\,\langle \mathbf{t}\,,\,\mathbf{m}\rangle - x^2\,\langle \mathbf{t}\,,\,\mathsf{D}\,\mathbf{t}\rangle/2\right)\ .$$

Specializing the above to $x = 1$ gives the result. □

Independence can be read very easily on Gaussian vectors.

**Theorem 4.51.** *A Gaussian family $(X_1, \ldots, X_n)$ is made of independent random variables if and only if its covariance matrix is diagonal.*

*Proof.* One direction is easy: if the random variables $X_i$ are independent then the covariances $\mathbb{E}\big[X_i\,X_j\big] - \mathbb{E}[X_i]\,\mathbb{E}[X_j]$ vanish for all $i \neq j$, by Proposition 4.31.

Conversely, if $\mathrm{Cov}(X)$ is diagonal then, by Proposition 4.50, the characteristic function of $X = (X_1, \ldots, X_n)$ factorizes into

$$\widehat{\mu}_X(\mathbf{t}) = \prod_{k=1}^{n} \exp\left(it_k\mathbb{E}[X_k] - t_k^2\,\mathrm{Var}(X_k)/2\right)\ .$$

It is the same characteristic function as the one of a family of independent Gaussian random variables with same respective expectations and variances. By uniqueness of the characteristic function (Proposition 4.25), we conclude. □

Gaussian vectors behave very well with respect to linear transforms.

**Proposition 4.52.** *Let $X$ be a Gaussian vector in $\mathbb{R}^n$, with expectation vector $\mathbf{m}$ and covariance matrix $\mathsf{D}$. Consider a linear application $\mathsf{A}$ from $\mathbb{R}^n$ to $\mathbb{R}^d$. Then the $d$-dimensional random variable $Y = \mathsf{A}\,X$ is also a Gaussian vector. Its expectation vector is $\mathsf{A}\,\mathbf{m}$ and its covariance matrix is $\mathsf{A}\,\mathsf{D}\,\mathsf{A}^*$.*

*Proof.* For all $\mathbf{t} \in \mathbb{R}^d$ the random variable $\langle \mathbf{t}\,,\,Y\rangle = \langle \mathbf{t}\,,\,\mathsf{A}\,X\rangle = \langle \mathsf{A}^*\,\mathbf{t}\,,\,X\rangle$ is Gaussian, for it is a linear combination of the coordinates of $X$. By Proposition 4.50 its expectation is $\langle \mathsf{A}^*\,\mathbf{t}\,,\,\mathbf{m}\rangle = \langle \mathbf{t}\,,\,\mathsf{A}\,\mathbf{m}\rangle$ and its variance is $\langle \mathsf{A}^*\,\mathbf{t}\,,\,\mathsf{D}\,\mathsf{A}^*\mathbf{t}\rangle = \langle \mathbf{t}\,,\,\mathsf{A}\,\mathsf{D}\,\mathsf{A}^*\,\mathbf{t}\rangle$. We now conclude easily with Proposition 4.50 again and the uniqueness of the characteristic function. □

**Corollary 4.53.** *For any vector $\mathbf{m} \in \mathbb{R}^n$ and any real positive $n \times n$-matrix $\mathsf{D}$, there exists a Gaussian vector $X$ in $\mathbb{R}^n$ with expectation $\mathbf{m}$ and covariance matrix $\mathsf{D}$.*

*Proof.* As $\mathsf{D}$ is a positive matrix it admits a square root $\mathsf{A}$, that is a positive matrix $\mathsf{A}$ such that $\mathsf{A}^2 = \mathsf{D}$. Let $U = (U_1, \dots, U_n)$ be a family of independent Gaussian random variables with law $\mathcal{N}(0,1)$ respectively. Then $U$ is a Gaussian vector, with expectation $\mathbf{0}$ and covariance matrix I. By Proposition 4.52, the random variable $X = \mathsf{A}\,U + \mathbf{m}$ has the required law. □

We end up this subsection with an important result which gives the explicit density of the laws of Gaussian vectors.

**Theorem 4.54.** *Let $X$ be a Gaussian vector in $\mathbb{R}^n$ with expectation vector $\mathbf{m}$ and covariance matrix $\mathsf{D}$. If $\operatorname{Det}\mathsf{D} \neq 0$ then the law of $X$ admits a density with respect to the Lebesgue measure of $\mathbb{R}^n$. This density is*

$$\frac{1}{\sqrt{(2\pi)^n \operatorname{Det}\mathsf{D}}} \exp\left(-\frac{1}{2}\left\langle (\mathbf{x}-\mathbf{m})\,,\, \mathsf{D}^{-1}(\mathbf{x}-\mathbf{m})\right\rangle\right).$$

*Proof.* Let $\mathsf{A}$ be the square root of $\mathsf{D}$, that is, the matrix $\mathsf{A}$ is positive and $\mathsf{A}^2 = \mathsf{D}$. In particular $\operatorname{Det}\mathsf{A} = \sqrt{\operatorname{Det}\mathsf{D}}$ and $\mathsf{A}$ is invertible too. Let $Y = (Y_1, \dots, Y_n)$ be a family of independent Gaussian random variables with law $\mathcal{N}(0,1)$ respectively. Then the gaussian vector $W = \mathbf{m} + \mathsf{A}\,Y$ has the same law as $X$, by Proposition 4.52. In particular, for any bounded Borel function $f$ on $\mathbb{R}^n$ we have, by the Transfer Theorem 4.17

$$\mathbb{E}[f(X)] = \mathbb{E}[f(\mathbf{m}+\mathsf{A}\,Y)] = \frac{1}{\sqrt{(2\pi)^n}} \int_{\mathbb{R}^n} f(\mathbf{m}+\mathsf{A}\,\mathbf{y}) \exp\left(-\frac{1}{2}\,\langle \mathbf{y}\,,\,\mathbf{y}\rangle\right) \mathrm{d}\mathbf{y}\,.$$

Consider the change of variable $\mathbf{x} = \mathbf{m} + \mathsf{A}\,\mathbf{y}$ in $\mathbb{R}^n$. We get

$$\begin{aligned}\mathbb{E}[f(X)] &= \\ &= \frac{1}{\sqrt{(2\pi)^n}} \int_{\mathbb{R}^n} f(\mathbf{x}) \exp\left(-\frac{1}{2}\,\left\langle \mathsf{A}^{-1}(\mathbf{x}-\mathbf{m})\,,\,\mathsf{A}^{-1}(\mathbf{x}-\mathbf{m})\right\rangle\right) \times \\ &\qquad\qquad \times \operatorname{Det}\mathsf{A}^{-1}\,\mathrm{d}\mathbf{x} \\ &= \frac{1}{\sqrt{(2\pi)^n \operatorname{Det}\mathsf{D}}} \int_{\mathbb{R}^n} f(\mathbf{x}) \exp\left(-\frac{1}{2}\,\left\langle (\mathbf{x}-\mathbf{m})\,,\,\mathsf{D}^{-1}(\mathbf{x}-\mathbf{m})\right\rangle\right) \mathrm{d}\mathbf{x}\,.\end{aligned}$$

It is easy to conclude now. □

## 4.4 Conditioning

In this section we present several really non-trivial probabilistic notions, which are all of main importance. They are all linked to the idea of conditioning, that is, to give the best estimate possible on the law of a random variable, given some information on another random variable.

### *4.4.1 Conditional Laws*

**Definition 4.55.** Consider two random variables $X$ and $Y$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in measurable spaces $(E, \mathcal{E})$ and $(F, \mathcal{F})$ respectively. Assume first that $E$ is at most countable. Then, for all $A \in \mathcal{F}$ and for all $i \in E$, consider the conditional probability

$$\mathbb{P}(i, A) = \mathbb{P}(Y \in A \,|\, X = i)\,.$$

Note that we assume that $\mathbb{P}(X = i) > 0$, for all $i$, for otherwise we reduce $E$ to the set of "true" values of $X$. Clearly $\mathbb{P}$ is a mapping from $E \times \mathcal{F}$ to $[0, 1]$ such that:

i) for each fixed $A \in \mathcal{F}$, the mapping $i \mapsto \mathbb{P}(i, A)$ is a function on $E$,

ii) for each fixed $i \in E$, the mapping $A \mapsto \mathbb{P}(i, A)$ is a probability measure on $(F, \mathcal{F})$.

The probability measure above is called the *conditional law of $Y$ knowing $X = i$*.

A mapping $\mathbb{P}$ from $E \times \mathcal{F}$ to $[0, 1]$ satisfying the conditions i) and ii) above is called a *transition kernel*. When it is associated to random variables $X$ and $Y$ as above, is also simply called the *conditional law of $Y$ knowing $X$*.

**Proposition 4.56.** *Let $X$ and $Y$ be two random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in $(E, \mathcal{E})$ and $(F, \mathcal{F})$ respectively, with $E$ being at most countable. Let $\mathbb{P}$ be a transition kernel on $E \times \mathcal{F}$. Then $\mathbb{P}$ is the conditional law of $Y$ knowing $X$ if and only if*

$$\mathbb{E}[f(X, Y)] = \int_E \int_F f(x, y)\, \mathbb{P}(x, \mathrm{d}y)\, \mu_X(\mathrm{d}x)\,, \tag{4.8}$$

*for all bounded measurable function $f$ from $E \times F$ to $\mathbb{R}$.*

*Proof.* If $\mathbb{P}$ is the transition kernel associated to the conditional law of $Y$ knowing $X$, then

$$\begin{aligned}
\mathbb{P}(X \in A, Y \in B) &= \sum_{i \in A} \mathbb{P}(X = i, Y \in B) \\
&= \sum_{i \in A} \mathbb{P}(Y \in B \,|\, X = i)\, \mathbb{P}(X = i) \\
&= \int_A \mathbb{P}(x, B)\, \mu_X(\mathrm{d}x) \\
&= \int_A \int_B \mathbb{P}(x, \mathrm{d}y)\, \mu_X(\mathrm{d}x)\,.
\end{aligned}$$

Hence Equation (4.8) is valid for $f$ of the form $f(x, y) = \mathbb{1}_A(x)\mathbb{1}_B(y)$. It is now easy to conclude, by an approximation argument, that Equation (4.8) is valid for all bounded $f$.

Conversely, assume Equation (4.8) is satisfied for all bounded $f$. Take $f(x,y) = \mathbb{1}_A(x)\mathbb{1}_B(y)$, then we have

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B \mathbb{P}(x, \mathrm{d}y)\, \mu_X(\mathrm{d}x)\,.$$

In particular,

$$\begin{aligned}\mathbb{P}(Y \in B \,|\, X = i) &= \frac{\mathbb{P}(X = i, Y \in B)}{\mathbb{P}(X = i)} \\ &= \frac{\int_B \mathbb{P}(i, \mathrm{d}y)\, \mu_X(\{i\})}{\mathbb{P}(X = i)} \\ &= \mathbb{P}(i, B)\,.\end{aligned}$$

This proves that $\mathbb{P}$ is the conditional law of $Y$ knowing $X$. □

In a general situation, where $X$ takes values in a non-discrete measurable set $(E, \mathcal{E})$, the problem of defining a conditional law is more delicate. One cannot directly consider conditional probabilities such as

$$\mathbb{P}(x, B) = \mathbb{P}(Y \in B \,|\, X = x)\,,$$

for the event $(X = x)$ is in general a null set for $\mu_X$.

The idea, in order to define the conditional law of $Y$ knowing $X$ in a general situation, is to keep Identity (4.8) as a general definition.

**Definition 4.57.** A *transition kernel* on $E \times \mathcal{F}$ is a mapping

$$\begin{aligned}\mathbb{P} : E \times \mathcal{F} &\to [0,1] \\ (x\,,\, A) &\mapsto \mathbb{P}(x, A)\end{aligned}$$

such that:

i) for each fixed $A \in \mathcal{F}$, the mapping $x \mapsto \mathbb{P}(x, A)$ is a measurable function on $E$,

ii) for each fixed $x \in E$, the mapping $A \mapsto \mathbb{P}(x, A)$ is a probability measure on $\mathcal{F}$.

We say that a transition kernel $\mathbb{P}$ is the *conditional law of $Y$ knowing $X$* if $\mathbb{P}$ satisfies

$$\mathbb{E}[f(X,Y)] = \int_E \int_F f(x,y)\, \mathbb{P}(x, \mathrm{d}y)\, \mu_X(\mathrm{d}x)\,, \tag{4.9}$$

for all bounded measurable function $f$ on $E \times F$.

The question of existence of such a general conditional law is much less trivial than in the discrete case. In full generality it is actually a very difficult theorem, which is a particular case of the so-called *measure desintegration*

property. We do not need to enter into the details of such a theory and we state without proof the main existence theorem (see [Vil06] for a complete proof). Note that the theorem we write here is not the one with the largest conditions possible, but it is simply stated and sufficient for our use.

**Theorem 4.58.** *Let $X, Y$ be random variables with values in Polish spaces $E$ and $F$, respectively, equipped with their Borel $\sigma$-field $\mathcal{E}$ and $\mathcal{F}$, respectively. Then there exists a transition kernel $\mathbb{P}$ on $E \times \mathcal{F}$, unique $\mu_X$-almost surely, such that*

$$\mathbb{E}[f(X,Y)] = \int_E \int_F f(x,y)\, \mathbb{P}(x, \mathrm{d}y)\, \mu_X(\mathrm{d}x)$$

*for all bounded measurable function $f$ on $E \times F$.*

This theorem is much easier to prove when specializing to particular cases. We have seen that $X$ being valued in a discrete set is one easy situation; here is another one.

**Theorem 4.59.** *Let $X, Y$ be random variables whose law $\mu_{(X,Y)}$ admits a density $h(x,y)$ with respect to the Lebesgue measure on $\mathbb{R}^2$. Then the law $\mu_X$ of $X$ admits a density $k(x)$ with respect to the Lebesgue measure on $\mathbb{R}$. Furthermore the conditional law of $Y$ knowing $X$ exists, it is given by the transition kernel*

$$\mathbb{P}(x, \mathrm{d}y) = p(x,y)\, \mathrm{d}y$$

*where*

$$p(x,y) = \begin{cases} \frac{h(x,y)}{k(x)} & \text{if } k(x) \neq 0, \\ 0 & \text{if } k(x) = 0\,. \end{cases}$$

*Proof.* The fact that the law of $X$ admits a density is very easy and left as an exercise, the density of $X$ is given by

$$k(x) = \int_{\mathbb{R}} h(x,y)\, \mathrm{d}y\,.$$

Note that $k(x) = 0$ implies $h(x,y) = 0$ for almost all $y$.

Define the function

$$p(x,y) = \begin{cases} h(x,y)/k(x) & \text{if } k(x) \neq 0, \\ 0 & \text{if } k(x) = 0\,. \end{cases}$$

Clearly $\mathbb{P}(x, \mathrm{d}y) = p(x,y)\, \mathrm{d}y$ defines a transition kernel $\mathbb{P}$.

By definition of the law $\mu_{(X,Y)}$ we have, for all bounded Borel function $f$ on $\mathbb{R}^2$

$$
\begin{aligned}
\mathbb{E}[f(X,Y)] &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\,\mu_{(X,Y)}(\mathrm{d}x,\mathrm{d}y) \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\,h(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\,h(x,y)\,\mathbb{1}_{k(x)\neq 0}\,\mathrm{d}x\,\mathrm{d}y + \\
&\qquad\qquad + \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\,h(x,y)\,\mathbb{1}_{k(x)=0}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\,\frac{h(x,y)}{k(x)}\,k(x)\,\mathbb{1}_{k(x)\neq 0}\,\mathrm{d}x\,\mathrm{d}y + 0 \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\,p(x,y)\,k(x)\,\mathbb{1}_{k(x)\neq 0}\,\mathrm{d}x\,\mathrm{d}y + \\
&\qquad\qquad + \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\,p(x,y)\,k(x)\,\mathbb{1}_{k(x)=0}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\,p(x,y)\,k(x)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} f(x,y)\,\mathbb{P}(x,\mathrm{d}y)\,\mu_X(\mathrm{d}x)\,.
\end{aligned}
$$

This proves that $\mathbb{P}(x,\mathrm{d}y) = p(x,y)\,dx$ is indeed the conditional law of $Y$ knowing $X$. □

### *4.4.2 Conditional Expectations*

The *conditional expectation* is an operation which has to do with conditioning, but it is more general than the conditional law. Conditional expectations are defined with respect to a $\sigma$-field, instead of a random variable. The idea is that a $\sigma$-field, in general a sub-$\sigma$-field of $\mathcal{F}$, represents some partial information that we have on the random phenomenon we study. The conditional expectation is then the best approximation one can have of a random variable, given the knowledge of this $\sigma$-field. We now turn to the definitions and properties of conditional expectations.

**Theorem 4.60.** *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. Let $\mathcal{G}\subset\mathcal{F}$ be a sub-$\sigma$-field of $\mathcal{F}$. Let $X$ be an integrable random variable on $(\Omega,\mathcal{F},\mathbb{P})$. Then there exists an integrable random variable $Y$ which is $\mathcal{G}$-measurable and which satisfies*

$$\mathbb{E}[\,X\mathbb{1}_A] = \mathbb{E}[\,Y\mathbb{1}_A] \quad \textit{for all} \quad A\in\mathcal{G}\,. \tag{4.10}$$

*Any other $\mathcal{G}$-measurable integrable random variable $Y'$ satisfying* (4.10) *is equal to $Y$ almost surely.*

*Proof.* Assume first that $X$ is square integrable, that is, $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. The space $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Let $Y$ be the orthogonal projection of $X$ onto $L^2(\Omega, \mathcal{G}, P)$. Then $Y$ belongs to $L^2(\Omega, \mathcal{G}, P)$ and satisfies, for all $Z \in L^2(\Omega, \mathcal{G}, P)$

$$\mathbb{E}[\,XZ\,] = \big\langle \overline{Z}\,,\, X\big\rangle_{L^2(\Omega,\mathcal{F},\mathbb{P})} = \big\langle \overline{Z}\,,\, Y\big\rangle_{L^2(\Omega,\mathcal{F},\mathbb{P})} = \mathbb{E}[\,YZ\,]\,. \tag{4.11}$$

In particular $\mathbb{E}[\,X\,\mathbb{1}_A] = \mathbb{E}[\,Y\mathbb{1}_A]$ for all $A \in \mathcal{G}$. This implies that $Y$ is real-valued (almost surely) for its integral against any set $A \in \mathcal{G}$ is real. This also implies that if $X$ is positive then so is $Y$ almost surely, for $Y$ has a positive integral on any set $A \in \mathcal{G}$.

Now, assume $X$ is only integrable. Its positive part $X^+$ is also integrable. For all $n$, the random variable $X^+ \wedge n$ is square integrable. Let $Y_n^+$ be associated to $X^+ \wedge n$ in the same way as above. Then, by the remark above on positivity, we have that $Y_n^+$ is positive and that the sequence $(Y_n^+)$ is increasing. The random variable $Y^+ = \lim_n Y_n^+$ is integrable, for $\mathbb{E}[\,Y_n^+] = \mathbb{E}[\,X^+ \wedge n] \leq \mathbb{E}[\,X^+]$ and by Fatou's Lemma. In the same way, associate $Y^-$ to $X^-$. The random variable $Y = Y^+ - Y^-$ answers our statements.

Note that if $X$ is positive then so is $Y$, almost surely.

Uniqueness is easy for the relation $\mathbb{E}[(Y - Y')\mathbb{1}_A] = 0$ for all $A \in \mathcal{G}$ implies that $Y - Y' = 0$ almost surely. □

**Definition 4.61.** The almost sure equivalence class of integrable, $\mathcal{G}$-measurable, random variables $Y$ such that (4.10) holds is called the *conditional expectation* of $X$ with respect to $\mathcal{G}$. It is denoted by $\mathbb{E}[\,X \,|\, \mathcal{G}]$. We may also denote by $\mathbb{E}[\,X \,|\, \mathcal{G}]$ a representative of the equivalence class $\mathbb{E}[\,X \,|\, \mathcal{G}]$.

Here are the main properties of conditional expectations.

**Theorem 4.62.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$ and $X$ be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.*

*1) The mapping $X \mapsto \mathbb{E}[\,X \,|\, \mathcal{G}]$ is linear.*

*2) We have*

$$\mathbb{E}\,[\,\mathbb{E}[\,X \,|\, \mathcal{G}]\,] = \mathbb{E}[X]\,. \tag{4.12}$$

*3) If $X$ is positive then so is $\mathbb{E}[\,X \,|\, \mathcal{G}]$.*

*4)* (Monotone Convergence) *If $(X_n)$ is an increasing sequence of positive random variables* a.s. *converging to an integrable random variable $X$ then $\mathbb{E}[\,X_n \,|\, \mathcal{G}]$ converges* a.s. *to $\mathbb{E}[\,X \,|\, \mathcal{G}]$.*

*5)* (Fatou's Lemma) *If $(X_n)$ is a sequence of positive random variables then*

$$\mathbb{E}\left[\,\liminf_n\, X_n \,\middle|\, \mathcal{G}\,\right] \leq \liminf_n\, \mathbb{E}\,[\,X_n \,|\, \mathcal{G}\,]\,.$$

*6)* (Dominated Convergence) *If $(X_n)$ is a sequence of integrable random variables* a.s. *converging to an integrable random variable $X$ and such that*

*$|X_n| \le G$ for all $n$, where $G$ is an integrable random variable, then $\mathbb{E}[\,X_n \,|\, \mathcal{G}]$ converges* a.s. *to $\mathbb{E}[\,X \,|\, \mathcal{G}]$ .*

*7)* (Jensen's Inequality) *If $f$ is a convex function on $\mathbb{R}$ and if $f(X)$ is integrable then*

$$f\,(\,\mathbb{E}[\,X \,|\, \mathcal{G}]\,) \le \mathbb{E}[\,f(X) \,|\, \mathcal{G}] \qquad \text{a.s.} \tag{4.13}$$

*8) If $X$ is $\mathcal{G}$-measurable then $\mathbb{E}[\,X \,|\, \mathcal{G}] = X$* a.s. *The converse is true if $\mathcal{G}$ is complete.*

*9) If $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$ are $\sigma$-fields then*

$$\mathbb{E}\,[\,\mathbb{E}\,[\,X \mid \mathcal{G}_2] \mid \mathcal{G}_1] = \mathbb{E}[\,X \,|\, \mathcal{G}_1] \qquad \text{a.s.} \tag{4.14}$$

*10) If $Y$ is $\mathcal{G}$-measurable and $XY$ is integrable then*

$$\mathbb{E}[\,XY \,|\, \mathcal{G}] = Y\,\mathbb{E}[\,X \,|\, \mathcal{G}] \qquad \text{a.s.} \tag{4.15}$$

*11) If $X$ belongs to $L^p(\Omega, \mathcal{F}, \mathbb{P})$, for some $1 \le p \le \infty$ , then $\mathbb{E}[\,X \,|\, \mathcal{G}]$ belongs to $L^p(\Omega, \mathcal{G}, P)$ and*

$$\|\,\mathbb{E}[\,X \,|\, \mathcal{G}]\,\|_p \le \|X\|_p \ .$$

*Proof.* Property 1) is obvious. For property 2) take $A = \Omega$ in (4.10). Property 3) has been already proved.

4) Let $Y = \lim_n \mathbb{E}[\,X_n \,|\, \mathcal{G}]$ (this limit exists for the sequence $(\mathbb{E}[\,X_n \,|\, \mathcal{G}])$ is increasing). Then

$$\begin{aligned}\big|\,\mathbb{E}[\,Y\,\mathbb{1}_A] - \mathbb{E}[\,X\,\mathbb{1}_A]\,\big| &\le \big|\,\mathbb{E}[\,Y\,\mathbb{1}_A] - \mathbb{E}\,[\,\mathbb{E}[\,X_n \,|\, \mathcal{G}]\,\mathbb{1}_A\,]\,\big| + \\ &\qquad + \big|\mathbb{E}\,[\,\mathbb{E}[\,X_n \,|\, \mathcal{G}]\,\mathbb{1}_A\,] - \mathbb{E}[\,X\,\mathbb{1}_A]\,\big| \\ &\le \mathbb{E}\,[\,|Y - \mathbb{E}[\,X_n \,|\, \mathcal{G}]|\,\mathbb{1}_A] + \mathbb{E}\,[\,|X_n - X|\,\mathbb{1}_A]\ .\end{aligned}$$

Both terms in the right hand side tend to 0 as $n$ goes to $+\infty$. This gives the result.

5) For all $n \in \mathbb{N}$, put $Y_n = \inf_{p \ge n} X_p$. Then $(Y_n)$ is an increasing sequence, converging to $\liminf_n X_n$. For all $p \ge n$ we have $Y_n \le X_p$. Taking the conditional expectation with respect to $\mathcal{G}$, we get

$$\mathbb{E}[\,Y_n \,|\, \mathcal{G}] \le \mathbb{E}[\,X_p \,|\, \mathcal{G}]\,.$$

In particular

$$\mathbb{E}[\,Y_n \,|\, \mathcal{G}] \le \inf_{p \ge n}\ \mathbb{E}[\,X_p \,|\, \mathcal{G}] \le \liminf_n\ \mathbb{E}[\,X_n \,|\, \mathcal{G}]\,.$$

Now, applying the Monotone Convergence Property 4), we get

$$\mathbb{E}[\liminf_n\ X_n \,|\, \mathcal{G}] = \lim_n\ \mathbb{E}[\,Y_n \,|\, \mathcal{G}] \le \liminf_n\ \mathbb{E}[\,X_n \,|\, \mathcal{G}]\,.$$

6) Consider the sequence $(G - X_n)$. It is positive and it converges to $G - X$. By Fatou's Lemma 5) we have

$$\mathbb{E}[\, G - X \,|\, \mathcal{G}] = \mathbb{E}[\liminf_n (G - X_n) \,|\, \mathcal{G}] \leq \liminf_n \mathbb{E}[\, G - X_n \,|\, \mathcal{G}]\,,$$

that is,

$$\mathbb{E}[\, X \,|\, \mathcal{G}] \geq \limsup_n \mathbb{E}[\, X_n \,|\, \mathcal{G}]\,.$$

Applying the same argument to the positive sequence $(G + X_n)$, we get

$$\mathbb{E}[\, G + X \,|\, \mathcal{G}] = \mathbb{E}[\liminf_n (G + X_n) \,|\, \mathcal{G}] \leq \liminf_n \mathbb{E}[\, G + X_n \,|\, \mathcal{G}]\,,$$

this is to say,

$$\mathbb{E}[\, X \,|\, \mathcal{G}] \leq \liminf_n \mathbb{E}[\, X_n \,|\, \mathcal{G}]\,.$$

This proves the convergence of $\mathbb{E}[\, X_n \,|\, \mathcal{G}]$ to $\mathbb{E}[\, X \,|\, \mathcal{G}]$.

7) The fact that $f$ is convex is equivalent to

$$\frac{f(t) - f(s)}{t - s} \leq \frac{f(u) - f(t)}{u - t}$$

for all $s < t < u$. Put $T = \mathbb{E}\,[X \,|\, \mathcal{G}]$ and put

$$B = \sup \left\{ \frac{f(T) - f(s)}{T - s}\,;\, s < T \right\}.$$

We have that

$$B \leq \frac{f(u) - f(T)}{u - T}$$

for all $u > T$. Altogether, for all $s$ we have

$$f(s) - f(T) \geq B(s - T)\,.$$

Applying this inequality to $s = X$ and taking the conditional expectation with respect to $\mathcal{G}$ gives

$$\mathbb{E}\,[f(X) \,|\, \mathcal{G}] - f\,(\mathbb{E}\,[X \,|\, \mathcal{G}]) \geq 0\,,$$

for $f(\mathbb{E}[X \,|\, \mathcal{G}])$ is $\mathcal{G}$-measurable.

8) If $X$ is already $\mathcal{G}$-measurable, then, obviously from the definition, we have $\mathbb{E}[\, X \,|\, \mathcal{G}] = X$ a.s..

Conversely, if $\mathbb{E}[\, X \,|\, \mathcal{G}] = X$ a.s. and $\mathcal{G}$ is complete, then $X$ is a.s. equal to a $\mathcal{G}$-measurable random variable, hence it is $\mathcal{G}$-measurable.

9) Take $A_1 \in \mathcal{G}_1$. Then $A_1$ belongs to $\mathcal{G}_2$ too and

$$\mathbb{E}[\, \mathbb{E}[\, X \,|\, \mathcal{G}_2]\, \mathbb{1}_{A_1}] = \mathbb{E}[\, X\, \mathbb{1}_{A_1}] = \mathbb{E}[\, \mathbb{E}[\, X \,|\, \mathcal{G}_1]\, \mathbb{1}_{A_1}]\,.$$

This proves the result.

10) For all $A, B \in \mathcal{G}$ we have

$$\mathbb{E}[X\, \mathbb{1}_B \mathbb{1}_A] = \mathbb{E}[\mathbb{1}_B\, \mathbb{E}[X \,|\, \mathcal{G}]\, \mathbb{1}_A]\,.$$

Hence $\mathbb{E}[X \mathbb{1}_B \,|\, \mathcal{G}] = \mathbb{1}_B\, \mathbb{E}[X \,|\, \mathcal{G}]$. The result therefore holds true for $Y$ being a simple random variable. For a general $Y$ one takes a monotone limit of simple random variables.

11) This is Jensen inequality applied to the convex functions $f = \|\cdot\|_p$. □

We end up this subsection on conditional expectations with a useful result concerning conditional expectations and uniform integrability.

**Proposition 4.63.** *Let $X$ be an integrable random variable. The set of random variables $\mathbb{E}[X \,|\, \mathcal{R}]$, where $\mathcal{R}$ runs over all the sub-$\sigma$-fields of $\mathcal{F}$, is a uniformly integrable family.*

*Proof.* Let $\mathcal{R}$ be a sub-$\sigma$-field of $\mathcal{F}$. Put $Y = \mathbb{E}[X \,|\, \mathcal{R}]$. We have

$$\mathbb{E}[|Y|\ \mathbb{1}_{|Y|\geq a}] \leq \mathbb{E}\left[\mathbb{E}[|X|\ |\, \mathcal{R}]\, \mathbb{1}_{|Y|\geq a}\right] = \mathbb{E}\left[|X|\ \mathbb{1}_{|Y|\geq a}\right]\,.$$

Furthermore we have

$$\mathbb{P}(|Y| \geq a) \leq a^{-1}\mathbb{E}[|Y|] \leq a^{-1}\mathbb{E}[|X|]\,.$$

The family reduced to the random variable $|X|$ alone is uniformly integrable, hence by Proposition 4.39, for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $\mathbb{P}(A) \leq \delta$ implies $\mathbb{E}[|X|\ \mathbb{1}_A] \leq \varepsilon$.

One can choose $a$ large enough such that $a^{-1}\mathbb{E}[|X|] \leq \delta$. For this $a$ we have $\mathbb{E}\left[|Y|\ \mathbb{1}_{|Y|\geq a}\right] \leq \varepsilon$, independently of $\mathcal{R}$. This proves the proposition. □

### *4.4.3 Conditional Expectations and Conditional Laws*

We shall now present some formulas connecting conditional expectations to conditional laws.

First of all note that transition kernels act naturally on bounded functions. The proof of the following is easy and left to the reader.

**Proposition 4.64.** *Let $\mathbb{P}(x, \mathrm{d}y)$ be a transition kernel on $E \times \mathcal{F}$. Let $f$ be a bounded (resp. positive) measurable function on $F$. Then the function*

$$\mathsf{P}f(x) = \int_F f(y)\, \mathbb{P}(x, \mathrm{d}y)$$

*is also bounded (*resp. *positive) and measurable on $E$.*

In some sense, we have

$$\mathsf{P}f(x) = \mathbb{E}[f(Y) \,|\, X = x]\,. \tag{4.16}$$

This has no direct rigorous meaning, for $(X = x)$ may be an event with null probability. There are many ways to understand this intuitive relation more rigorously. The following result is one of them.

**Proposition 4.65.** *For every bounded measurable function $f$ on $F$ we have*

$$\mathbb{E}[f(Y) \,|\, \sigma(X)] = \mathsf{P}f(X)\,.$$

*Proof.* For all bounded function $g$ we have

$$\mathbb{E}[\, g(X)f(Y)] = \mathbb{E}\,[\,\mathbb{E}[\, f(Y) \,|\, \sigma(X)]\, g(X)]$$

by Theorem 4.62. But the left hand side is equal to

$$\int_E \int_F g(x)f(y)\, \mu_{(X,Y)}(\mathrm{d}x, \mathrm{d}y) = \int_E \int_F g(x)f(y)\, \mathbb{P}(x, \mathrm{d}y)\, \mu_X(\mathrm{d}x)$$

by definition of the conditional law. Consider the (measurable and bounded) function

$$h(x) = \int_F f(y)\, \mathbb{P}(x, dy) = \mathsf{P}f(x)\,.$$

Then

$$\int_E \int_F g(x)\, f(y)\, \mathbb{P}(x, \mathrm{d}y)\, \mu_X(\mathrm{d}x) = \int_E g(x)\, h(x)\, \mu_X(\mathrm{d}x) = \mathbb{E}[\, g(X)\, h(X)]\,.$$

We have established the identity

$$\mathbb{E}[\, g(X)\, f(Y)] = \mathbb{E}[\, g(X)\, h(X)]$$

which shows that

$$\mathbb{E}[f(Y) \,|\, \sigma(X)] = h(X)$$

almost surely (take $g(X) = \mathbb{1}_A(X)$). □

Transition kernels also act naturally on probability measures. The following is trivial and left to the reader.

**Proposition 4.66.** *Let $\mathbb{P}$ be a transition kernel on $E \times \mathcal{F}$ and $\mu$ a probability measure on $(E, \mathcal{E})$. Then the mapping $\mu\mathbb{P}$, from $\mathcal{F}$ to $[0, 1]$, defined by*

$$A \mapsto \mu\mathbb{P}(A) = \int_E \mathbb{P}(x, A)\, \mu(\mathrm{d}x)\,,$$

*is a probability measure on $(F, \mathcal{F})$.*

The connection with conditional laws is the following.

**Corollary 4.67.** *If $\mathbb{P}(x, dy)$ is the conditional law of $Y$ knowing $X$ and if $\mu_X$ is the law of $X$, then the law of $Y$ is $\mu_X\mathbb{P}$.*

*Proof.* By definition of the conditional law we have

$$\mathbb{E}[f(Y)] = \int_E \int_F f(y)\, \mathbb{P}(x, dy)\, \mu_X(\mathrm{d}x) = \int_F f(y)\, \mu_X\mathbb{P}(\mathrm{d}y)\,.$$

This clearly means that $\mu_X\mathbb{P}$ is the law of $Y$. □

**Definition 4.68.** We end up this section with some notations that are very often used in Probability Theory. The notation $\mathbb{P}(Y \in A \,|\, \mathcal{G})$ means $\mathbb{E}[\, \mathbb{1}_{(Y\in A)} \,|\, \mathcal{G}]$. We denote by $\mathbb{E}[\, Y \,|\, X]$ the conditional expectation $\mathbb{E}[\, Y \,|\, \sigma(X)]$.

## 4.5 Stochastic Processes

The main goal of Stochastic Process Theory is to study the behavior of families $(X_t)_{t\in I}$ of random variables $X_t$ indexed by a subset $I$ of $\mathbb{R}$ and defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In the sequel we shall mainly consider the cases where $I$ is $\mathbb{R}^+$ or an interval of $\mathbb{R}^+$. In the discrete time setup we are interested in $I = \mathbb{N}$. We shall give here a rigorous framework for the study of stochastic processes and their laws.

### *4.5.1 Products of measurable spaces*

**Definition 4.69.** Let $(E, \mathcal{E})$ be a measurable space and let $T$ be a set. On the set $E^T$ of all functions $\omega : T \to E$, we consider, for all $t \in T$, the *evaluation maps*

$$\begin{aligned} \Pi_t : E^T &\to E \\ \omega &\mapsto \omega(t)\,. \end{aligned}$$

We denote by $\mathcal{E}^T$ the $\sigma$-field generated by the evaluation maps $\{\Pi_t\,;\ t \in T\}$.

For any non-empty $S \subset T$ we denote by $\Pi_S$ the *restriction map*

$$\begin{aligned} \Pi_S : E^T &\to E^S \\ \omega &\mapsto \omega|_S\,. \end{aligned}$$

In the same way, for $U \subset V \subset T$ we denote by $\Pi_U^V$ the restriction map from $E^V$ onto $E^U$.

A subset $F$ of $E^T$ is a *finite cylinder* if it is of the form $F = \Pi_S^{-1}(A) = A \times E^{T\setminus S}$ for some finite $S \subset T$ and some $A \in E^S$. It is a *$\sigma$-cylinder* if it is of the same form but with $S$ countable.

The following characterization is easy to prove and left to the reader.

**Proposition 4.70.** *The $\sigma$-field $\mathcal{E}^T$ is generated by the finite cylinders. It exactly coincides with the set of all $\sigma$-cylinders of $E^T$.*

### *4.5.2 Generalities*

**Definition 4.71.** Let $T$ be a subset of $\mathbb{R}$. A *process* (or *stochastic process*) *indexed by* $T$ is a family $(X_t)_{t\in T}$ of random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. When $T$ is $\mathbb{R}^+$ one simply says a *process* and $(X_t)_{t\in T}$ may be denoted by $X$ simply. The random variables $X_t$ may be valued in any measurable space $(E, \mathcal{E})$. If $E$ is not mentioned then the process $X$ is understood to be valued in $\mathbb{R}$.

**Definition 4.72.** Let $X = (X_t)_{t\in T}$ be a process valued in $(E, \mathcal{E})$. The *law* of $X$ is the measure $\mu_X$ on $(E^T, \mathcal{E}^T)$ given by

$$\mu_X = \mathbb{P} \circ X^{-1} \ ,$$

the image of $\mathbb{P}$ under the mapping $X : \omega \mapsto (X_t(\omega))_{t\in T}$ from $\Omega$ to $E^T$.

The process $X$ has finite dimensional distributions. Indeed, for each finite subset $S = \{t_1, \dots, t_n\}$ of $T$ one denotes by $\mu_{X,S}$ the law of the $n$-tuple $(X_{t_1}, \dots, X_{t_n})$ on $(E^S, \mathcal{E}^S)$. That is,

$$\mu_{X,S} = \mu_X \circ \Pi_S^{-1} = \mathbb{P} \circ (\Pi_S \circ X)^{-1} \ .$$

Clearly, the measures $\mu_{X,S}$ satisfy the *compatibility condition*

$$\mu_{X,U} = \mu_{X,V} \circ \left(\Pi_U^V\right)^{-1}$$

for all finite $U \subset V \subset T$.

**Proposition 4.73.** *The law $\mu_X$ of a process $X = (X_t)_{t\in T}$ is determined by its finite dimensional marginals*

$$\{\mu_{X,S} \,;\ S \subset T \,,\ S \text{ finite}\} \ .$$

*Proof.* The restriction of $\mu_X$ to $\mathcal{E}^S$ is the measure $\mu_{X,S}$. Hence, the finite dimensional marginals determine the values of $\mu$ on finite cylinders. By Proposition 4.70 the finite cylinders generate $\mathcal{E}^T$. The finite cylinders clearly form a $\pi$-system $\mathcal{C}$. By Corollary 4.5, the measure $\mu_X$ is determined by its values on $\mathcal{C}$. □

### 4.5.3 Daniell-Kolmogorov Theorem

More interesting is the converse property. When is a given family $(\mu_S)$ of finite dimensional distributions on $(E^S, \mathcal{E}^S)$, respectively, determining a measure on $(E^T, \mathcal{E}^T)$? This is an interesting question for in general it is difficult to describe a process by specifying its law $\mu_X$. It is much easier to describe the finite dimensional distributions of $X$ and pray for they determine a true process. The answer to this reciprocal is given by the famous *Daniell-Kolmogorov Theorem*, also often called *Kolmogorov Consistency Theorem*.

**Definition 4.74.** Let $T$ be a subset of $\mathbb{R}$. Suppose we are given for any finite subset $S$ of $T$ a probability measure $\mu_S$ on $\mathbb{R}^S$. The family

$$\{\mu^S\,;\ S \text{ finite subset of } T\}$$

is called *compatible* if, for all finite subsets $U, V$ of $T$ such that $U \subset V$, we have

$$\mu_U = \mu_V \circ \left(\Pi_U^V\right)^{-1}\ .$$

**Definition 4.75.** A measurable space $(E, \mathcal{E})$ is a *Lusin space* if it is homeomorphic to the Borel subset of a compact metrisable space $J$. Note that the spaces $\mathbb{R}^n$ are obviously Lusin spaces.

We can now give the famous theorem. We shall not develop the proof here, it is rather long. But the attentive reader would have easily guessed that it consists in an application of Carathéodory's Extension Theorem 4.7: the compatible measures $\mu_S$ determine an additive measure $\mu_0$ on the finite cylinder field $\mathcal{C}$, the only difficulty is to prove that $\mu_0$ is $\sigma$-additive on $\mathcal{C}$. We refer to [RW00], Theorem II.31.1 for a complete proof.

**Theorem 4.76 (**Daniell-Kolmogorov Theorem). *Let $(E, \mathcal{E})$ be a Lusin space together with its Borel $\sigma$-field. Let $T$ be a subset of $\mathbb{R}$. If for all finite subsets $S$ of $T$ there exists a probability measure $\mu_S$ on $(E^S, \mathcal{E}^S)$ such that the family $\{\mu_S\,;\ S$ finite subset of $T\}$ is compatible, then there exists a unique probability measure $\mu$ on $(E^T, \mathcal{E}^T)$ such that $\mu_S = \mu \circ \Pi_S^{-1}$, for all finite $S \subset T$.*

### 4.5.4 Canonical Version

**Definition 4.77.** If $(X_t)_{t\in I}$ and $(Y_t)_{t\in I}$ are two processes indexed by the same set $I$ and if $\mu_{X,T} = \mu_{Y,T}$ for all finite subsets $T$ of $I$, we say that $(X_t)_{t\in I}$ and $(Y_t)_{t\in I}$ are *versions* of each other.

As a consequence of Proposition 4.73 we get the following.

**Proposition 4.78.** *Two processes $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ are versions of each other if and only if they have the same law.*

**Definition 4.79.** Let $(X_t)_{t\in T}$ be a process on $(\Omega,\mathcal{F},\mathbb{P})$, valued in $(E,\mathcal{E})$. For all $t\in T$, consider the evaluation maps $\Pi_t$ from $E^T$ to $E$ (recall that $\Pi_t(\omega)=\omega(t)$). They define a process $\Pi=(\Pi_t)_{t\in T}$ on the new probability space $(\Omega',\mathcal{F}',\mathbb{P}')=(E^T,\mathcal{E}^T,\mu_X)$ and we have, for $S=\{t_1,\dots,t_n\}\subset T$

$$\begin{aligned}\mu_{\Pi,S}(A_{t_1}\times\cdots\times A_{t_n}) &= \mu_X\left(\Pi_{t_1}^{-1}(A_{t_1})\cap\cdots\cap\Pi_{t_n}^{-1}(A_{t_n})\right)\\ &= \mu_X\left(\{x\in\mathbb{R}^T\,;\,x(t_1)\in A_{t_1},\dots,x(t_n)\in A_{t_n}\}\right)\\ &= \mu_{X,S}(A_{t_1}\times\cdots\times A_{t_n})\,.\end{aligned}$$

Thus $(\Pi_t)_{t\in T}$ is a version of $(X_t)_{t\in T}$. We call it the *canonical version* of $(X_t)_{t\in T}$.

The probability space $(E^T,\mathcal{E}^T,\mu_X)$ is called the *canonical space* of $(X_t)_{t\in T}$.

### *4.5.5 Modifications*

From now on, unless otherwise stated, all the processes are indexed by $\mathbb{R}^+$.

**Definition 4.80.** Two processes $X$ and $Y$ defined on the same probability space are *modifications* of each other if, for all $t\in\mathbb{R}^+$, $X_t$ equals $Y_t$ almost surely.

They are *indistinguishable* if, for almost all $\omega\in\Omega$, $X_t(\omega)$ equals $Y_t(\omega)$ for all $t$.

Note the subtle difference between the two definitions. "Modification" and "indistinguishable" both mean that $Y_t=X_t$ for all $t$ and almost all $\omega$, but in the first case the null set of $\omega$ may depend on $t$, whereas in the second defintion it does not depend on $t$.

**Definition 4.81.** A subset $A$ of $\mathbb{R}^+\times\Omega$ is *evanescent* if there exists a negligible subset $B$ of $\Omega$ such that $A\subset\mathbb{R}^+\times B$.

From all these definitions one can check easily the following properties (left to the reader).

**Proposition 4.82.**

*1) If $X$ and $Y$ are indistinguishable then they are modifications of each other.*

*2) If $X$ and $Y$ are modifications of each other then they are versions of each other.*

*3) $X$ and $Y$ are indistinguishable if and only if the set of $(\omega,t)$ such that $X_t(\omega)\neq Y_t(\omega)$ is evanescent.*

**Definition 4.83.** Let $X$ be a process. For all $\omega \in \Omega$, the mapping $t \mapsto X_t(\omega)$ defines a function on $\mathbb{R}^+$. These functions are called the *paths* of the process $X$.

One says that the process $X$ has *continuous paths* (or simply, is *continuous*) if for almost all $\omega \in \Omega$ the path $X_\cdot(\omega)$ is continuous on $\mathbb{R}^+$.

In the same way one defines *right-continuous processes*, *left-continuous processes* (we take the convention that any process is left-continuous at 0).

A process is said to be *càdlàg*[1] if its paths are right-continuous and admit left limits at all points.

**Proposition 4.84.** *Let $X$ and $Y$ be two right-continuous (resp. left-continuous) processes. If they are modifications of each other then they are indistinguishable.*

*Proof.* There exists a negligible set $\mathcal{N}$ such that for all $\omega \notin \mathcal{N}$ the paths $X_\cdot(\omega)$ and $Y_\cdot(\omega)$ are right-continuous and $X_t(\omega) = Y_t(\omega)$ for all $t$ rational. Thus passing to the limit we have $X_t(\omega) = Y_t(\omega)$ for all $t$. □

**Definition 4.85.** A process $X$ is *measurable* if it is measurable as a mapping from $\mathbb{R}^+ \times \Omega$ to $\mathbb{R}$, where $\mathbb{R}^+ \times \Omega$ is equipped with the $\sigma$-algebra $\mathbb{R}(\mathbb{R}^+) \otimes \mathcal{F}$.

### *4.5.6 Regularization*

At that stage of the lecture we have to discuss a rather fine point about the definition of the canonical version of a stochastic process. This section can be avoided at first reading. We do not give all details.

There is quite an important problem attached to the way the canonical space of a process $X$ has been constructed. Indeed, the canonical space furnished by Kolmogorov's Theorem 4.76 is not good enough, for the $\sigma$-field $\mathcal{F} = \mathcal{E}^I$ associated to this construction is too poor. Recall that this $\sigma$-algebra is the one generated by finite support cylinders in $\Omega = E^I$. We have shown that the events of $\mathcal{E}$ are those which depend only on a countable set of coordinates of $\omega \in \Omega$. As a consequence, one cannot "ask" if the process we are interested in is continuous or not. Indeed, the set

$$\{\omega\,;\ t \mapsto X_t(\omega) \text{ is continuous } \}$$

has no reason in general to be in $\mathcal{F}$, for it cannot be expressed by a countable number of constraints of the $X_t$'s.

The same problems hold when one wants to study simple objects such as

[1] The name càdlàg comes from the french "continu à droite avec des limites à gauche" which means "right-continuous with left-limits" (the english version of càdlàg should then be "rcll". It has also been sometimes called "corlol" for "continuous on the right with limits on the left, but "càdlàg" is clearly much easier to pronounce!

$$\inf\{t; X_t(\omega) > 0\}\,, \quad \lim_{s\to t} X_s(\omega)\,, \quad \sup_{s\leq t} |X_s| \ ... \tag{4.17}$$

This clearly implies strong restrictions for whom wants to study the properties of a given stochastic process!

The way this problem is solved in general follows the following path. For most of the processes one is interested in (for example martingales, ...) we are able to prove that the process in question $X$ admits a modification $Y$ which is càdlàg (or sometimes better: continuous). Denote by $\mathcal{D}$ the set of càdlàg functions from $\mathbb{R}^+$ to $\mathbb{R}$. This is a subset of $E = \mathbb{R}^{\mathbb{R}^+}$ which is not measurable for the cylinder $\sigma$-field $\mathcal{F} = \mathrm{Bor}(\mathbb{R})^{\mathbb{R}^+}$.

**Lemma 4.86.** *If a process $X$ admits a càdlàg version then the measure $\mathbb{P}_X$ of any $\mathcal{F}$-measurable set $F$ such that $F \supset \mathcal{D}$ is equal to 1.*

In general measure theory, the property described above means that the *outer measure* of $\mathbb{P}$ is equal to 1 on $\mathcal{D}$.

*Proof.* Assume that on some probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ there exists a process $Y$ which is a version of $X$ and which has almost all its paths càdlàg.

Let $F$ be a $\mathcal{F}$-measurable set such that $F$ contains $\mathcal{D}$. The probability $\mathbb{P}_X(F)$ is the probability that a certain countable set of constraints on $X_t$, $t \in \mathbb{R}^+$, is satisfied. It is then equal to the probability $\mathbb{P}'(F')$ of the same set of constraints applied to $Y_t$, $t \in \mathbb{R}^+$. But the fact that $F$ contains $\mathcal{D}$ implies that $F'$ is the whole of $\Omega'$ (up to a null set maybe). Hence the probability is equal to 1. □

Let us give an example in order to make the proof above more clear. Let $F$ be for example the set

$$F = \{\omega \in \Omega; \lim_{\substack{s\to 0\\ s\in\mathbb{Q}}} \omega(s) = \omega(0)\}\,.$$

This set is $\mathcal{F}$-measurable, for it can be written as

$$F = \bigcap_{\substack{\epsilon>0\\ \epsilon\in\mathbb{Q}}} \bigcup_{\substack{\eta>0\\ \eta\in\mathbb{Q}}} \bigcap_{\substack{s>0\\ s\in\mathbb{Q}}} \{\omega \in \Omega; |\omega(s) - \omega(0)| \leq \epsilon\}\,.$$

Hence we have

$$\begin{aligned}
\mathbb{P}_X(F) &= \mathbb{P}_X\left(\bigcap_{\substack{\varepsilon>0\\ \varepsilon\in\mathbb{Q}}}\bigcup_{\substack{\eta>0\\ \eta\in\mathbb{Q}}}\bigcap_{\substack{s>0\\ s\in\mathbb{Q}}}\{\omega\in\Omega;|\omega(s)-\omega(0)|\le\varepsilon\}\right)\\
&= \mathbb{P}\left(\bigcap_{\substack{\varepsilon>0\\ \varepsilon\in\mathbb{Q}}}\bigcup_{\substack{\eta>0\\ \eta\in\mathbb{Q}}}\bigcap_{\substack{s>0\\ s\in\mathbb{Q}}}\{\omega\in\Omega;|X_s(\omega)-X_0(\omega)|\le\varepsilon\}\right)\\
&= \mathbb{P}'\left(\bigcap_{\substack{\varepsilon>0\\ \varepsilon\in\mathbb{Q}}}\bigcup_{\substack{\eta>0\\ \eta\in\mathbb{Q}}}\bigcap_{\substack{s>0\\ s\in\mathbb{Q}}}\{\omega'\in\Omega';|Y_s(\omega')-Y_0(\omega')|\le\varepsilon\}\right)\\
&= \mathbb{P}'(\{\omega'\in\Omega';\lim_{\substack{s\to 0\\ s\in\mathbb{Q}}}Y_s(\omega')=Y_0(\omega')\})\,.
\end{aligned}$$

But as the path of $Y$ are almost all càdlàg we have $\lim_{\substack{s\to 0\\ s\in\mathbb{Q}}} Y_s(\omega') = Y_0(\omega')$ to be satisfied by almost all $\omega'$. Hence the above probability is equal to 1.

When the situation $\mathbb{P}(F) = 1$ for all $F \in \mathcal{F}$, $F \supset \mathcal{D}$, occurs we have the following result, whose easy proof is left to the reader.

**Lemma 4.87.** *On the space $\mathcal{D}$ define the coordinate mappings $Y_t(\omega) = \omega(t)$. Then the $\sigma$-algebra $\mathcal{F}_\mathcal{D}$ generated by the $Y_t$, $t \in \mathbb{R}^+$ coincides with $\mathcal{F} \cap \mathcal{D}$. Furthermore, putting, for all $F \in \mathcal{F}_\mathcal{D}$*

$$\mathbb{Q}(F) = \mathbb{P}_X(\widehat{F})$$

*for any $\widehat{F} \in \mathcal{F}$ such that $F = \widehat{F} \cap \mathcal{D}$, defines a probability measure on $(\mathcal{D}, \mathcal{F}_\mathcal{D})$.*

Hence we have defined a version $(\mathcal{D}, \mathcal{F}_\mathcal{D}, \mathbb{Q}, Y)$ of $X$ whose paths are all càdlàg. On that probability space, elements such as in (4.17) are now all well-defined and measurable.

Actually the $\sigma$-algebra $\mathcal{F}_\mathcal{D}$ is much richer than was $\mathcal{F}$. One can for example prove the following. The space $\mathcal{D}$ admits a natural topology, called the *Skorohod topology*, that we shall not describe here. Let $\mathcal{C}$ be the space of continuous functions from $\mathbb{R}^+$ to $\mathbb{R}$. The Skorohod topology coincides on $\mathcal{C}$ with the topology of uniform convergence on compact sets. Both $\mathcal{D}$ and $\mathcal{C}$ are Polish spaces with these topologies.

**Theorem 4.88.** *On $\mathcal{D}$ (resp. $\mathcal{C}$) the $\sigma$-algebra $\mathcal{F}_\mathcal{D}$ (resp. $\mathcal{F}_\mathcal{C}$) coincide with the Borel $\sigma$-algebra* $\mathrm{Bor}(\mathcal{D})$ *(resp.* $\mathrm{Bor}(\mathcal{C})$*) associated to the Skorohod topology.*

## 4.6 Brownian Motion

In this section and in the next one, we shall consider two fundamental examples of stochastic processes: *Brownian motion* Brownian motion and the *Poisson process*. They are fundamental for they are cornerstones of classical Stochastic Calculus, but they also appear to be key processes in the probabilistic interpretations of Quantum Stochastic Calculus. We first start with the Brownian motion.

### *4.6.1 Construction*

**Definition 4.89.** A *Brownian motion* is a process $(W_t)_{t\in\mathbb{R}^+}$ with values in $\mathbb{R}$ such that:

i) for all $s<t$, the random variable $W_t - W_s$ is independent of the random variables $W_u$, $u\leq s$,

ii) for all $t\geq 0$, the random variable $W_t$ follows the Gaussian $\mathcal{N}(0,t)$ probability law,

iii) a.s. the paths $t\mapsto W_t$ are continuous.

The question of existence of such a process is not trivial, let us construct it.

We start with the following, which is an easy application of Daniell-Kolmogorov (Theorem 4.76).

**Theorem 4.90.** *For any given probability measure $\mu$ on $\mathbb{R}$, there exists a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a sequence $(X_n)$ of random variables on $\Omega$ such that the random variables $X_n$ are independent of each other and all have the same individual law $\mu$.* □

We can now prove the following.

**Theorem 4.91.** *Let $\mathcal{H}$ be a separable Hilbert space. There exists a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a family $X(h)$, $h\in\mathcal{H}$, of random variables on $\Omega$ such that*

*i) the mapping $h\longmapsto X(h)$ is linear,*

*ii) each random variable $X(h)$ follows the Gaussian law $\mathcal{N}(0,\|h\|)$.*

*Proof.* Let $(e_n)$ be an orthonormal basis of $\mathcal{H}$. By Theorem 4.90, there exists a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a sequence $(Z_n)$ of independent random variables on $\Omega$, with individual law $\mathcal{N}(0,1)$.

For any $h\in\mathcal{H}$, put $X(h)=\sum_n\langle e_n\,,\,h\rangle\,Z_n$. This series is convergent in $L^2(\Omega,\mathcal{F},\mathbb{P})$ and defines a random variable $X(h)$ on $\Omega$. The family $X(h)$, $h\in\mathcal{H}$, satisfies the assumptions of the theorem, as can be easily checked. □

Note that, as a consequence of the properties i) and ii) above, we have

$$\langle X(h')\,,\,X(h)\rangle_{L^2(\Omega,\mathcal{F},\mathbb{P})} = \langle h'\,,\,h\rangle\,,$$

as can be checked easily by computing $\mathbb{E}[X(h+h')^2]$ in two different ways.

Also note that the set $\{X(h),\ h\in\mathcal{H}\}$ is a *Gaussian subspace* of $L^2(\Omega,\mathcal{F},\mathbb{P})$, that is, any linear combination of the $X(h)$ is a Gaussian random variable. In other words, this means that every finite family $(X(h_1)\dots X(h_n))$ is Gaussian vector.

Now we can start the construction of a Brownian motion. Take $\mathcal{H} = L^2(\mathbb{R}^+)$ and construct a family $X(h)$, $h\in\mathcal{H}$, such as in Theorem 4.91. We claim that the process $W_t = X(\mathbb{1}_{[0,t]})$, $t\in\mathbb{R}^+$, is a Brownian motion. Indeed, the random variable $W_t$ follows the law $\mathcal{N}(0,t)$ by Property ii) of Theorem 4.91 and thus satisfies condition ii) in the definition of Brownian motion. Furthermore, we have, for all $u\le s\le t$

$$\langle W_t - W_s\,,\,W_u\rangle_{L^2(\Omega,\mathcal{F},P)} = \langle \mathbb{1}_{[s,t]}\,,\,\mathbb{1}_{[0,u]}\rangle_{\mathcal{H}} = 0\,.$$

But the pair $(W_t - W_s, W_u)$ is Gaussian, hence by Theorem 4.51 they are independent random variables. This gives the condition i) in the definition of a Brownian motion.

We still need to prove that the process we obtained is continuous, or at least can be modified into a continuous process.

### *4.6.2 Regularity of the Paths*

Actually, we shall establish a stronger property for the paths: the Hölder continuity of order $\alpha$ for every $\alpha < 1/2$. This property is based on a general criterion due to Kolmogorov that we state here without proof, for it is rather long and would take us too far here (see [RY99], Theorem 2.1).

**Theorem 4.92.** [Kolmogorov criterion] *Let $X$ be a process such that there exists strictly positive constants $\gamma$, $C$ and $\varepsilon$ such that*

$$\mathbb{E}[|X_t - X_s|^\gamma] \le C\,|t-s|^{1+\varepsilon}\,,$$

*for all $s,t\in\mathbb{R}^+$. Then there exists a modification of $X$ whose paths are Hölder continuous of order $\alpha$ for every $\alpha\in[0,\varepsilon/\gamma]$.* □

Once this is admited, the fact that the increments of the Brownian motion are Gaussian gives

$$\mathbb{E}\left[(W_t - W_s)^{2p}\right] = C_p\,|t-s|^p$$

for every integer $p>0$ (by Proposition 4.48). This immediately yields the following.

**Theorem 4.93.** *Up to modification, the paths of Brownian motion are locally Hölder continuous of order $\alpha$ for every $\alpha < 1/2$. In particular they are continuous.* □

We have proved that the process we have constructed admits a continuous modification. Up to the change of probability space described in Subsection 4.5.6, we have constructed a Brownian motion.

### *4.6.3 Quadratic Variations*

Another very important property of the Brownian motion is that it admits a non-trivial quadratic variation.

**Theorem 4.94.** *For any $t \in \mathbb{R}^+$, the quantity*

$$\sum_{i; t_i \leq t} \big(W_{t_{i+1}} - W_{t_i}\big)^2$$

*converges to $t$ in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, when the diameter $\delta$ of the partition $\{t_i\,;\, i \in \mathbb{N}\}$ tends to 0.*

*Proof.* Let us compute the $L^2$-norm of the difference (for simplicity, we assume that the partition ends at $t$):

$$\begin{aligned}\left\| \sum_{i;\, t_i \leq t} (W_{t_{i+1}} - W_{t_i})^2 - t \right\|^2 &= \mathbb{E}\left[ \left( \sum_{i;\, t_i \leq t} (W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \right)^2 \right] \\ &= \sum_{i;\, t_i \leq t} \mathbb{E}\left[ \big( (W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i) \big)^2 \right] ,\end{aligned}$$

where we used the independence of the increments and the fact that

$$\mathbb{E}[(W_{t_{i+1}} - W_{t_i})^2 - (t_{i+1} - t_i)] = 0\,.$$

Furthermore, for a Gaussian random variable $Y$ with mean 0, we have $\mathbb{E}[Y^4] = 3\mathbb{E}[Y^2]^2$. This gives

$$\left\| \sum_{i;\, t_i \leq t} (W_{t_{i+1}} - W_{t_i})^2 - t \right\|^2 = 2 \sum_{i;\, t_i \leq t} (t_{i+1} - t_i)^2$$

which converges to 0 with the diameter of the partition. □

It is actually possible to prove the same result for almost sure convergence, but we do not give a proof of this fact here.

This quadratic variation property has many important consequences that we shall not develop in this lecture, but one very important application of it is the following.

**Theorem 4.95.** *The paths of the Brownian motion are almost surely of infinite variation on any interval.*

*The paths of the Brownian motion are almost surely nowhere locally Hölder continuous of order $\alpha$ for $\alpha > 1/2$.*

*In particular, the paths of the Brownian motion are almost surely nowhere differentiable.*

*Proof.* Note that almost surely we have

$$\sum_{i;\, t_i \le t} \left(W_{t_{i+1}}(\omega) - W_{t_i}(\omega)\right)^2 \le$$
$$\le \left(\sup_{i;\, t_i \le t} \left|W_{t_{i+1}}(\omega) - W_{t_i}(\omega)\right|\right) \left(\sum_{i;\, t_i \le t} \left|W_{t_{i+1}}(\omega) - W_{t_i}(\omega)\right|\right).$$

The first term in the right hand side converges to 0 by the continuity of Brownian motion. The second term is dominated by the total variation of the Brownian path. As the left hand side converges to a finite quantity, this forces the total variation to be infinite.

The case of the non-Hölder property is treated following a similar idea: for all $\alpha > 1/2$ we have

$$\sum_{i;t_i \le t} (W_{t_{i+1}}(\omega) - W_{t_i}(\omega))^2 \le$$
$$\le t \left(\sup_{i;t_i \le t} |t_{i+1} - t_i|^{2\alpha - 1}\right) \left(\sup_{i;t_i \le t} \frac{|W_{t_{i+1}}(\omega) - W_{t_i}(\omega)|^2}{|t_{i+1} - t_i|^{2\alpha}}\right).$$

If the Brownian paths were Hölder of order $\alpha$ the last term above would be dominated independently of the partition. The rest of the right hand side converges to 0. This contradicts the fact that the left hand side converges to $t$. This proves the non Hölderian character of Brownian motion for $\alpha > 1/2$.

Non-differentiability is immediate now. □

We have not yet said if the Brownian paths are Hölder-continuous of order exactly 1/2 or not. It so happens that they are not, but this result needs further developments; we just mention it as a remark.

## 4.7 Poisson Processes

### *4.7.1 Definition*

We now get interested in our second fundamental stochastic process, the *Poisson process*.

**Definition 4.96.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $(T_n)$ be a strictly increasing sequence of positive random variables. The $T_n$'s are thought as arrival times. A process $X$ such that, for all $t \in \mathbb{R}^+$

$$X_t = \sum_n \mathbb{1}_{T_n \leq t}$$

is called a *counting process associated to* $(T_n)$. It is valued in $\mathbb{N} \cup \{+\infty\}$. If $\sup_n T_n = \infty$ a.s. one says that $X$ is a *non-exploding counting process*.

A *Poisson process* is a non-exploding counting process $N$ whose increments are independent and stationary. That is,

i) $N_t - N_s$ is independent of all the random variables $N_u$, $u \leq s$

ii) $N_t - N_s$ has the same law as $N_{t+h} - N_{s+h}$ for all $t \geq s \geq 0$ and $h \geq 0$.

### *4.7.2 Existence*

**Theorem 4.97.** *Poisson processes exist and they are all of the following form: there exists $\lambda \in \mathbb{R}^+$ such that*

$$\mathbb{P}(N_t = n) = \frac{(\lambda t)^n}{n!}\, e^{-\lambda t} \tag{4.18}$$

*for all $n \in \mathbb{N}$, all $t \in \mathbb{R}^+$. In other words, the associated sequence $(T_n)$ consists of sums of independent times each of which follows an exponential distribution with parameter $\lambda$.*

*Proof.* One direction is easy. If $(T_n)$ is the sequence given by $T_n = \sum_{m \leq n} S_m$ where the $S_m$'s are independent, identically distributed random variables following the exponential law with parameter $\lambda \geq 0$, we have already seen in Subsection 4.3.2 that the associated counting process

$$N_t = \sum_{n=0}^{\infty} \mathbb{1}_{T_n \leq t}$$

is a Poisson process and follows the Poisson law (4.18).

The converse is more difficult. From the hypothesis we have

$$\mathbb{P}(N_t = 0) = \mathbb{P}(N_s = 0)\,\mathbb{P}(N_t - N_s = 0) = \mathbb{P}(N_s = 0)\,\mathbb{P}(N_{t-s} = 0)$$

and thus

$$\mathbb{P}(N_t = 0) = e^{-\lambda t}$$

for some $\lambda \geq 0$ and for all $t \in \mathbb{R}^+$.

We now claim that $\mathbb{P}(N_t \geq 2) = o(t)$. Indeed, divide $[0,1]$ into $n$ intervals of the same length. Let $S_n$ be the number of subintervals which contain at least two times of the sequence $(T_m)$. Clearly $S_n$ has a binomial distribution $\mathcal{B}(n, \mathbb{P}(N_{1/n} \geq 2))$. Therefore $\mathbb{E}[S_n] = n\,\mathbb{P}(N_{1/n} \geq 2)$. For a fixed $\omega$, for $n$ sufficiently large there is no interval with more than one random time $T_i$. Thus $\lim_{n\to+\infty} S_n(\omega) = 0$ a.s. We now wish to apply the dominated convergence theorem in order to conclude that $\lim_{n\to+\infty} \mathbb{E}(S_n) = 0$ and hence the announced estimate. As we clearly have $S_n \leq N_1$, we just need to prove that $\mathbb{E}[N_1] < \infty$. The intervals $T_{n+1} - T_n$ between the jumps are independent random variables, with the same law: the one of $T_1$. Hence

$$\mathbb{E}[e^{-T_n}] = \mathbb{E}[e^{-T_1}]^n = \alpha^n.$$

This proves that

$$\mathbb{P}(|N_t| > n) \leq \mathbb{P}(T_n < t) \leq \frac{\mathbb{E}[e^{-T_n}]}{e^{-t}} \leq e^t \alpha^n.$$

That is, $N_t$ admits exponential moments and we have proved our claim.

Now, we have

$$\mathbb{P}(N_t = 1) = 1 - \mathbb{P}(N_t = 0) - \mathbb{P}(N_t \geq 2)$$

and thus

$$\lim_{t\to 0} \frac{1}{t} P(N_t = 1) = \lim_{t\to 0} \frac{1 - e^{-\lambda t} + o(t)}{t} = \lambda\,.$$

Finally, for $\beta \in [0,1]$ put $f(t) = \mathbb{E}[\beta^{N_t}]$. Clearly $f(t+s) = f(t)f(s)$ and $f$ is of the form $f(t) = e^{t\,g(\beta)}$. But, we also have

$$\begin{aligned} f(t) &= \sum_{n=0}^{\infty} \beta^n\,\mathbb{P}(N_t = n) \\ &= \mathbb{P}(N_t = 0) + \beta\,\mathbb{P}(N_t = 1) + \sum_{n=2}^{\infty} \beta^n\,\mathbb{P}(N_t = n) \end{aligned}$$

and $g(\beta) = f'(0)$. This gives

$$\begin{aligned} g(\beta) &= \lim_{t\to 0} \frac{\mathbb{P}(N_t = 0) - 1}{t} + \frac{\beta\,\mathbb{P}(N_t = 1)}{t} + \frac{1}{t}\,o(t) \\ &= -\lambda + \lambda\,\beta, \end{aligned}$$

so

$$f(t) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n \beta^n}{n!}$$

and the required result follows. □

**Definition 4.98.** The parameter $\lambda$ is called the *intensity* of $N$. In particular we have

$$\mathbb{E}[N_t] = \lambda t$$
$$\mathrm{Var}[N_t] = \lambda t\,.$$

We shall finish this lecture with two fundamental notions which are at the heart of Stochastic Process Theory. They are very specific to this domain of mathematics, they have now true equivalent in other domains. They are rather connected and both deal with a notion of information growing with time.

### *4.7.3 Filtrations*

When discussing the conditional expectations, we have already discussed the fact that sub-$\sigma$-fields $\mathcal{G}$ of $\mathcal{F}$ represent a certain information one has gained on some random event, for example by knowing a certain random variable ($\mathcal{G} = \sigma(X)$). The notion of filtration is in the same spirit but one step further: one is speaking here of information browning with time.

Recall that all our processes are indexed by $\mathbb{R}^+$ here.

**Definition 4.99.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *filtration* of the space $(\Omega, \mathcal{F}, \mathbb{P})$ is a family $(\mathcal{F}_t)$ of sub-$\sigma$-algebras of $\mathcal{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s \leq t$.

We shall denote the filtration by $(\mathcal{F}_t)$ simply (we cannot simplify the notation to $\mathcal{F}$ as for processes, because there would be too much confusion with the $\sigma$-algebra $\mathcal{F}$).

The quadruple $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ is called a *filtered probability space*.

A filtration $(\mathcal{F}_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is called *complete* if $\mathcal{F}_0$ (and hence each $\mathcal{F}_t$, $t \in \mathbb{R}^+$) contains the negligible sets of $(\Omega, \mathcal{F}, \mathbb{P})$. If it is not the case we make it complete by adding the negligible sets to $\mathcal{F}_0$ in the same way as in Section 4.2.

**Definition 4.100.** Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space. For all $t \geq 0$ one defines

$$\mathcal{F}_{t-} = \bigvee_{s<t} \mathcal{F}_s, \quad \mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$$

with the convention $\mathcal{F}_{0-} = \mathcal{F}_0$. One also puts

$$\mathcal{F}_\infty = \bigvee_{t\geq 0} \mathcal{F}_t.$$

This way one defines two new filtrations of $(\Omega, \mathcal{F}, \mathbb{P})$, namely $(\mathcal{F}_{t-})_{t\in\mathbb{R}^+}$ and $(\mathcal{F}_{t+})_{t\in\mathbb{R}^+}$.

A filtration $\mathcal{F}$ is *right-continuous* if $\mathcal{F}_t = \mathcal{F}_{t+}$ for all $t \geq 0$. The filtration $(\mathcal{F}_{t+})_{t\in\mathbb{R}^+}$ is always right-continuous. In the same way one can speak of *left-continuous* filtrations or *continuous* filtrations.

From now on, all the filtered probability spaces are supposed to be complete and right-continuous (one replaces $(\mathcal{F}_t)$ by $(\mathcal{F}_{t+})_{t\in\mathbb{R}^+}$ if necessary).

**Definition 4.101.** A process $X$, defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, is *adapted* if $X_t$ is $\mathcal{F}_t$-measurable for all $t \in \mathbb{R}^+$.

The *natural filtration* of a process $X$ is the filtration generated by $X$, that is, the filtration $(\mathcal{F}_t)$ where $\mathcal{F}_t = \sigma\{X_s\,;\ s \leq t\}$ is the $\sigma$-algebra generated by all the random variables $X_s$, $s \leq t$.

Once again, the natural filtration of a process is understood to be made complete and right-continuous.

The natural filtration of a process is the smallest (complete and right-continuous) filtration that makes this process measurable and adapted (exercise).

**Definition 4.102.** A process $X$ is *progressive* if for all $t \in \mathbb{R}^+$ the mapping $(s, \omega) \mapsto X_s(\omega)$ on $[0, t] \times \Omega$ is measurable for $\mathrm{Bor}([0, t]) \otimes \mathcal{F}_t$.

A subset $A$ of $\mathbb{R}^+ \times \Omega$ is *progressive* if $(s, \omega) \mapsto \mathbb{1}_A(s, \omega)$ is a progressive process.

The progressive subsets of $\mathbb{R}^+ \times \Omega$ form a $\sigma$-algebra of $\mathbb{R}^+ \times \Omega$. This $\sigma$-algebra is called the *progressive $\sigma$-algebra*. A process is progressive if and only if it is measurable with respect to this $\sigma$-algebra.

It is clear that a progressive process is measurable and adapted, but the converse is not true (cf [Del72], p. 47), one needs a little more regularity.

**Proposition 4.103.** *An adapted process with right-continuous paths (or with left-continuous paths) is progressive.*

*Proof.* For all $n \in \mathbb{N}$ let

$$X_t^n = \sum_{k=0}^{\infty} X_{(k+1)2^{-n}} \mathbb{1}_{[k2^{-n},(k+1)2^{-n}[}(t).$$

As $X$ is right-continuous, $X_t^n$ converges to $X_t$ for all $t$. But the process $X^n$ is clearly progressive with respect to the filtration $(\mathcal{F}_{t+2^{-n}})_{t\in\mathbb{R}^+}$. Consequently $X$ is progressive with respect to the filtration $(\mathcal{F}_{t+\varepsilon})_{t\in\mathbb{R}^+}$ for any $\varepsilon > 0$.

For all $s \leq t$ we have

$$X_s = \lim_{\varepsilon\to 0} X_s \mathbb{1}_{[0,t-\varepsilon]}(s) + X_t \mathbb{1}_{\{t\}}(s).$$

The term within the limit symbol is measurable with respect to $\mathrm{Bor}([0,t]) \otimes \mathcal{F}_t$, thus so is $X_s$. This proves that $X$ is progressive.

The case where $X$ has left-continuous paths is treated in the same way. □

### *4.7.4 Stopping Times*

In the theory of stochastic processes some random times occur very often and very naturally. For example "the time when the process reaches a certain value, or the boundary of a certain domain". The point is that among all random times associated to a filtered probability space, only a certain family of random times have a nice behavior, in particular when stopping the processes at these times. They are the *stopping times*.

**Definition 4.104.** Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space. A *stopping time* is a measurable map $T : \Omega \mapsto \mathbb{R}^+ \cup \{+\infty\}$ such that for all $t \in \mathbb{R}^+$ the event $(T \leq t)$ belongs to $\mathcal{F}_t$.

As the filtration $(\mathcal{F}_t)$ is right-continuous one may equivalently replace the condition "$(T \leq t) \in \mathcal{F}_t$ for all $t$" by the condition "$(T < t) \in \mathcal{F}_t$ for all $t$" (exercise).

The constant times: $T(\omega) = t$ for all $\omega$, are stopping times. More generally, for any stopping time $T$ and any $t \in \mathbb{R}^+$, then $T + t$ is a stopping time.

**Definition 4.105.** Let $T$ be a stopping time. The set of $A \in \mathcal{F}$ such that

$$A \cap (T \leq t) \text{ belongs to } \mathcal{F}_t \text{ for all } t$$

is a $\sigma$-algebra (exercise). We denote it by $\mathcal{F}_T$ and call it the *$\sigma$-algebra of events anterior to $T$*. This $\sigma$-algebra coincides with the $\sigma$-algebra of $A \in \mathcal{F}$ such that

$$A \cap (T < t) \text{ belongs to } \mathcal{F}_t \text{ for all } t\,.$$

The terminology for $\mathcal{F}_T$ comes from the fact that its definition generalizes the idea that $\mathcal{F}_t$ is the $\sigma$-algebra of events occurring before the time $t$. Indeed, the constant stopping time $T(\omega) = t$ has its anterior $\sigma$-algebra $\mathcal{F}_T$ which coincides with $\mathcal{F}_t$.

**Definition 4.106.** One denotes by $\mathcal{F}_{T-}$ the $\sigma$-algebra generated by $\mathcal{F}_0$ and the events of the form

$$A \cap (T > t), \quad t \geq 0, \quad A \in \mathcal{F}_t.$$

The $\sigma$-algebra $\mathcal{F}_{T-}$ is called the *$\sigma$-algebra of events strictly anterior to $T$*. When $T(\omega) = t$, then clearly $\mathcal{F}_{T-}$ coincides with $\mathcal{F}_{t-}$.

**Definition 4.107.** A stopping time is *discrete* if the set of its values is (at most) countable.

**Proposition 4.108.** *Every stopping time is the limit of a decreasing sequence of discrete stopping times.*

*Proof.* Let $T$ be a stopping time. For all $n \in \mathbb{N}$ put

$$S_n = +\infty \, \mathbb{1}_{T=+\infty} + \sum_{k\in\mathbb{N}} k2^{-n} \, \mathbb{1}_{(k-1)2^{-n}<T\leq k2^{-n}}.$$

Then the sequence $(S_n)$ satisfies the statements. □

Here is a list of the main properties of stopping times and of their associated $\sigma$-algebras. Note that some of them generalize to stopping times some properties that were known for constant times.

**Theorem 4.109.**

*1) If $S, T$ are stopping times then so are $S \wedge T$ and $S \vee T$.*

*2) Let $(S_n)$ be a monotonic sequence of stopping times and put $S = \lim_{n\to\infty} S_n$, then $S$ is also a stopping time.*

*3) For all stopping time $T$ we have $\mathcal{F}_{T-} \subset \mathcal{F}_T$ and $T$ is $\mathcal{F}_{T-}$-measurable.*

*4) If $S, T$ are two stopping times and if $S \leq T$ then $\mathcal{F}_S \subset \mathcal{F}_T$ and $\mathcal{F}_{S-} \subset \mathcal{F}_{T-}$; if $S < T$ then $\mathcal{F}_S \subset \mathcal{F}_{T-}$ .*

*5) For all stopping times $S, T$ and all $A \in \mathcal{F}_S$ we have $A \cap (S \leq T) \in \mathcal{F}_T$ and $A \cap (S < T) \in \mathcal{F}_{T-}$ . In particular $(S \leq T)$ belongs to $\mathcal{F}_S$ and $\mathcal{F}_T$, the event $(S = T)$ belongs to $\mathcal{F}_S$ and $\mathcal{F}_T$, finally the event $(S < T)$ belongs to $\mathcal{F}_S$ and $\mathcal{F}_{T-}$ .*

*Proof.* All the proofs are easy from the definitions and left to the reader. We just precise that for proving *d)* and *e)* it is useful to notice the following identities:

$$(S < T) = \bigcup_{r\in\mathbb{Q}^+} (S < r) \cap (r < T) \tag{4.19}$$

$$(S \leq T) = \bigcap_{r\in\mathbb{Q}^+} (S \leq r) \cup (r \leq T). \qquad \square \tag{4.20}$$

**Proposition 4.110.** *For all $A \in \mathcal{F}_\infty$ and all stopping time $T$, the set $A \cap (T = \infty)$ belongs to $\mathcal{F}_{T-}$. In particular the events $(T = \infty)$, $(T < \infty)$ belong to $\mathcal{F}_{T-}$.*

*Proof.* As the set of $A \in \mathcal{F}_\infty$ such that $A \cap (T = \infty)$ belongs to $\mathcal{F}_{T-}$ is a $\sigma$-algebra, it is sufficient to prove the result for all $A \in \mathcal{F}_n$, $n \in \mathbb{N}$. But, in this case, $A \cap (T = \infty)$ is equal to $\cap_{m\geq n}\{A \cap (T > m)\}$ which clearly belongs to $\mathcal{F}_{T-}$. □

**Theorem 4.111.** *Let $(T_n)$ be a monotonic sequence of stopping times. Let $T = \lim_n T_n$.*
*1) If $(T_n)$ is decreasing then*

$$\mathcal{F}_T = \bigcap_n \mathcal{F}_{T_n} .$$

*2) If $(T_n)$ is increasing then*

$$\mathcal{F}_{T-} = \bigvee_n \mathcal{F}_{T_n-} .$$

*Proof.* 1) Clearly $\mathcal{F}_T \subset \cap_n \mathcal{F}_{T_n}$. Now let $A \in \cap_n \mathcal{F}_{T_n}$. We have $A \cap (T < t) = \cup_n A \cap (T_n < t)$, which is an element of $\mathcal{F}_t$. Thus $A$ belongs to $\mathcal{F}_T$.

2) Clearly $\mathcal{F}_{T-}$ contains $\vee_n \mathcal{F}_{T_n-}$. Now consider $A \cap (t < T)$, with $A \in \mathcal{F}_t$, a typical generator of $\mathcal{F}_{T-}$. This set also writes as $\cup_n A \cap (t < T_n)$, thus it belongs to $\vee_n \mathcal{F}_{T_n-}$. □

**Definition 4.112.** Let $S, T$ be two stopping times such that $S \leq T$. One denotes by $[\![S, T]\!]$ the following subset of $\mathbb{R} \times \Omega$:

$$\{(t, \omega) \text{ such that } t \in [S(\omega), T(\omega)]\}. \tag{4.21}$$

One defines in an analogous way the intervals $[\![S, T[\![$, $]\!]S, T]\!]$, $]\!]S, T[\![$. All these particular subsets of $\mathbb{R} \times \Omega$ are called *stochastic intervals*.

The stochastic interval $[\![S, S]\!]$ is denoted $[\![S]\!]$ and is called the *graph* of $S$; it corresponds to the set of $(t, \omega)$ in $\mathbb{R} \times \Omega$ such that $S(\omega) = t$.

**Proposition 4.113.** *Every stochastic interval is a progressive subset of $\mathbb{R}^+ \times \Omega$.*

*Proof.* The indicator function of $[\![S, T[\![$ is adapted and right-continuous, thus progressive. The indicator function of $]\!]S, T]\!]$ is adapted and left-continuous, thus progressive. Furthermore $]\!]S, T[\![ = [\![S, T[\![ \cap ]\!]S, T]\!]$ and $[\![S, T]\!] = [\![0, S[\![^c \cap ]\!]T, +\infty[\![^c$. Thus every stochastic interval is progressive. □

**Definition 4.114.** On a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, if $X$ is a stochastic process and $T$ is a stopping time, we define the random variable $X_T$ by

$$[X_T](\omega) = X_{T(\omega)}(\omega) ,$$

for all $\omega \in \Omega$.

**Proposition 4.115.** *If $X$ is a progressive process and $T$ is a finite stopping time, then $X_T$ is $\mathcal{F}_T$-measurable.*

*Proof.* We have to prove that for every Borel set $A$ the set $(X_T \in A) \cap (T \leq t)$ belongs to $\mathcal{F}_t$. But this set is equal to $(X_{T \wedge t} \in A) \cap (T \leq t)$. Let us consider

the stopping time $S = T \wedge t$, it is $\mathcal{F}_t$-measurable. As $X$ is progressive then $X_S$ is $\mathcal{F}_t$-measurable for it is the composition of the mappings $\omega \mapsto (S(\omega), \omega)$ and $(s, \omega) \mapsto X_s(\omega)$. $\square$

## 4.8 Markov Chains

In this section we just give an introduction to *Markov chains*. These discrete time stochastic processes are the simplest ones and certainly the most studied of all stochastic processes. The basic theory of Markov chains is rather simple, but some of its developments may give rise to very deep results.

### *4.8.1 Basic definitions*

**Definition 4.116.** A random process $(X_n)_{n\in\mathbb{N}}$, defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and with values in $(E, \mathcal{E})$ is a *Markov chain* if, for all $n \in \mathbb{N}$ and all $A \in \mathcal{E}$, we have, putting $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$

$$\mathbb{P}(X_{n+1} \in A \,|\, \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in A \,|\, X_n)\,.$$

Denote by $\mathbb{P}_n$ the conditional law of $X_{n+1}$ knowing $X_n$. An easy consequence of the above is that, if $\mu$ is the law of $X_0$, then the law of $(X_0, \dots, X_n)$ is the measure

$$\nu_n(\mathrm{d}x_0, \mathrm{d}x_1, \dots, \mathrm{d}x_n) = \mu(\mathrm{d}x_0)\mathbb{P}_0(x_0, \mathrm{d}x_1) \dots \mathbb{P}_{n-1}(x_{n-1}, \mathrm{d}x_n)\,.$$

**Definition 4.117.** In this introduction we are interested only in the case of *homogeneous Markov chains* that is, such that $\mathbb{P}_n$ does not depend on $n$. In that case we put

$$\mathbb{P}_n(x, \mathrm{d}y) = \mathbb{P}(x, \mathrm{d}y)$$

for all $n$.

An homogeneous Markov chain is thus determined by a single transition kernel $\mathbb{P}$; this transition kernel is the conditional law of $X_1$ knowing $X_0$.

**Definition 4.118.** If $\mathbb{P}$ and $\mathbb{Q}$ are two transition kernels, then one can check easily that

$$\mathbb{P}\mathbb{Q}(x, \mathrm{d}y) = \int_E \mathbb{P}(x, \mathrm{d}z)\, \mathbb{Q}(z, \mathrm{d}y)$$

defines a new transition kernel on $E \times \mathcal{E}$. This way, one can speak of $\mathbb{P}^n$ in the sense of this composition.

All the definitions and notations we put above are resumed in the following proposition whose proof is almost immediate.

**Proposition 4.119.** *Let $(X_n)$ be a homogeneous Markov chain with transition kernel $\mathbb{P}$ and initial measure $\mu$. The law of $X_n$ is then $\mu\mathbb{P}^n$ and for every bounded measurable function $f$ on $E$, all $n, m \in \mathbb{N}$, we have*

$$\mathbb{E}[f(X_{n+m}) \,|\, X_n] = \mathbb{P}^m f(X_n)\,.$$

### *4.8.2 Existence*

Consider a given transition kernel $\mathbb{P}$ and a measure $\mu$, it is natural to wonder if it is always possible to construct a homogeneous Markov chain $(X_n)$ for which $\mathbb{P}$ would be the transition kernel and $\mu$ the initial measure. The answer is actually an easy application of Daniell-Kolmogorov Theorem. Indeed, the family of measures

$$\nu_n(\mathrm{d}x_0, \ldots, \mathrm{d}x_n) = \mu(\mathrm{d}x_0)\,\mathbb{P}(x_0, \mathrm{d}x_1)\,\ldots\,\mathbb{P}(x_{n-1}, \mathrm{d}x_n)$$

is consistant. We thus get the following result.

**Theorem 4.120.** *If $E$ is countable or is a locally compact topological space, equipped with its Borel $\sigma$-field, then on the space $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}})$ there exists a unique measure $\mathbb{P}_\mu$ such that the coordinate process $(X_n)$ is a Markov chain with initial measure $\mu$ and transition kernel $\mathbb{P}$. In other words*

$$\mathbb{P}_\mu\left(X_0 \in A_0, \ldots, X_n \in A_n\right) = \int_{A_0\times\ldots\times A_n} \mu(\mathrm{d}x_0)\,\mathbb{P}(x_0, \mathrm{d}x_1)\,\ldots\,\mathbb{P}(x_{n-1}, \mathrm{d}x_n)\,.$$

**Definition 4.121.** The above construction is called the *canonical version* of the Markov chain with kernel $\mathbb{P}$ and initial measure $\mu$.

When the initial measure is $\mu = \delta_x$, the process starts with a deterministic initial value. We denote by $\mathbb{P}_x$ the associated canonical measure.

These measures $\mathbb{P}_x$ are enough to determine the measures $\mathbb{P}_\mu$, as appears clearly with the following result.

**Proposition 4.122.** *If $\mu$ is some initial measure and if $A \in \mathcal{E}^{\mathbb{N}}$, then*

$$\mathbb{P}_\mu(A) = \int_E \mathbb{P}_x(A)\,\mu(\mathrm{d}x)\,.$$

*Proof.* By definition we have

$$\mathbb{P}_x(A_0 \times \ldots \times A_n) = \mathbb{1}_{A_0}(x) \int_{E^n} \mathbb{1}_{A_1\times\ldots\times A_n}\,\mathbb{P}(x, \mathrm{d}x_1)\,\ldots\,\mathbb{P}(x_{n-1}, \mathrm{d}x_n)$$

and

$$\mathbb{P}_\mu(A_0 \times \ldots \times A_n) = \int_{E^{n+1}} \mathbb{1}_{A_0 \times A_1 \times \ldots \times A_n} \, \mu(\mathrm{d}x)\mathbb{P}(x, \mathrm{d}x_1) \ldots \mathbb{P}(x_{n-1}, \mathrm{d}x_n).$$

One can easily conclude now. □

**Definition 4.123.** In what follows we denote by $\mathbb{E}_\mu$ the expectation with respect to $\mathbb{P}_\mu$ and by $\mathbb{E}_x$ the expectation with respect to $\mathbb{P}_x$.

If $X$ and $F$ are random variables the notation

$$\mathbb{E}_X[F]$$

means the random variable

$$\omega \mapsto \mathbb{E}_{X(\omega)}[F]\,.$$

**Theorem 4.124.** *If $(X_n)$ is a Markov chain, on its canonical space $(E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}, \mathbb{P}_\mu)$ and canonical filtration $(\mathcal{F}_n)$, if $f$ is any bounded measurable function on $E^{m+1}$, then for all $n \in \mathbb{N}$ we have*

$$\mathbb{E}_\mu\left[f(X_n, \ldots, X_{n+m}) \,|\, \mathcal{F}_n\right] = \mathbb{E}_{X_n}\left[f(X_0, \ldots, X_m)\right]\,.$$

*Proof.* In the case $f$ is of the form

$$f(x_0, \ldots, x_m) = f_0(x_0) \ldots f(x_m)$$

we get

$$\begin{aligned}
&\mathbb{E}_\mu\left[f(X_n, \ldots, X_{n+m}) \,|\, \mathcal{F}_n\right] \\
&\quad = \mathbb{E}\left[f_0(X_n) \ldots f_{m-1}(X_{n+m-1})\mathbb{E}\left[f_m(X_{n+m}) \,|\, \mathcal{F}_{n+m-1}\right] \,|\, \mathcal{F}_n\right] \\
&\quad = \mathbb{E}\left[f_0(X_n) \ldots f_{m-1}(X_{n+m-1})\, \mathbb{P}f_m(X_{n+m-1}) \,|\, \mathcal{F}_n\right] \\
&\quad = \mathbb{E}\left[f_0(X_n) \ldots \widetilde{f}_{m-1}(X_{n+m-1}) \,|\, \mathcal{F}_n\right]\,.
\end{aligned}$$

Iterating this idea we finally obtain a certain explicit function $h(X_n)$ at the end. Had we computed

$$\mathbb{E}_x\left[f(X_0, \ldots, X_m)\right]$$

with the same procedure, we would have obtained $h(x)$. This proves the theorem for such product functions. The conclusion comes easily by some usual approximation argument. □

### *4.8.3 Strong Markov Property*

**Definition 4.125.** The canonical space $(\Omega, \mathcal{F}, \mathbb{P}) = (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}, \mathbb{P}_\mu)$ of $(X_n)$ is advantageous for it carries a natural *shift*. We put

$$\theta((\omega_n)_{n\in\mathbb{N}}) = (\omega_{n+1})_{n\in\mathbb{N}}.$$

For every $n \in \mathbb{N}^*$ we put $\theta_n = \theta^n$ and $\theta_0 = I$. One can easily check that

$$\theta_n(\mathcal{F}) = \sigma(X_k; k \geq n)$$

and

$$\theta_n(X_p) = X_{p+n}\,.$$

For every stopping time $T$, we define on $(T < \infty)$ the operator $\theta_T$ by $\theta_T = \theta_n$ on $(T = n)$.

In terms of $\theta$, the Markov property reads as follows.

**Theorem 4.126.** *Let $(X_n)$ be a canonical Markov process, with initial measure $\mu$. For every bounded (*resp. *positive) measurable function $F$ on $\Omega$, we have*

$$\mathbb{E}_\mu[F \circ \theta_n \,|\, \mathcal{F}_n] = \mathbb{E}_{X_n}[\,F\,]\,.$$

*Proof.* Let us denote by $\mathcal{F}_n$ the $\sigma$-field generated by $X_0, \ldots, X_n$. We have, using Theorem 4.124

$$\begin{aligned}\mathbb{E}_\mu[\mathbb{1}_{(X_{k_1}\in A_1,\ldots,X_{k_m}\in A_m)} \circ \theta_n \,|\, \mathcal{F}_n] \\ &= \mathbb{E}_\mu[\mathbb{1}_{(X_{n+k_1}\in A_1,\ldots,X_{n+k_m}\in A_m)} \,|\, \mathcal{F}_n] \\ &= \mathbb{E}_{X_n}[\mathbb{1}_{(X_{n+k_1}\in A_1,\ldots,X_{n+k_m}\in A_m)}]\,.\end{aligned}$$

Hence the theorem is proved for those particular functions. We conclude by a monotone class argument. □

The main result of this subsection is that the above theorem extends to stopping times. This is the so-called *Strong Markov Property*.

**Theorem 4.127 (Strong Markov Property).** *Let $(X_n)$ be a canonical Markov chain, with initial measure $\mu$ and let $T$ be a stopping time valued in $\mathbb{N} \cap \{+\infty\}$. For every bounded (*resp. *positive) measurable function $F$ on $\Omega$, we have*

$$\mathbb{E}_\mu[\mathbb{1}_{(T<\infty)}\, F \circ \theta_T \,|\, \mathcal{F}_T] = \mathbb{1}_{(T<\infty)}\, \mathbb{E}_{X_T}[\,F\,].$$

*Proof.* We have

$$\mathbb{E}_\mu[\mathbb{1}_{(T<\infty)}\, F \circ \theta_T \,|\, \mathcal{F}_T] = \sum_{n\in\mathbb{N}} \mathbb{E}_\mu[\mathbb{1}_{(T=n)}\, F \circ \theta_n \,|\, \mathcal{F}_T]\,.$$

We need here the following lemma.

**Lemma 4.128.** *If $F$ is an integrable random variable and if $T$ is a discrete stopping time, then*

$$\mathbb{E}[\mathbb{1}_{(T=a)}\, F \,|\, \mathcal{F}_T] = \mathbb{1}_{(T=a)}\, \mathbb{E}[F \,|\, \mathcal{F}_a]\,.$$

*Proof (of the lemma).* If $A$ is any event of $\mathcal{F}_T$ then

$$\mathbb{E}[\mathbb{1}_{(T=a)}\, F\, \mathbb{1}_A] = \mathbb{E}[\mathbb{1}_{(T=a)\cap A}\, F]\,.$$

But $(T=a)\cap A$ belongs to $\mathcal{F}_a$ and hence

$$\mathbb{E}[\mathbb{1}_{(T=a)\cap A}\, F] = \mathbb{E}[\mathbb{1}_{(T=a)\cap A}\, \mathbb{E}[F\,|\,\mathcal{F}_a]\,] = \mathbb{E}[\mathbb{1}_A\, \mathbb{1}_{(T=a)}\, \mathbb{E}[F\,|\,\mathcal{F}_a]\,]\,.$$

But $\mathbb{1}_{(T=a)}\, \mathbb{E}[F\,|\,\mathcal{F}_a]$ is $\mathcal{F}_T$-measurable and we have proved the lemma.

Coming back to the theorem, we have, using this lemma and Theorem 4.126

$$\begin{aligned}\mathbb{E}_\mu[\mathbb{1}_{(T<\infty)}\, F\circ\theta_T\,|\,\mathcal{F}_T] &= \sum_{n\in\mathbb{N}} \mathbb{1}_{(T=n)}\, \mathbb{E}_\mu[F\circ\theta_n\,|\,\mathcal{F}_n] \\ &= \sum_{n\in\mathbb{N}} \mathbb{1}_{(T=n)}\, \mathbb{E}_{X_n}[\,F\,] \\ &= \mathbb{1}_{(T<\infty)}\, \mathbb{E}_{X_T}[\,F\,]\,.\end{aligned}$$

This proves the theorem. □

## Notes

The litterature on Probability Theory and Stochastic Processes Theory is huge. Hundreds of new books are appearing every year. It is thus difficult to guide the reader in this profusion.

For those interested in general books on probability theory we can recommend the two old volumes by Feller [Fel68] and [Fel71]. They are not so modern in their presentation, but they stay an unavoidable reference. More recent and considered as a must-have by most of the probabilists is the book by Billingsley [Bil95]. Also very complete and pleasant to read, we recommend Loeve's book [Loè77].

For the reader willing to enter more deeply in the general theory of stochastic processes, Brownian motion, Poisson process ... (but without entering into stochastic calculus yet) once again the choice is incredible. Our favorites are the very complete book by Roger and Williams [RW00] (not for an introductory book, but essential for a probabilist's personal library), Karlin and Taylor's book [KT75] is also a reference.

Otherwise there are plenty of very famous references on stochastic processes, but which include the theory of stochastic integration: Revuz-Yor [RY99], Protter [Pro05] and the two first volumes of Dellacherie-Meyer [DM75], [DM80].

However, when writing this lecture, we have been inspired by several references that we quote here. We have used the chapter II of Roger and Williams

[RW00] for the Monotone Class Theorems. Our presentation of Gaussian families is taken from Neveu [Nev68]. The section on conditional expectations is inspired from Dellacherie-Meyer [DM75]. The discussion on regularization of stochastic processes follows Revuz and Yor [RY99] and Roger and Williams [RW00]. Our presentation of Brownian motion and its properties follows Revuz and Yor [RY99], the presentation of Poisson processes is taken from Protter [Pro05]. The section on filtrations takes a lot from Dellacherie-Meyer [DM75], whereas for stopping times we have mainly followed the (must-have) little book of Dellacherie [Del72].

## References


[Bil95] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[Del72] Claude Dellacherie. *Capacités et processus stochastiques*. Springer-Verlag, Berlin, 1972. Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 67.

[DM75] Claude Dellacherie and Paul-André Meyer. *Probabilités et potentiel*. Hermann, Paris, 1975. Chapitres I à IV, Édition entièrement refondue, Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. XV, Actualités Scientifiques et Industrielles, No. 1372.

[DM80] Claude Dellacherie and Paul-André Meyer. *Probabilités et potentiel. Chapitres V à VIII*, volume 1385 of *Actualités Scientifiques et Industrielles [Current Scientific and Industrial Topics]*. Hermann, Paris, revised edition, 1980. Théorie des martingales. [Martingale theory].

[Fel68] William Feller. *An introduction to probability theory and its applications. Vol. I*. Third edition. John Wiley & Sons Inc., New York, 1968.

[Fel71] William Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York, 1971.

[KT75] Samuel Karlin and Howard M. Taylor. *A first course in stochastic processes*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, second edition, 1975.

[Loè77] Michel Loève. *Probability theory. I*. Springer-Verlag, New York, fourth edition, 1977. Graduate Texts in Mathematics, Vol. 45.

[Nev68] Jacques Neveu. *Processus aléatoires gaussiens*. Séminaire de Mathématiques Supérieures, No. 34 (Été, 1968). Les Presses de l'Université de Montréal, Montreal, Que., 1968.

[Pro05] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.

[RW00] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Foundations, Reprint of the second (1994) edition.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[Vil06] Cedric Villani. *Intégration et Analyse de Fourier*. Course on internet: cedricvillani.org/wp-content/uploads/2013/03/IAF.pdf. Ecole Normale Supérieure de Lyon, 2006.