Stochastic Processes and Applications

Jean-François Delmas

September 15, 2023

Contents

Chapter 1

A starter on measure theory and random variables

In this chapter, we present in Section 1.1 a basic tool kit in measure theory with in mind the applications to probability theory. In Section 1.2, we develop the corresponding integration and expectation. The presentation of this chapter follows closely [1], see also [2].

We use the following convention $\mathbb{N} = \{0, 1, \ldots\}$ is the set of non-negative integers, $\mathbb{N}^* = \mathbb{N} \bigcap (0, +\infty)$, and for $m < n \in \mathbb{N}$, we set $[\![m, n]\!] = [m, n] \bigcap \mathbb{N}$. We shall consider $\overline{\mathbb{R}} = \mathbb{R} \bigcup \{\pm\infty\} = [-\infty, +\infty]$, and for $a, b \in \overline{\mathbb{R}}$, we write $a \vee b = \max(a, b)$, $a^+ = a \vee 0$ the positive part of a, and $a^- = (-a)^+$ its negative part.

For two sets $A \subset E$, the function $\mathbf{1}_A$ defined on E taking values in $\mathbb{R}$ is equal to 1 on A and to 0 on $E \setminus A$.

1.1 Measures and measurable functions

1.1.1 Measurable space

Let Ω be a set also called a space. A measure on a set Ω is a function which gives the "size" of subsets of Ω. We shall see that, if one asks the measure to satisfy some natural additive properties, it is not always possible to define the measure of any subsets of Ω. For this reason, we shall consider families of sub-sets of Ω called σ-fields. We denote by $\mathcal{P}(\Omega) = \{A; A \subset \Omega\}$ the set of all subsets of Ω.

Definition 1.1. *A collection of subsets of Ω, $\mathcal{F} \subset \mathcal{P}(\Omega)$, is called a σ-field on Ω if:*

(i) $\Omega \in \mathcal{F}$;

(ii) $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$;

(iii) if $(A_i, i \in I)$ *is a finite or countable collection of elements of* $\mathcal{F}$, *then* $\bigcup_{i \in I} A_i \in \mathcal{F}$.

We call $(\Omega, \mathcal{F})$ *a measurable space and a set* $A \in \mathcal{F}$ *is said to be* $\mathcal{F}$*-measurable.*

When there is no ambiguity on the σ-field we shall simply say that A is measurable instead of $\mathcal{F}$-measurable. In probability theory a measurable set is also called an event. Properties

(i) and (ii) implies that $\emptyset$ is measurable. Notice that $\mathcal{P}(\Omega)$ and $\{\emptyset, \Omega\}$ are σ-fields. The latter is called the trivial σ-field. When Ω is at most countable, unless otherwise specified, we shall consider the σ-field $\mathcal{P}(\Omega)$.

Proposition 1.2. *Let $\mathcal{C} \subset \mathcal{P}(\Omega)$. There exists a smallest σ-field on Ω which contains $\mathcal{C}$.*

The smallest σ-field which contains $\mathcal{C}$ is denoted by $\sigma(\mathcal{C})$ and is also called the σ-field generated by $\mathcal{C}$.

Proof. Let $(\mathcal{F}_j, j \in J)$ be the collection of all the σ-fields on Ω containing $\mathcal{C}$. This collection is not empty as it contains $\mathcal{P}(\Omega)$. It is left to the reader to check that $\bigcap_{j\in J} \mathcal{F}_j$ is a σ-field. Clearly, this is the smallest σ-field containing $\mathcal{C}$. □

Remark 1.3. In this remark we give an explicit description of a σ-field generated by a finite number of sets. Let $\mathcal{C} = \{A_1, \ldots, A_n\}$, with $n \in \mathbb{N}^*$, be a finite collection of subsets of Ω. It is elementary to check that $\mathcal{F} = \left\{\bigcup_{I\in\mathcal{I}} C_I; \mathcal{I} \subset \mathcal{P}([\![1, n]\!])\right\}$, with $C_I = \bigcap_{i\in I} A_i \bigcap_{j\notin I} A_j^c$ and $I \subset [\![1, n]\!]$, is a σ-field. Notice that $C_I \bigcap C_J = \emptyset$ for $I \neq J$. Thus, the subsets C_I are atoms of $\mathcal{F}$ in the sense that if $B \in \mathcal{F}$, then $C_I \bigcap B$ is equal to C_I or to $\emptyset$.

We shall prove that $\sigma(\mathcal{C}) = \mathcal{F}$. Since by construction $C_I \in \sigma(\mathcal{C})$ for all $I \subset [\![1, n]\!]$, we deduce that $\mathcal{F} \subset \sigma(\mathcal{C})$. On the other hand, for all $i \in [\![1, n]\!]$, we have $A_i = \bigcup_{I\subset[\![1,n]\!],\, i\in I} C_I$. This gives that $\mathcal{C} \subset \mathcal{F}$, and thus $\sigma(\mathcal{C}) \subset \mathcal{F}$. In conclusion, we get $\sigma(\mathcal{C}) = \mathcal{F}$. ♢

If $\mathcal{F}$ and $\mathcal{H}$ are σ-fields on Ω, we denote by $\mathcal{F} \vee \mathcal{H} = \sigma(\mathcal{F} \bigcup \mathcal{H})$ the σ-field generated by $\mathcal{F}$ and $\mathcal{H}$. More generally, if $(\mathcal{F}_i, i \in I)$ is a collection of σ-fields on Ω, we denote by $\bigvee_{i\in I} \mathcal{F}_i = \sigma(\bigcup_{i\in I} \mathcal{F}_i)$ the σ-field generated by $(\mathcal{F}_i, i \in I)$.

We shall consider product of measurable spaces. If $(A_i, i \in I)$ is a collection of sets, then its product is denoted by $\prod_{i\in I} A_i = \{(\omega_i, i \in I); \omega_i \in A_i \quad \forall i \in I\}$.

Definition 1.4. *Let $((\Omega_i, \mathcal{F}_i), i \in I)$ be a collection of measurable spaces. The product σ-field $\bigotimes_{i\in I} \mathcal{F}_i$ on the product space $\prod_{i\in I} \Omega_i$ is the σ-field generated by all the sets $\prod_{i\in I} A_i$ such $A_i \in \mathcal{F}_i$ for all $i \in I$ and $A_i = \Omega_i$ for all $i \in I$ but for a finite number of indices.*

When all the measurable spaces $(\Omega_i, \mathcal{F}_i)$ are the same for all $i \in I$, say $(\Omega, \mathcal{F})$, then we also write the product space $\prod_{i\in I} \Omega_i = \Omega^I$ and the product σ-field $\bigotimes_{i\in I} \mathcal{F}_i = \mathcal{F}^{\otimes I}$.

We recall a topological space $(E, \mathcal{O})$ is a space E and a collection $\mathcal{O}$ of subsets of E such that: $\emptyset$ and E belongs to $\mathcal{O}$, any (finite or infinite) union of elements of $\mathcal{O}$ belongs to $\mathcal{O}$, and the intersection of any finite number of elements of $\mathcal{O}$ belongs to $\mathcal{O}$. The elements of $\mathcal{O}$ are called the open sets, and $\mathcal{O}$ is called a topology on E. There is a very natural σ-field on a topological space.

Definition 1.5. *If $(E, \mathcal{O})$ is a topological space, then the Borel σ-field, $\mathcal{B}(E) = \sigma(\mathcal{O})$, is the σ-field generated by all the open sets. An element of $\mathcal{B}(E)$ is called a Borel set.*

Usually the Borel σ-field on E is different from $\mathcal{P}(E)$.

Remark 1.6. Since all the open subsets of $\mathbb{R}$ can be written as the union of a countable number of bounded open intervals, we deduce that the Borel σ-field is generated by all the intervals (a, b) for $a < b$. It is not trivial to exhibit a set which is not a Borel set; an example was provided by Vitali[1].

[1] J. Stern. "Le problème de la mesure." Séminaire Bourbaki 26 (1983-1984): 325-346. `http://eudml.org/doc/110033`.

Similarly to the one dimensional case, as all the open sets of $\mathbb{R}^d$ can be written as a countable union of open rectangles, the Borel σ-field on $\mathbb{R}^d$, $d \geq 1$, is generated by all the rectangles $\prod_{i=1}^d (a_i, b_i)$ with $a_i < b_i$ for $1 \leq i \leq d$. In particular, we get that the Borel σ-field of $\mathbb{R}^d$ is the product[2] of the d Borel σ-fields on $\mathbb{R}$. $\Diamond$

1.1.2 Measures

We give in this section the definition and some properties of measures and probability measures.

Definition 1.7. *Let $(\Omega, \mathcal{F})$ be a measurable space.*

(i) A $[0, +\infty]$-valued function μ defined on $\mathcal{F}$ is σ-additive if for all finite or countable collection $(A_i, i \in I)$ of measurable pairwise disjoint sets, that is $A_i \in \mathcal{F}$ for all $i \in I$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$, we have:

$$\mu\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mu(A_i). \tag{1.1}$$

(ii) A measure μ on $(\Omega, \mathcal{F})$ is a σ-additive $[0, +\infty]$-valued function defined on $\mathcal{F}$ such that $\mu(\emptyset) = 0$. We call $(\Omega, \mathcal{F}, \mu)$ a measured space. A measurable set A is μ-null if $\mu(A) = 0$.

(iii) A measure μ on $(\Omega, \mathcal{F})$ is σ-finite if there exists a sequence of measurable sets $(\Omega_n, n \in \mathbb{N})$ such that $\bigcup_{n\in\mathbb{N}} \Omega_n = \Omega$ and $\mu(\Omega_n) < +\infty$ for all $n \in \mathbb{N}$.

(iv) A probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is a measure on $(\Omega, \mathcal{F})$ such that $\mathbb{P}(\Omega) = 1$. The measured space $(\Omega, \mathcal{F}, \mathbb{P})$ is also called a probability space.

We refer to Section 7.1 for the construction of measures such as the Lebesgue measure, see Proposition 7.4 and Remark 7.6, and the product probability measure, see Proposition 7.7.

Example 1.8. We give some examples of measures (check these are indeed measures). Let Ω be a space.

(i) The counting measure Card is defined by $A \mapsto \text{Card}\,(A)$ for $A \subset \Omega$, with $\text{Card}\,(A)$ the cardinal of A. It is σ-finite if and only if Ω is at most countable.

(ii) Let $\omega \in \Omega$. The Dirac measure at ω, δ_ω, is defined by $A \mapsto \delta_\omega(A) = \mathbf{1}_A(\omega)$ for $A \subset \Omega$. It is a probability measure.

(iii) The Bernoulli probability distribution with parameter $p \in [0, 1]$, $\mathrm{P}_{\mathcal{B}(p)}$, is a probability measure on $(\mathbb{R}, \mathcal{B}(R))$ given by $\mathrm{P}_{\mathcal{B}(p)} = (1 - p)\delta_0 + p\delta_1$.

---

[2]Let $(E_1, \mathcal{O}_1)$ and $(E_2, \mathcal{O}_2)$ be two topological spaces. Let $\mathcal{C} = \{O_1 \times O_2;\, O_1 \in \mathcal{O}_1, O_2 \in \mathcal{O}_1\}$ be the set of product of open sets. By definition, $\mathcal{B}(E_1) \otimes \mathcal{B}(E_2)$ is the σ-field generated by $\mathcal{C}$. The product topology $\mathcal{O}_1 \otimes \mathcal{O}_2$ on $E_1 \times E_2$ is defined as the smallest topology on $E_1 \times E_2$ containing $\mathcal{C}$. The Borel σ-field on $E_1 \times E_2$, $\mathcal{B}(E_1 \times E_2)$, is the σ-field generated by $\mathcal{O}_1 \otimes \mathcal{O}_2$. Since $\mathcal{C} \subset \mathcal{O}_1 \otimes \mathcal{O}_2$, we deduce that $\mathcal{B}(E_1) \otimes \mathcal{B}(E_2) \subset \mathcal{B}(E_1 \times E_2)$. Since $\mathcal{O}_1 \otimes \mathcal{O}_2$ is stable by infinite (even uncountable) union, it might happens that the previous inclusion is not an equality, see Theorem 4.44 p. 149 from C. Aliprantis and K. Border. *Infinite Dimensional Analysis*. Springer, 2006.

(iv) The Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a measure characterized by $\lambda([a,b]) = b-a$ for all $a < b$. In particular, any finite set or (by σ-additivity) any countable set is λ-null[3]. The Lebesgue measure is σ-finite.

△

Let us mention that assuming only the additivity property (that is I is assumed to be finite in (1.1)), instead of the stronger σ-additivity property, for the definition of measures[4] leads to a substantially different and less efficient approach. We give elementary properties of measures.

Proposition 1.9. *Let μ be a measure on $(\Omega, \mathcal{F})$. We have the following properties.*

(i) Additivity: $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$, for all $A, B \in \mathcal{F}$.

(ii) Monotonicity: $A \subset B$ implies $\mu(A) \leq \mu(B)$, for all $A, B \in \mathcal{F}$.

(iii) Monotone convergence: If $(A_n, n \in \mathbb{N})$ is a sequence of elements of $\mathcal{F}$ such that $A_n \subset A_{n+1}$ for all $n \in \mathbb{N}$, then, we have:

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

(iv) If $(A_i, i \in I)$ is a finite or countable collection of elements of $\mathcal{F}$, then we have the inequality $\mu\left(\bigcup_{i\in I} A_i\right) \leq \sum_{i\in I} \mu(A_i)$. In particular a finite or countable union of μ-null sets is μ-null.

Proof. We prove (i). The sets $A \cap B^c$, $A \cap B$ and $A^c \cap B$ are measurable and pairwise disjoint. Using the additivity property three times, we get:

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A \cap B^c) + 2\mu(A \cap B) + \mu(A^c \cap B) = \mu(A) + \mu(B).$$

We prove (ii). As $A^c \cap B \in \mathcal{F}$, we get by additivity that $\mu(B) = \mu(A) + \mu(A^c \cap B)$. Then use that $\mu(A^c \cap B) \geq 0$, to conclude.

We prove (iii). We set $B_0 = A_0$ and $B_n = A_n \cap A_{n-1}^c$ for all $n \in \mathbb{N}^*$ so that $\bigcup_{n\leq m} B_n = A_m$ for all $m \in \mathbb{N}^*$ and $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n$. The sets $(B_n, n \geq 0)$ are measurable and pairwise disjoint. By σ-additivity, we get $\mu(A_m) = \mu(\bigcup_{n\leq m} B_n) = \sum_{n\leq m} \mu(B_n)$ and $\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{n\in\mathbb{N}} \mu(B_n)$. Use the convergence of the partial sums $\sum_{n\leq m} \mu(B_n)$, whose terms are non-negative, towards $\sum_{n\in\mathbb{N}} \mu(B_n)$ as m goes to infinity to conclude.

Property (iv) is a direct consequence of properties (i) and (iii). □

We give a property for probability measures, which is deduced from (i) of Proposition 1.9.

Corollary 1.10. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A \in \mathcal{F}$. We have $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.*

[3]A set $A \subset \mathbb{R}$ is negligible if there exists a λ-null set B such that $A \subset B$ (notice that A might not be a Borel set). Let N_λ be the sets of negligible sets. The Lebesgue σ-field, $\mathcal{B}^\lambda(\mathbb{R})$, on $\mathbb{R}$ is the σ-field generated by the Borel σ-field and N_λ. By construction, we have $\mathcal{B}(\mathbb{R}) \subset \mathcal{B}^\lambda(\mathbb{R}) \subset \mathcal{P}(\mathbb{R})$. It can be proved that those two inclusions are strict.

[4]H. Föllmer and A. Schied. *Stochastic finance. An introduction in discrete time.* De Gruyter, 2011.

We end this section with the definition of independent events.

Definition 1.11. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The events $(A_i, i \in I)$ are independent if for all finite subset $J \subset I$, we have:*

$$\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j).$$

The σ-fields $(\mathcal{F}_i, i \in I)$ are independent if for all $A_i \in \mathcal{F}_i \subset \mathcal{F}$, $i \in I$, the events $(A_i, i \in I)$ are independent.

1.1.3 Characterization of probability measures

In this section, we prove that if two probability measures coincide on a sufficiently large family of events, then they are equal, see the main results of Corollaries 1.14 and 1.15. After introducing a λ-system (or monotone class), we prove the monotone class theorem.

Definition 1.12. *A collection $\mathcal{A}$ of sub-sets of Ω is a λ-system (or monotone class) if:*

(i) $\Omega \in \mathcal{A}$*;*

(ii) $A, B \in \mathcal{A}$ *and* $A \subset B$ *imply* $B \cap A^c \in \mathcal{A}$*;*

(iii) if $(A_n, n \in \mathbb{N})$ *is an increasing sequence of elements of* $\mathcal{A}$*, then we have* $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Theorem 1.13 (Mononote class Theorem)**.** *Let $\mathcal{C}$ be a collection of sub-sets of Ω stable by finite intersection (also called a π-system). All λ-system (or monotone class) containing $\mathcal{C}$ also contains $\sigma(\mathcal{C})$.*

Proof. Notice that $\mathcal{P}(\Omega)$ is λ-system containing *cc*. Let $\mathcal{A}$ be the intersection of all λ-systems containing $\mathcal{C}$. It is easy to check that $\mathcal{A}$ is the smallest λ-system containing $\mathcal{C}$. It is clear that $\mathcal{A}$ satisfies properties (i) and (ii) from Definition 1.1. To check that property (iii) from Definition 1.1 holds also, so that $\mathcal{A}$ is a σ-field, it is enough, according to property (iii) from Definition 1.12, to check that $\mathcal{A}$ is stable by finite union or equivalently by finite intersection, thanks to property (ii) of Definition 1.12.

For $B \subset \Omega$, set $\mathcal{A}_B = \{A \subset \Omega; A \cap B \in \mathcal{A}\}$. Assume that $B \in \mathcal{C}$. It is easy to check that $\mathcal{A}_B$ is a λ-system, as $\mathcal{C}$ is stable by finite intersection, and that it contains $\mathcal{C}$ and thus $\mathcal{A}$. Therefore, for all $B \in \mathcal{C}$, $A \in \mathcal{A}$, we get $A \in \mathcal{A}_B$ and thus $A \cap B \in \mathcal{A}$.

Assume now that $B \in \mathcal{A}$. It is easy to check that $\mathcal{A}_B$ is a λ-system. According to the previous part, it contains $\mathcal{C}$ and thus $\mathcal{A}$. In particular, for all $B \in \mathcal{A}$, $A \in \mathcal{A}$, we get $A \in \mathcal{A}_B$ and thus $A \cap B \in \mathcal{A}$. We deduce that $\mathcal{A}$ is stable by finite intersection and is therefore a σ-field. To conclude, notice that $\mathcal{A}$ contains $\mathcal{C}$ and thus $\sigma(\mathcal{C})$ also. □

Corollary 1.14. *Let $\mathbb{P}$ and $\mathbb{P}'$ be two probability measures defined on a measurable space $(\Omega, \mathcal{F})$ Let $\mathcal{C} \subset \mathcal{F}$ be a collection of events stable by finite intersection. If $\mathbb{P}(A) = \mathbb{P}'(A)$ for all $A \in \mathcal{C}$, then we have $\mathbb{P}(B) = \mathbb{P}'(B)$ for all $B \in \sigma(\mathcal{C})$.*

Proof. Notice that $\{A \in \mathcal{F}; \mathbb{P}(A) = \mathbb{P}'(A)\}$ is a λ-system. It contains $\mathcal{C}$. By the monotone class theorem, it contains $\sigma(\mathcal{C})$. □

The next corollary is an immediate consequence of Definition 1.5 and Corollary 1.14.

Corollary 1.15. *Let $(E, \mathcal{O})$ be a topological space. Two probability measures on $(E, \mathcal{B}(E))$ which coincide on the open sets $\mathcal{O}$ are equal.*

1.1.4 Measurable functions

Let $(S, \mathcal{S})$ and $(E, \mathcal{E})$ be two measurable spaces. Let f be a function defined on S and taking values in E. For $A \subset E$, we set $\{f \in A\} = f^{-1}(A) = \{x \in S; f(x) \in A\}$. It is easy to check that for $A \subset E$ and $(A_i, i \in I)$ a collection of subsets of E, we have:

$$f^{-1}(A^c) = f^{-1}(A)^c, \quad f^{-1}\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f^{-1}(A_i) \quad \text{and} \quad f^{-1}\left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} f^{-1}(A_i). \tag{1.2}$$

We deduce from the properties (1.2) the following elementary lemma.

Lemma 1.16. *Let f be a function from S to E and $\mathcal{E}$ a σ-field on E. The collection of sets $\left\{f^{-1}(A);\, A \in \mathcal{E}\right\}$ is a σ-field on S.*

The σ-field $\left\{f^{-1}(A);\, A \in \mathcal{E}\right\}$, denoted by $\sigma(f)$, is also called the σ-field generated by f.

Definition 1.17. *A function f defined on a space S and taking values in a space E is measurable from $(S, \mathcal{S})$ to $(E, \mathcal{E})$ if $\sigma(f) \subset \mathcal{S}$.*

When there is no ambiguity on the σ-fields $\mathcal{S}$ and $\mathcal{E}$, we simply say that f is measurable.

Example 1.18. Let $A \subset S$. The function $\mathbf{1}_A$ is measurable from $(S, \mathcal{S})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ if and only if A is measurable as $\sigma(\mathbf{1}_A) = \{\emptyset, S, A, A^c\}$. △

The next proposition is useful to prove that a function is measurable.

Proposition 1.19. *Let $\mathcal{C}$ be a collection of subsets of E which generates the σ-field $\mathcal{E}$ on E. A function f from S to E is measurable from $(S, \mathcal{S})$ to $(E, \mathcal{E})$ if and only if for all $A \in \mathcal{C}$, $f^{-1}(A) \in \mathcal{S}$.*

Proof. We denote by $\mathcal{G}$ the σ-field generated by $\left\{f^{-1}(A), A \in \mathcal{C}\right\}$. We have $\mathcal{G} \subset \sigma(f)$. It is easy to check that the collection $\left\{A \in E; f^{-1}(A) \in \mathcal{G}\right\}$ is a σ-field on E. It contains $\mathcal{C}$ and thus $\mathcal{E}$. This implies that $\sigma(f) \subset \mathcal{G}$ and thus $\sigma(f) = \mathcal{G}$. We conclude using Definition 1.17. □

We deduce the following result, which is important in practice.

Corollary 1.20. *A continuous function defined on a topological space and taking values in a topological space is measurable with respect to the Borel σ-fields.*

The next proposition concerns function taking values in product spaces.

Proposition 1.21. *Let $(S, \mathcal{S})$ and $((E_i, \mathcal{E}_i), i \in I)$ be measurable spaces. For all $i \in I$, let f_i be a function defined on S taking values in E_i and set $f = (f_i, i \in I)$. The function f is measurable from $(S, \mathcal{S})$ to $(\prod_{i\in I} E_i, \bigotimes_{i\in I} \mathcal{E}_i)$ if and only if for all $i \in I$, the function f_i is measurable from $(S, \mathcal{S})$ to $(E_i, \mathcal{E}_i)$.*

Proof. By definition, the σ-field $\bigotimes_{i\in I}\mathcal{E}_i$ is generated by $\prod_{i\in I} A_i$ with $A_i \in \mathcal{E}_i$ and for all $i \in I$ but one, $A_i = E_i$. Let $\prod_{i\in I} A_i$ be such a set. Assume it is not equal to $\prod_{i\in I} E_i$ and let i_0 denote the only index such that $A_{i_0} \neq E_{i_0}$. We have $f^{-1}\left(\prod_{i\in I} A_i\right) = f_{i_0}^{-1}(A_{i_0}) \in \mathcal{S}$. Thus if f is measurable so is f_{i_0}. The converse is a consequence of Proposition 1.19. $\square$

The proof of the next proposition is immediate.

Proposition 1.22. *Let $(\Omega, \mathcal{F})$, $(S, \mathcal{S})$, $(E, \mathcal{E})$ be three measurable spaces, f a measurable function defined on Ω taking values in S and g a measurable function defined on S taking values in E. The composed function $g \circ f$ defined on Ω and taking values in E is measurable.*

We shall consider functions taking values in $\overline{\mathbb{R}}$. The Borel σ-field on $\overline{\mathbb{R}}$, $\mathcal{B}(\overline{\mathbb{R}})$, is by definition the σ-field generated by $\mathcal{B}(\mathbb{R})$, $\{+\infty\}$ and $\{-\infty\}$ or equivalently by the family $([-\infty, a), a \in \mathbb{R})$. We say a function (resp. a sequence) is real-valued if it takes values in (resp. its elements belong to) $\overline{\mathbb{R}}$. With the convention $0 \cdot \infty = 0$, the product of two real-valued functions is always defined. The sum of two functions f and g taking values in $\overline{\mathbb{R}}$ is well defined if (f, g) does not take the values $(+\infty, -\infty)$ or $(-\infty, +\infty)$.

Corollary 1.23. *Let f and g be real-valued measurable functions defined on the same measurable space. The functions fg, $f \vee g = \max(f, g)$ are measurable. If (f, g) does not take the values $(+\infty, -\infty)$ and $(-\infty, +\infty)$, then the function $f + g$ is measurable.*

Proof. The $\overline{\mathbb{R}}^2$-valued functions defined on $\overline{\mathbb{R}}^2$ by $(x, y) \mapsto xy$, $(x, y) \mapsto x \vee y$ and $(x, y) \mapsto (x + y)\mathbf{1}_{\{(x,y)\in\overline{\mathbb{R}}^2\setminus\{(-\infty,+\infty),(+\infty,-\infty)\}\}}$ are continuous on $\mathbb{R}^2$ and thus measurable on $\mathbb{R}^2$ according to Corollary 1.20. Thus, they are also measurable on $\overline{\mathbb{R}}^2$. The corollary is then a consequence of Proposition 1.22. $\square$

If $(a_n, n \in \mathbb{N})$ is a real-valued sequence then its lower and upper limit are defined by:

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} \inf\{a_k,\, k \geq n\} \quad \text{and} \quad \limsup_{n\to\infty} a_n = \lim_{n\to\infty} \sup\{a_k,\, k \geq n\}$$

and they belong to $\overline{\mathbb{R}}$. Notice that:

$$\limsup_{n\to\infty} a_n = \limsup_{n\to\infty} a_n^+ - \liminf_{n\to\infty} a_n^-.$$

The sequence $(a_n, n \in \mathbb{N})$ converges (in $\overline{\mathbb{R}}$) if $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$ and this common value, denoted by $\lim_{n\to\infty} a_n$, belongs to $\overline{\mathbb{R}}$.

The next proposition asserts in particular that the limit of measurable functions is measurable.

Proposition 1.24. *Let $(f_n, n \in \mathbb{N})$ be a sequence of real-valued measurable functions defined on a measurable space $(S, \mathcal{S})$. The functions $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ are measurable. The set of convergence of the sequence, $\{x \in S;\, \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x)\}$, is measurable. In particular, if the sequence $(f_n, n \in \mathbb{N})$ converges, then its limit, denoted by $\lim_{n\to\infty} f_n$, is also measurable.*

Proof. For $a \in \mathbb{R}$, we have:

$$\left\{\limsup_{n\to\infty} f_n < a\right\} = \bigcup_{k\in\mathbb{N}^*} \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} \left\{f_n \leq a - \frac{1}{k}\right\}.$$

Since the functions f_n are measurable, we deduce that $\{\limsup_{n\to\infty} f_n < a\}$ is also measurable. Since the σ-field $\mathcal{B}(\overline{\mathbb{R}})$ is generated by $[-\infty, a)$ for $a \in \mathbb{R}$, we deduce from Proposition 1.19 that $\limsup_{n\to\infty} f_n$ is measurable. Since $\liminf_{n\to\infty} f_n = -\limsup_{n\to\infty}(-f_n)$, we deduce that $\liminf_{n\to\infty} f_n$ is measurable.

Let $h = \limsup_{n\to\infty} f_n - \liminf_{n\to\infty} f_n$, with the convention $+\infty - \infty = 0$. The function h is measurable thanks to Corollary 1.23. Since the set of convergence is equal to $h^{-1}(\{0\})$ and that $\{0\}$ is a Borel set, we deduce that the set of convergence is measurable. □

We end this section with a very useful result which completes Proposition 1.22.

Proposition 1.25. *Let $(\Omega, \mathcal{F})$, $(S, \mathcal{S})$ be measurable spaces, f a measurable function defined on Ω taking values is S and φ a measurable function from $(\Omega, \sigma(f))$ to $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Then, there exists a real-valued measurable function g defined on S such that $\varphi = g \circ f$.*

Proof. By simplicity, we assume that φ takes its values in $\mathbb{R}$ instead of $\overline{\mathbb{R}}$. For all $k \in \mathbb{Z}$, $n \in \mathbb{N}$ the sets $A_{n,k} = \varphi^{-1}([k2^{-n}, (k+1)2^{-n}))$ are $\sigma(f)$-measurable. Thus, for all $n \in \mathbb{N}$, there exists a collection $(B_{n,k}, k \in \mathbb{Z})$ of sets of S pairwise disjoint such that $\bigcup_{k\in\mathbb{Z}} B_{n,k} = S$, $B_{n,k} \in \mathcal{S}$ and $f^{-1}(B_{n,k}) = A_{n,k}$ for all $k \in \mathbb{Z}$. For all $n \in \mathbb{N}$, the real-valued function $g_n = 2^{-n} \sum_{k\in\mathbb{Z}} k \mathbf{1}_{B_{n,k}}$ defined on S is measurable, and we have $g_n \circ f \leq \varphi \leq g_n \circ f + 2^{-n_0}$ for $n \geq n_0 \geq 0$. The function $g = \limsup_{n\to\infty} g_n$ is measurable according to Proposition 1.24, and we have $g \circ f \leq \varphi \leq g \circ f + 2^{-n_0}$ for all $n_0 \in \mathbb{N}$. This implies that $g \circ f = \varphi$. □

1.1.5 Probability distribution and random variables

We first start with the definition of the image measure (or push-forward measure) which is obtained by transferring a measure using a measurable function. The proof of the next Lemma is elementary and left to the reader.

Lemma 1.26. *Let $(E, \mathcal{E}, \mu)$ be a measured space, $(S, \mathcal{S})$ a measurable space, and f a measurable function defined on E and taking values in S. We define the function μ_f on $\mathcal{S}$ by $\mu_f(A) = \mu(f^{-1}(A))$ for all $A \in \mathcal{S}$. Then μ_f is a measure on $(S, \mathcal{S})$.*

The measure μ_f is called the push-forward measure or image measure of μ by f.

In what follow, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.27. *Let $(E, \mathcal{E})$ be a measurable space. An E-valued random variable X defined on Ω is a measurable function from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$. Its probability distribution or law is the image probability measure $\mathbb{P}_X$.*

At some point we shall specify the σ-field $\mathcal{F}$ on Ω, and say that X is $\mathcal{F}$-measurable.

We say two E-valued random variables X and Y are equal in distribution, and we write $X \overset{(d)}{=} Y$, if $\mathbb{P}_X = \mathbb{P}_Y$. For $A \in \mathcal{E}$, we recall we write $\{X \in A\} = \{\omega; X(\omega) \in A\} = X^{-1}(A)$. Two random variables X and Y defined on the same probability space are equal a.s., and we write $X \overset{\text{a.s.}}{=} Y$, if $\mathbb{P}(X = Y) = 1$. Notice that if X and Y are equal a.s., then they have the same probability distribution.

Remark 1.28. Let X be a real-valued random variable. Its cumulative distribution function F_X is defined by $F_X(x) = \mathbb{P}(X \leq x)$ for all $x \in \mathbb{R}$. It is easy to deduce from Exercise 8.1

that if X and Y are real-valued random variables, then X and Y are equal in distribution if and only if $F_X = F_Y$. $\lozenge$

The next lemma gives a characterization of the distribution of a family of random variables.

Lemma 1.29. *Let $((E_i, \mathcal{E}_i), i \in I)$ be a collection of measurable spaces and $X = (X_i, i \in I)$ a random variable taking values in the product space $\prod_{i\in I} E_i$ endowed with the product σ-field. The distribution of X is characterized by the family of distributions of $(X_j, j \in J)$ where J runs over all finite subsets of I.*

According to Proposition 1.21, in the above lemma the X_j is an E_j-valued random variable and its marginal probability distribution can be recovered from the distribution of X as:

$$\mathbb{P}(X_j \in A_j) = \mathbb{P}\left(X \in \prod_{i\in I} A_i\right) \quad \text{with } A_i = E_i \text{ for } i \neq j.$$

Proof. According to Definition 1.4, the product σ-field $\mathcal{E}$ on the product space $E = \prod_{i\in I} E_i$ is generated by the family $\mathcal{C}$ of product sets $\prod_{i\in I} A_i$ such $A_i \in \mathcal{E}_i$ for all $i \in I$ and $A_i = E_i$ for all $i \notin J$, with $J \subset I$ finite. Notice then that $\mathbb{P}_X(\prod_{i\in I} A_i) = \mathbb{P}(X_j \in A_j \text{ for } j \in J)$. Since $\mathcal{C}$ is stable by finite intersection, we deduce from the monotone class theorem, and more precisely Corollary 1.14, that the probability measure $\mathbb{P}_X$ on E is uniquely characterized by $\mathbb{P}(X_j \in A_j \text{ for } j \in J)$ for all J finite subset of I and all $A_j \in \mathcal{E}_j$ for $j \in J$. $\square$

We first give the definition of a random variable independent from a σ-field.

Definition 1.30. *A random variable X taking values in a measurable space $(E, \mathcal{E})$ is independent from a σ-field $\mathcal{H} \subset \mathcal{F}$ if $\sigma(X)$ and $\mathcal{H}$ are independent or equivalently if, for all $A \in \mathcal{E}$ and $B \in \mathcal{H}$, the events $\{X \in A\}$ and B are independent.*

We now give the definition of independent random variables.

Definition 1.31. *The random variables $(X_i, i \in I)$ are independent if the σ-fields $(\sigma(X_i), i \in I)$ are independent. Equivalently, the random variables $(X_i, i \in I)$ are independent if for all finite subset $J \subset I$, all $A_j \in \mathcal{E}_j$ with $j \in J$, we have:*

$$\mathbb{P}(X_j \in A_j \textit{ for all } j \in J) = \prod_{j\in J} \mathbb{P}(X_j \in A_j).$$

We deduce from this definition that if the marginal probability distributions P_i of all the random variables X_i for $i \in I$ are known and if $(X_i, i \in I)$ are independent, then the distribution of X is the product probability $\bigotimes_{i\in I} \mathrm{P}_i$ introduced in Proposition 7.7.

We end this section with the Bernoulli scheme.

Theorem 1.32. *Let $(E, \mathcal{E}, \mathrm{P})$ be a probability space. Let I be a set of indices. Then, there exists a probability space and a sequence $(X_i, i \in I)$ of E-valued random variables defined on this probability space which are independent and with the same distribution probability P.*

When P is the Bernoulli probability distribution and $I = \mathbb{N}^*$, then $(X_n, n \in \mathbb{N}^*)$ is called a Bernoulli scheme.

Proof. For $i \in I$, set $\Omega_i = E$, $\mathcal{F}_i = \mathcal{E}$ and $\mathbb{P}_i = \mathrm{P}$. According to Proposition 7.7, we can consider the product space $\Omega = \prod_{i\in I} \Omega_i$ with the product σ-field and the product probability $\bigotimes_{i\in I} \mathbb{P}_i$. For all $i \in I$, we consider the random variable: $X_i(\omega) = \omega_i$ where $\omega = (\omega_i, i \in I)$. Using Definition 1.31, we deduce that the random variables $(X_i, i \in I)$ are independent with the same probability distribution P. □

1.2 Integration and expectation

Using the results from the integration theory of Sections 1.2.1 and 1.2.2, we introduce in Section 1.2.5 the expectation of real-valued or $\mathbb{R}^d$-valued random variables and give some well known inequalities. We study the properties of the L^p spaces in Section 1.2.3 and prove the Fubini theorem in Section 1.2.4. In Section 1.2.6 we collect some further results on independence.

1.2.1 Integration: construction and properties

Let $(S, \mathcal{S}, \mu)$ be a measured space. The set $\overline{\mathbb{R}}$ is endowed with its Borel σ-field. We use the convention $0 \cdot \infty = 0$. A function f defined on S is simple if it is real-valued, measurable and if there exists $n \in \mathbb{N}^*$, $\alpha_1, \ldots, \alpha_n \in [0, +\infty]$, $A_1, \ldots, A_n \in \mathcal{S}$ such that we have the representation $f = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$. The integral of f with respect to μ, denoted by $\mu(f)$ or $\int f \, \mathrm{d}\mu$ or $\int f(x)\mu(\mathrm{d}x)$, is defined by:

$$\mu(f) = \sum_{k=1}^n \alpha_k \, \mu(A_k) \in [0, +\infty].$$

Lemma 1.33. *Let f be a simple function defined on S. The integral $\mu(f)$ does not depend on the choice of its representation.*

Proof. Consider two representations for f: $f = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k} = \sum_{\ell=1}^m \beta_\ell \mathbf{1}_{B_\ell}$, with $n, m \in \mathbb{N}^*$ and $A_1, \ldots, A_n, B_1, \ldots, B_m \in \mathcal{S}$. We shall prove that $\sum_{k=1}^n \alpha_k \mu(A_k) = \sum_{\ell=1}^m \beta_\ell \mu(B_\ell)$.

According to Remark 1.3, there exits a finite family of measurable sets $(C_I, I \in \mathcal{P}([\![1, n+m]\!]))$ such that $C_I \bigcap C_J = \emptyset$ if $I \neq J$ and for all $k \in [\![1, n]\!]$ and $\ell \in [\![1, m]\!]$ there exists $\mathcal{I}_k \subset [\![1, n]\!]$ and $\mathcal{J}_\ell \subset [\![1, m]\!]$ such that $A_k = \bigcup_{I\in\mathcal{I}_k} C_I$ and $B_\ell = \bigcup_{I\in\mathcal{J}_\ell} C_I$. We deduce that:

$$f = \sum_{I\in\mathcal{P}([\![1,n+m]\!])} \left(\sum_{k=1}^n \alpha_k \mathbf{1}_{\{I\in\mathcal{I}_k\}} \right) \mathbf{1}_{C_I} = \sum_{I\in\mathcal{P}([\![1,n+m]\!])} \left(\sum_{\ell=1}^m \beta_\ell \mathbf{1}_{\{I\in\mathcal{J}_\ell\}} \right) \mathbf{1}_{C_I}$$

and thus $\sum_{k=1}^n \alpha_k \mathbf{1}_{\{I\in\mathcal{I}_k\}} = \sum_{\ell=1}^m \beta_\ell \mathbf{1}_{\{I\in\mathcal{J}_\ell\}}$ for all I such that $C_I \neq \emptyset$. We get:

$$\sum_{k=1}^n \alpha_k \mu(A_k) = \sum_I \left(\sum_{k=1}^n \alpha_k \mathbf{1}_{\{I\in\mathcal{I}_k\}} \right) \mu(C_I) = \sum_I \left(\sum_{\ell=1}^m \beta_\ell \mathbf{1}_{\{I\in\mathcal{J}_\ell\}} \right) \mu(C_I) = \sum_{\ell=1}^m \beta_\ell \mu(B_\ell),$$

where we used the additivity of μ for the first and third equalities. This ends the proof. □

It is elementary to check that if f and g are simple functions, then we get:

$$\mu(af + bg) = a\mu(f) + b\mu(g) \quad \text{for } a, b \in [0, +\infty) \text{ (linearity)}, \tag{1.3}$$
$$f \leq g \Rightarrow \mu(f) \leq \mu(g) \quad \text{(monotonicity)}. \tag{1.4}$$

Definition 1.34. *Let f be a $[0,+\infty]$-valued measurable function defined on S. We define the integral of f with respect to the measure μ by:*

$$\mu(f) = \sup\{\mu(g);\ g \text{ simple such that } 0 \le g \le f\}.$$

The next lemma gives a representation of $\mu(f)$ using that a non-negative measurable function f is the non-decreasing limit of a sequence of simple functions. Such sequence exists. Indeed, one can define for $n \in \mathbb{N}^*$ the simple function f_n by $f_n(x) = \min(n, 2^{-n}\lfloor 2^n f(x)\rfloor)$ for $x \in S$ with the convention $\lfloor +\infty \rfloor = +\infty$. Then, the functions $(f_n, n \in \mathbb{N}^*)$ are measurable and their non-decreasing limit is f.

Lemma 1.35. *Let f be a $[0,+\infty]$-valued function defined on S and $(f_n, n \in \mathbb{N})$ a non-decreasing sequence of simple functions such that $\lim_{n\to\infty} f_n = f$. Then, we have that $\lim_{n\to\infty} \mu(f_n) = \mu(f)$.*

Proof. It is enough to prove that for all non-decreasing sequence of simple functions $(f_n, n \in \mathbb{N})$ and simple function g such that $\lim_{n\to\infty} f_n \ge g$, we have $\lim_{n\to\infty} \mu(f_n) \ge \mu(g)$. We deduce from the proof of Lemma 1.33 that there exists a representation of g such that $g = \sum_{k=1}^N \alpha_k \mathbf{1}_{A_k}$ and the measurable sets $(A_k, 1 \le k \le N)$ are pairwise disjoint. Using this representation and the linearity, we see it is enough to consider the particular case $g = \alpha \mathbf{1}_A$, with $\alpha \in [0,+\infty]$, $A \in \mathcal{S}$ and $f_n \mathbf{1}_{A^c} = 0$ for all $n \in \mathbb{N}$.

By monotonicity, the sequence $(\mu(f_n), n \in \mathbb{N})$ is non-decreasing and thus $\lim_{n\to\infty} \mu(f_n)$ is well defined, taking values in $[0,+\infty]$.

The result is clear if $\alpha = 0$. We assume that $\alpha > 0$. Let $\alpha' \in [0, \alpha)$. For $n \in \mathbb{N}$, we consider the measurable sets $B_n = \{f_n \ge \alpha'\}$. The sequence $(B_n, n \in \mathbb{N})$ is non-decreasing with A as limit because $\lim_{n\to\infty} f_n \ge g$. The monotone property for measure, see property (iii) from Proposition 1.9, implies that $\lim_{n\to\infty} \mu(B_n) = \mu(A)$. As $\mu(f_n) \ge \alpha' \mu(B_n)$, we deduce that $\lim_{n\to\infty} \mu(f_n) \ge \alpha' \mu(A)$ and that $\lim_{n\to\infty} \mu(f_n) \ge \mu(g)$ as $\alpha' \in [0, \alpha)$ is arbitrary. □

Corollary 1.36. *The linearity and monotonicity properties, see (1.3) and (1.4), also hold for $[0,+\infty]$-valued measurable functions f and g defined on S.*

Proof. Let $(f_n, n \in \mathbb{N})$ and $(g_n, n \in \mathbb{N})$ be two non-decreasing sequences of simple functions converging respectively towards f and g. Let $a, b \in [0,+\infty)$. The non-decreasing sequence $(af_n + bg_n, n \in \mathbb{N})$ of simple functions converges towards $af + bg$. By linearity, we get:

$$\mu(af + bg) = \lim_{n\to\infty} \mu(af_n + bg_n) = a \lim_{n\to\infty} \mu(f_n) + b \lim_{n\to\infty} \mu(g_n) = a\mu(f) + b\mu(g).$$

Assume $f \le g$. The non-decreasing sequence $((f_n \vee g_n), n \in \mathbb{N})$ of simple functions converges towards g. By monotonicity, we get $\mu(f) = \lim_{n\to\infty} \mu(f_n) \le \lim_{n\to\infty} \mu(f_n \vee g_n) = \mu(g)$. □

Recall that for a function f, we write $f^+ = f \vee 0 = \max(f, 0)$ and $f^- = (-f)^+$.

Definition 1.37. *Let f be a real-valued measurable function defined on S. The integral of f with respect to the measure μ is well defined if $\min(\mu(f^+), \mu(f^-)) < +\infty$ and it is given by:*

$$\mu(f) = \mu(f^+) - \mu(f^-).$$

The function f is μ-integrable if $\max(\mu(f^+), \mu(f^-)) < +\infty$ (i.e. $\mu(|f|) < +\infty$).

We also write $\mu(f) = \int f \, \mathrm{d}\mu = \int f(x) \, \mu(\mathrm{d}x)$. A property holds μ-almost everywhere (μ-a.e.) if it holds on a measurable set B such that $\mu(B^c) = 0$. If μ is a probability measure, then one says μ-almost surely (μ-a.s.) for μ-a.e.. We shall omit μ and write a.e. or a.s. when there is no ambiguity on the measure.

Lemma 1.38. *Let $f \geq 0$ be a real-valued measurable function defined on S. We have:*

$$\mu(f) = 0 \Longleftrightarrow f = 0 \quad \mu\text{-a.e.}.$$

Proof. The equivalence is clear if f is simple.

When f is not simple, consider a non-decreasing sequence of simple (non-negative) functions $(f_n, n \in \mathbb{N})$ which converges towards f. As $\{f \neq 0\}$ is the non-decreasing limit of the measurable sets $\{f_n \neq 0\}$, $n \in \mathbb{N}$, we deduce from the monotonicity property of Proposition 1.9, that $f = 0$ a.e. if and only if $f_n = 0$ a.e. for all $n \in \mathbb{N}$. We deduce from the first part of the proof that $f = 0$ a.e. if and only if $\mu(f_n) = 0$ for all $n \in \mathbb{N}$. As $(\mu(f_n), n \in \mathbb{N})$ is non-decreasing and converges towards $\mu(f)$, we get that $\mu(f_n) = 0$ for all $n \in \mathbb{N}$ if and only if $\mu(f) = 0$. We deduce that $f = 0$ a.e. if and only if $\mu(f) = 0$. □

The next corollary asserts that it is enough to know f a.e. to compute its integral.

Corollary 1.39. *Let f and g be two real-valued measurable functions defined on S such that $\mu(f)$ and $\mu(g)$ are well defined. If a.e. $f = g$, then we have $\mu(f) = \mu(g)$.*

Proof. Assume first that $f \geq 0$ and $g \geq 0$. By hypothesis the measurable set $A = \{f \neq g\}$ is μ-null. We deduce that a.e. $f\mathbf{1}_A = 0$ and $g\mathbf{1}_A = 0$. This implies that $\mu(f\mathbf{1}_A) = \mu(g\mathbf{1}_A) = 0$. By linearity, see Corollary 1.36, we get:

$$\mu(f) = \mu(f\mathbf{1}_{A^c}) + \mu(f\mathbf{1}_A) = \mu(g\mathbf{1}_{A^c}) = \mu(g\mathbf{1}_{A^c}) + \mu(g\mathbf{1}_A) = \mu(g).$$

To conclude notice that $f = g$ a.e. implies that $f^+ = g^+$ a.e. and $f^- = g^-$ a.e. and then use the first part of the proof to conclude. □

The relation $f = g$ a.e. is an equivalence relation on the set of real-valued measurable functions defined on S. We shall identify a function f with its equivalent class of all measurable functions g such that μ-a.e., $f = g$. Notice that if f is μ-integrable, then μ-a.e. $|f| < +\infty$. In particular, if f and g are μ-integrable, we shall write $f + g$ for any element of the equivalent class of $f\mathbf{1}_{\{|f|<+\infty\}} + g\mathbf{1}_{\{|g|<+\infty\}}$. Using this remark, we conclude this section with the following immediate corollary.

Corollary 1.40. *The linearity property, see (1.3) with $a, b \in \mathbb{R}$, and the monotonicity property (1.4), where $f \leq g$ can be replaced by $f \leq g$ a.e., hold for real-valued measurable μ-integrable functions f and g defined on S.*

We deduce that the set of $\mathbb{R}$-valued μ-integrable functions defined on S is a vector space. The linearity property (1.3) holds also on the set of real-valued measurable functions h such that $\mu(h^+) < +\infty$ and on the set of real-valued measurable functions h such that $\mu(h^-) < +\infty$. The monotonicity property holds also for real-valued measurable functions f and g such that $\mu(f)$ and $\mu(g)$ are well defined.

1.2.2 Integration: convergence theorems

The a.e. convergence for sequences of measurable functions introduced below is weaker than the simple convergence and adapted to the convergence of integrals. Let $(S, \mathcal{S}, \mu)$ be a measured space.

Definition 1.41. *Let $(f_n, n \in \mathbb{N})$ be a sequence of real-valued measurable functions defined on S. The sequence converges a.e. if a.e. $\liminf_{n\to\infty} f_n = \limsup_{n\to\infty} f_n$. We denote by $\lim_{n\to\infty} f_n$ any element of the equivalent class of the measurable functions which are a.e. equal to $\liminf_{n\to\infty} f_n$.*

Notice that Proposition 1.24 assures indeed that $\liminf_{n\to\infty}$ is measurable. We thus deduce the following corollary.

Corollary 1.42. *If a sequence of real-valued measurable functions defined on S converges a.e., then its limit is measurable.*

We now give the three main results on the convergence of integrals for sequence of converging functions.

Theorem 1.43 (Monotone convergence theorem)**.** *Let $(f_n, n \in \mathbb{N})$ be a sequence of real-valued measurable functions defined on S such that for all $n \in \mathbb{N}$, a.e. $0 \leq f_n \leq f_{n+1}$. Then, we have:*

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \int \lim_{n\to\infty} f_n \, \mathrm{d}\mu.$$

Proof. The set $A = \{x; f_n(x) < 0 \text{ or } f_n(x) > f_{n+1}(x) \text{ for some } n \in \mathbb{N}\}$ is μ-null as countable union of μ-null sets. Thus, we get that a.e. $f_n = f_n \mathbf{1}_{A^c}$ for all $n \in \mathbb{N}$. Corollary 1.39 implies that, replacing f_n by $f_n \mathbf{1}_{A^c}$ without loss of generality, it is enough to prove the theorem under the stronger conditions: for all $n \in \mathbb{N}$, $0 \leq f_n \leq f_{n+1}$. We set $f = \lim_{n\to\infty} f_n$ the non-decreasing (everywhere) limit of $(f_n, n \in \mathbb{N})$.

For all $n \in \mathbb{N}$, let $(f_{n,k}, k \in \mathbb{N})$ be a non-decreasing sequence of simple functions which converges towards f_n. We set $g_n = \max\{f_{j,n}; 1 \leq j \leq n\}$. The non-decreasing sequence of simple functions $(g_n, n \in \mathbb{N})$ converges to f and thus $\lim_{n\to\infty} \int g_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$. By monotonicity, $g_n \leq f_n \leq f$ implies $\int g_n \, \mathrm{d}\mu \leq \int f_n \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\mu$. Taking the limit, we get $\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$. $\square$

The proof of the next corollary is left to the reader (hint: use the monotone convergence theorem to get the σ-additivity).

Corollary 1.44. *Let f be a real-valued measurable function defined on S such that a.e. $f \geq 0$. Then the function $f\mu$ defined on $\mathcal{S}$ by $f\mu(A) = \int \mathbf{1}_A f \, \mathrm{d}\mu$ is a measure on $(S, \mathcal{S})$.*

We shall say that the measure $f\mu$ has density f with respect to the reference measure μ.

Fatou's lemma will be used for the proof of the dominated convergence theorem, but it is also interesting by itself.

Lemma 1.45 (Fatou's lemma)**.** *Let $(f_n, n \in \mathbb{N})$ be a sequence of real-valued measurable functions defined on S such that a.e. $f_n \geq 0$ for all $n \in \mathbb{N}$. Then, we have the lower* *semi-continuity property:*

$$\liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu \geq \int \liminf_{n\to\infty} f_n \, \mathrm{d}\mu.$$

Proof. The function $\liminf_{n\to\infty} f_n$ is the non-decreasing limit of the sequence $(g_n, n \in \mathbb{N})$ with $g_n = \inf_{k\geq n} f_k$. We get:

$$\int \liminf_{n\to\infty} f_n \, \mathrm{d}\mu = \lim_{n\to\infty} \int g_n \, \mathrm{d}\mu \leq \lim_{n\to\infty} \inf_{k\geq n} \int f_k \, \mathrm{d}\mu = \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu,$$

where we used the monotone convergence theorem for the first equality and the monotonicity property of the integral for the inequality. □

The next theorem and the monotone convergence theorem are very useful to exchange integration and limit.

Theorem 1.46 (Dominated convergence theorem)**.** *Let $f, g, (f_n, n \in \mathbb{N})$ and $(g_n, n \in \mathbb{N})$ be real-valued measurable functions defined on S. We assume that a.e.: $|f_n| \leq g_n$ for all $n \in \mathbb{N}$, $f = \lim_{n\to\infty} f_n$ and $g = \lim_{n\to\infty} g_n$. We also assume that $\lim_{n\to\infty} \int g_n \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu$ and $\int g \, \mathrm{d}\mu < +\infty$. Then, we have:*

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \int \lim_{n\to\infty} f_n \, \mathrm{d}\mu.$$

Taking $g_n = g$ for all $n \in \mathbb{N}$ in the above theorem gives the following result.

Corollary 1.47 (Lebesgue's dominated convergence theorem)**.** *Let f, g and $(f_n, n \in \mathbb{N})$ be real-valued measurable functions defined on S. We assume that a.e.: $|f_n| \leq g$ for all $n \in \mathbb{N}$, $f = \lim_{n\to\infty} f_n$ and $\int g \, \mathrm{d}\mu < +\infty$. Then, we have:*

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu = \int \lim_{n\to\infty} f_n \, \mathrm{d}\mu.$$

Proof of Theorem 1.46. As a.e. $|f| \leq g$ and $\int g \, \mathrm{d}\mu < +\infty$, we get that the function f is integrable. The functions $g_n + f_n$ and $g_n - f_n$ are a.e. non-negative. Fatou's lemma with $g_n + f_n$ and $g_n - f_n$ gives:

$$\int g \, \mathrm{d}\mu + \int f \, \mathrm{d}\mu = \int (g + f) \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int (g_n + f_n) \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu + \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu,$$
$$\int g \, \mathrm{d}\mu - \int f \, \mathrm{d}\mu = \int (g - f) \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int (g_n - f_n) \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu - \limsup_{n\to\infty} \int f_n \, \mathrm{d}\mu.$$

Since $\int g \, \mathrm{d}\mu$ is finite, we deduce from those inequalities that $\int f \, \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \, \mathrm{d}\mu$ and that $\limsup_{n\to\infty} \int f_n \, \mathrm{d}\mu \leq \int f \, \mathrm{d}\mu$. Thus, the sequence $(\int f_n \, \mathrm{d}\mu, n \in \mathbb{N})$ converges towards $\int f \, \mathrm{d}\mu$. □

We shall use the next Corollary in Chapter 5, which is a direct consequence of Fatou's lemma and dominated convergence theorem.

Corollary 1.48. *Let $f, g, (f_n, n \in \mathbb{N})$ be real-valued measurable functions defined on S. We assume that a.e. $f_n^+ \leq g$ for all $n \in \mathbb{N}$, $f = \lim_{n\to\infty} f_n$ and that $\int g \, \mathrm{d}\mu < +\infty$. Then, we have that $(\mu(f_n), n \in \mathbb{N})$ and $\mu(f)$ are well defined as well as:*

$$\limsup_{n\to\infty} \int f_n \, \mathrm{d}\mu \leq \int \lim_{n\to\infty} f_n \, \mathrm{d}\mu.$$

1.2.3 The L^p space

Let $(S, \mathcal{S}, \mu)$ be a measured space. We start this section with very useful inequalities.

Proposition 1.49. *Let f and g be two real-valued measurable functions defined on S.*

- ***Hölder inequality.*** *Let $p, q \in (1, +\infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Assume that $|f|^p$ and $|g|^q$ are integrable. Then fg is integrable and we have:*

$$\int |fg| \, \mathrm{d}\mu \leq \left(\int |f|^p \, \mathrm{d}\mu \right)^{1/p} \left(\int |g|^q \, \mathrm{d}\mu \right)^{1/q}.$$

The Hölder inequality is an equality if and only if there exists $c, c' \in [0, +\infty)$ such that $(c, c') \neq (0, 0)$ and a.e. $c|f|^p = c'|g|^q$.

- ***Cauchy-Schwarz inequality.*** *Assume that f^2 and g^2 are integrable. Then fg is integrable and we have:*

$$\int |fg| \, \mathrm{d}\mu \leq \left(\int f^2 \, \mathrm{d}\mu \right)^{1/2} \left(\int g^2 \, \mathrm{d}\mu \right)^{1/2}.$$

Furthermore, we have $\int fg \, \mathrm{d}\mu = \left(\int f^2 \, \mathrm{d}\mu \right)^{1/2} \left(\int g^2 \, \mathrm{d}\mu \right)^{1/2}$ if and only there exist $c, c' \in [0, +\infty)$ such that $(c, c') \neq (0, 0)$ and a.e. $c\, f = c'\, g$.

- ***Minkowski inequality.*** *Let $p \in [1, +\infty)$. Assume that $|f|^p$ and $|g|^p$ are integrable. We have:*

$$\left(\int |f + g|^p \, \mathrm{d}\mu \right)^{1/p} \leq \left(\int |f|^p \, \mathrm{d}\mu \right)^{1/p} + \left(\int |g|^p \, \mathrm{d}\mu \right)^{1/p}.$$

Proof. Hölder inequality. We recall the convention $0 \cdot +\infty = 0$. The Young inequality states that for $a, b \in [0, +\infty]$, $p, q \in (1, +\infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$, we have:

$$ab \leq \frac{1}{p}\, a^p + \frac{1}{q}\, b^q.$$

Indeed, this inequality is obvious if a or b belongs to $\{0, +\infty\}$. For $a, b \in (0, +\infty)$, using the convexity of the exponential function, we get:

$$ab = \exp\left(\frac{\log(a^p)}{p} + \frac{\log(b^q)}{q} \right) \leq \frac{1}{p} \exp\left(\log(a^p)\right) + \frac{1}{q} \exp\left(\log(b^q)\right) = \frac{1}{p}\, a^p + \frac{1}{q}\, b^q.$$

If $\mu(|f|^p) = 0$ or $\mu(|g|^q) = 0$, the Hölder inequality is trivially true as a.e. $fg = 0$ thanks to Lemma 1.38. If this is not the case, then integrating with respect to μ in the Young inequality with $a = |f|/\mu(|f|^p)^{1/p}$ and $b = |g|/\mu(|g|^q)^{1/q}$ gives the result. Because of the strict convexity of the exponential, if a and b are finite, then the Young inequality is an equality if and only if a^p and b^q are equal. This implies that, if $|f|^p$ and $|g|^q$ are integrable, then the Hölder inequality is an equality if and only there exist $c, c' \in [0, +\infty)$ such that $(c, c') \neq (0, 0)$ and a.e. $c|f|^p = c'|g|^q$.

The Cauchy-Schwarz inequality is the Hölder inequality with $p = q = 2$. If $\int fg \, \mathrm{d}\mu = \left(\int f^2 \, \mathrm{d}\mu \right)^{1/2} \left(\int g^2 \, \mathrm{d}\mu \right)^{1/2}$, then since $\int fg \, \mathrm{d}\mu \leq \int |fg| \, \mathrm{d}\mu$, the equality holds also in the

Cauchy-Schwarz inequality. Thus there exists $c, c' \in [0, +\infty)$ such that $(c, c') \neq (0, 0)$ and $c|f| = c'|g|$. Notice also that $\int (|fg| - fg)\, \mathrm{d}\mu = 0$. Then use Lemma 1.38 to conclude that a.e. $|fg| = fg$ and thus a.e. $c\, f = c'\, g$.

Let $p \geq 1$. From the convexity of the function $x \mapsto |x|^p$, we get $(a+b)^p \leq 2^{p-1}(a^p + b^p)$ for all $a, b \in [0, +\infty]$. We deduce that $|f+g|^p$ is integrable. The case $p = 1$ of the Minkowski inequality comes from the triangular inequality in $\overline{\mathbb{R}}$. Let $p > 1$. We assume that $\int |f + g|^p\, \mathrm{d}\mu > 0$, otherwise the inequality is trivial. Using Hölder inequality, we get:

$$\begin{aligned} \int |f+g|^p \, \mathrm{d}\mu &\leq \int |f||f+g|^{p-1} \, \mathrm{d}\mu + \int |g||f+g|^{p-1} \, \mathrm{d}\mu \\ &\leq \left(\left(\int |f|^p \, \mathrm{d}\mu \right)^{1/p} + \left(\int |g|^p \, \mathrm{d}\mu \right)^{1/p} \right) \left(\int |f+g|^p \, \mathrm{d}\mu \right)^{(p-1)/p} . \end{aligned}$$

Dividing by $\left(\int |f+g|^p \, \mathrm{d}\mu\right)^{(p-1)/p}$ gives the Minkowski inequality. □

For $p \in [1, +\infty)$, let $\mathcal{L}^p((S, \mathcal{S}, \mu))$ denote the set of $\mathbb{R}$-valued measurable functions f defined on S such that $\int |f|^p \, \mathrm{d}\mu < +\infty$. When there is no ambiguity on the underlying space, resp. space and measure, we shall simply write $\mathcal{L}^p(\mu)$, resp. $\mathcal{L}^p$. Minkowski inequality and the linearity of the integral yield that $\mathcal{L}^p$ is a vector space and the map $\|\cdot\|_p$ from $\mathcal{L}^p$ to $[0, +\infty)$ defined by $\|f\|_p = \left(\int |f|^p \, \mathrm{d}\mu\right)^{1/p}$ is a semi-norm (that is $\|f+g\|_p \leq \|f\|_p + \|g\|_p$ and $\|af\|_p \leq |a| \; \|f\|_p$ for $f, g \in \mathcal{L}^p$ and $a \in \mathbb{R}$). Notice that $\|f\|_p = 0$ implies that a.e. $f = 0$ thanks to Lemma 1.38. Recall that the relation "a.e. equal to" is an equivalence relation on the set of real-valued measurable functions defined on S. We deduce that the space $(L^p, \|\cdot\|_p)$, where L^p is the space $\mathcal{L}^p$ quotiented by the equivalence relation "a.e. equal to", is a normed vector space. We shall use the same notation for an element of $\mathcal{L}^p$ and for its equivalent class in L^p. If we need to stress the dependence of on the measure μ of the measured space $(S, \mathcal{S}, \mu)$ we may write $L^p(\mu)$ and even $L^p(S, \mathcal{S}, \mu)$ for L^p.

The next proposition asserts that the normed vector space $(L^p, \|\cdot\|_p)$ is complete and, by definition, is a Banach space. We recall that a sequence $(f_n, n \in \mathbb{N})$ of elements of L^p converges in L^p to a limit, say f, if $f \in L^p$, and $\lim_{n\to\infty} \|f_n - f\|_p = 0$.

Proposition 1.50. *Let $p \in [1, +\infty)$. The normed vector space $(L^p, \|\cdot\|_p)$ is complete. That is every Cauchy sequence of elements of L^p converges in L^p: if $(f_n, n \in \mathbb{N})$ is a sequence of elements of L^p such that $\lim_{\min(n,m)\to\infty} \|f_n - f_m\|_p = 0$, then there exists $f \in L^p$ such that $\lim_{n\to\infty} \|f_n - f\|_p = 0$.*

Proof. Let $(f_n, n \in \mathbb{N})$ be a Cauchy sequence of elements of L^p, that is $f_n \in L^p$ for all $n \in \mathbb{N}$ and $\lim_{\min(n,m)\to\infty} \|f_n - f_m\|_p = 0$. Consider the sub-sequence $(n_k, k \in \mathbb{N})$ defined by $n_0 = 0$ and for $k \geq 1$, $n_k = \inf\{m > n_{k-1}; \|f_i - f_j\|_p \leq 2^{-k}$ for all $i \geq m, j \geq m\}$. In particular, we have $\left\|f_{n_{k+1}} - f_{n_k}\right\|_p \leq 2^{-k}$ for all $k \geq 1$. Minkowski inequality and the monotone convergence theorem imply that $\left\|\sum_{k\in\mathbb{N}} |f_{n_{k+1}} - f_{n_k}|\right\|_p < +\infty$ and thus $\sum_{k\in\mathbb{N}} |f_{n_{k+1}} - f_{n_k}|$ is a.e. finite. The series with general term $(f_{n_{k+1}} - f_{n_k})$ is a.e. absolutely converging. By considering the convergence of the partial sums, we get that the sequence $(f_{n_k}, k \in \mathbb{N})$ converges a.e. towards a limit, say f. This limit is a real-valued measurable function, thanks to Corollary 1.42. We deduce from Fatou lemma that:

$$\|f_m - f\|_p \leq \liminf_{k\to\infty} \|f_m - f_{n_k}\|_p \, .$$

This implies that $\lim_{m\to\infty} \|f_m - f\|_p = 0$, and Minkowski inequality gives that $f \in L^p$. □

We give an elementary criterion for the L^p convergence for a.e. converging sequences.

Lemma 1.51. *Let $p \in [1, +\infty)$. Let $(f_n, n \in \mathbb{N})$ be a sequence of elements of L^p which converges a.e. towards $f \in L^p$. The convergence holds in L^p (i.e. $\lim_{n\to\infty} \|f - f_n\|_p = 0$) if and only if $\lim_{n\to\infty} \|f_n\|_p = \|f\|_p$.*

Proof. Assume that $\lim_{n\to\infty} \|f - f_n\|_p = 0$. Using Minkowski inequality, we deduce that $\left| \|f\|_p - \|f_n\|_p \right| \leq \|f - f_n\|_p$. This proves that $\lim_{n\to\infty} \|f_n\|_p = \|f\|_p$.

On the other hand, assume that $\lim_{n\to\infty} \|f_n\|_p = \|f\|_p$. Set $g_n = 2^{p-1}(|f_n|^p + |f|^p)$ and $g = 2^p |f|^p$. Since the function $x \mapsto |x|^p$ is convex, we get $|f_n - f|^p \leq g_n$ for all $n \in \mathbb{N}$. We also have $\lim_{n\to\infty} g_n = g$ a.e. and $\lim_{n\to\infty} \int g_n \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu < +\infty$. The dominated convergence Theorem 1.46 gives then that $\lim_{n\to\infty} \int |f_n - f|^p \, \mathrm{d}\mu = \int \lim_{n\to\infty} |f_n - f|^p \, \mathrm{d}\mu = 0$. This ends the proof. □

1.2.4 Fubini theorem

Let $(E, \mathcal{E}, \nu)$ and $(S, \mathcal{S}, \mu)$ be two measured spaces. The product space $E \times S$ is endowed with the product σ-field $\mathcal{E} \otimes \mathcal{S}$. We give a preliminary lemma.

Lemma 1.52. *Assume that ν and μ are σ-finite measures. Let f be a real-valued measurable function defined on $E \times S$.*

(i) For all $x \in E$, the function $y \mapsto f(x, y)$ defined on S is measurable and for all $y \in S$, the function $x \mapsto f(x, y)$ defined on E is measurable.

(ii) Assume that $f \geq 0$. The function $x \mapsto \int f(x, y) \, \mu(\mathrm{d}y)$ defined on E is measurable and the function $y \mapsto \int f(x, y) \, \nu(\mathrm{d}x)$ defined on S is measurable.

Proof. It is not difficult to check that the set $\mathcal{A} = \{C \in \mathcal{E} \otimes \mathcal{F};\, \mathbf{1}_C$ satisfies (i) and (ii)$\}$ is a λ-system, thanks to Corollary 1.23 and Proposition 1.24. (Hint: consider first the case μ and ν finite, and then extend to the case that μ and ν σ-finite, to prove that $\mathcal{A}$ satisfies property (ii) from Definition 1.12 of a λ-system.) Since $\mathcal{A}$ trivially contains $\mathcal{C} = \{A \times B;\, A \in \mathcal{E}$ and $B \in \mathcal{S}\}$ which is stable by finite intersection, we deduce from the monotone class theorem that $\mathcal{A}$ contains $\sigma(\mathcal{C}) = \mathcal{E} \otimes \mathcal{S}$. We deduce that (i) holds for any real-valued measurable functions, as they are limits of difference of simple functions, see the comments after Definition 1.34. We also deduce that (ii) holds for every simple function, and then for every $[0, +\infty]$-valued measurable functions thanks to Proposition 1.24 and the dominated convergence theorem. □

The next theorem allows to define the integral of a real-valued function with respect to the product of σ-finite[5] measures.

Theorem 1.53 (Fubini's theorem)**.** *Assume that ν and μ are σ-finite measures.*

(i) There exists a unique measure on $(E \times S, \mathcal{E} \otimes \mathcal{S})$, denoted by $\nu \otimes \mu$ and called product

[5]When the measures ν and μ are not σ-finite, the Fubini's theorem may fail essentially because the product measure might not be well defined. Consider the measurable space $([0,1], \mathcal{B}([0,1]))$ with λ the Lebesgue measure and μ the counting measure (which is not σ-finite), and the measurable function $f \geq 0$ defined by $f(x, y) = \mathbf{1}_{\{x=y\}}$, so that $\int \left(\int f(x, y) \, \mu(\mathrm{d}y) \right) \, \lambda(\mathrm{d}x) = 1$ and $\int \left(\int f(x, y) \, \lambda(\mathrm{d}x) \right) \, \mu(\mathrm{d}y) = 0$ are not equal.

measure such that:

$$\nu \otimes \mu(A \times B) = \nu(A)\mu(B) \quad \text{for all } A \in \mathcal{E},\ B \in \mathcal{S}. \tag{1.5}$$

(ii) Let f be a $[0, +\infty]$-valued measurable function defined on $E \times S$. We have:

$$\int f(x, y)\ \nu \otimes \mu(\mathrm{d}x, \mathrm{d}y) = \int \left(\int f(x, y)\ \mu(\mathrm{d}y) \right)\ \nu(\mathrm{d}x) \tag{1.6}$$

$$= \int \left(\int f(x, y)\ \nu(\mathrm{d}x) \right)\ \mu(\mathrm{d}y). \tag{1.7}$$

Let f be a real-valued measurable function defined $E \times S$. If $\nu \otimes \mu(f)$ is well defined, then the equalities (1.6) and (1.7) hold with their right hand-side being well defined.

We shall write $\nu(\mathrm{d}x)\mu(\mathrm{d}y)$ for $\nu \otimes \mu(\mathrm{d}x, \mathrm{d}y)$. If ν and μ are probabilities measures, then the definition of the product measure $\nu \otimes \mu$ coincide with the one given in Proposition 7.7.

Proof. For all $C \in \mathcal{E} \otimes \mathcal{S}$, we set $\nu \otimes \mu(C) = \int \left(\int \mathbf{1}_C(x, y)\ \mu(\mathrm{d}y) \right)\ \nu(\mathrm{d}x)$. The σ-additivity of ν and μ and the dominated convergence implies that $\nu \otimes \mu$ is a measure on $(E \times S, \mathcal{E} \otimes \mathcal{S})$. It is clear that (1.5) holds. Since ν and μ are σ-finite, we deduce that $\nu \otimes \mu$ is σ-finite. Using Exercise 8.2 based on the monotone class theorem and that $\{A \times B;\ A \in \mathcal{E}, B \in \mathcal{S}\}$ generates $\mathcal{E} \otimes \mathcal{S}$, we get that (1.5) characterizes uniquely the measure $\nu \otimes \mu$. This ends the proof of property (i).

Property (ii) holds clearly for functions $f = \mathbf{1}_C$ with $C = A \times B$, $A \in \mathcal{E}$ and $B \in \mathcal{S}$. Exercise 8.2, Corollary 1.23, Proposition 1.24, the monotone convergence theorem and the monotone class theorem imply that (1.6) and (1.7) hold also for $f = \mathbf{1}_C$ with $C \in \mathcal{E} \otimes \mathcal{S}$. We deduce that (1.6) and (1.7) hold for all simple functions thanks to Corollary 1.23, and then for all $[0, +\infty]$-valued measurable functions defined on $E \times S$ thanks to Proposition 1.24 and the monotone convergence theorem.

Let f be a real-valued measurable function defined $E \times S$ such that $\nu \otimes \mu(f)$ is well defined. Without loss of generality, we can assume that $\nu \otimes \mu(f^+)$ is finite. We deduce from (1.6) and then (1.7) with f replaced by f^+ that $N_E = \{x \in E;\ \int f(x, y)^+\ \mu(\mathrm{d}y) = +\infty\}$ is ν-null, and then that $N_S = \{y \in S;\ \int f(x, y)^+\ \nu(\mathrm{d}x) = +\infty\}$ is μ-null. We set $g = f^+ \mathbf{1}_{N_E^c \times N_S^c}$. It is now legitimate to subtract (1.6) with f replaced by f^- to (1.6) with f replaced by g in order to get (1.6) with f replaced by $g - f^-$. Since $\nu \otimes \mu$-a.e. $f^+ = g$ and thus $f = g - f^-$, Lemma 1.38 implies that (1.6) holds. Equality (1.7) is deduced by symmetry. □

Remark 1.54. Notice that the proof of (i) of Fubini theorem gives an alternative construction of the product of two σ-finite measures to the one given in Proposition 7.7 for the product of probability measures. Thanks to (i) of Fubini theorem, the Lebesgue measure on $\mathbb{R}^d$ can be seen as the product measure of d times the one-dimensional Lebesgue measure. ◊

1.2.5 Expectation, variance, covariance and inequalities

We consider the particular case of probability measure. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let X be a real-valued random variable. The expectation of X is by definition the integral of X with respect to the probability measure $\mathbb{P}$ and is denoted by $\mathbb{E}[X] = \int X(\omega)\,\mathbb{P}(\mathrm{d}\omega)$. We recall the expectation $\mathbb{E}[X]$ is well defined if $\min(\mathbb{E}[X^+], \mathbb{E}[X^-])$ is finite, where $X^+ = X \vee 0$ and $X^- = (-X) \vee 0$, and that X is integrable if $\max(\mathbb{E}[X^+], \mathbb{E}[X^-])$ is finite.

Example 1.55. If A is an event, then $\mathbf{1}_A$ is a random variable and we have $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$. Taking $A = \Omega$, we get obviously that $\mathbb{E}[\mathbf{1}_\Omega] = \mathbb{E}[\mathbf{1}] = 1$. △

The next elementary lemma is very useful to compute expectation in practice. Recall the distribution of X, denoted by P_X, has been introduced in Definition 1.27.

Lemma 1.56. *Let X be a random variable taking values in a measured space $(E, \mathcal{E})$. Let φ be a real-valued measurable function defined on $(E, \mathcal{E})$. If $\mathbb{E}[\varphi(X)]$ is well defined, or equivalently if $\int \varphi(x)\,\mathrm{P}_X(\mathrm{d}x)$ is well defined, then we have $\mathbb{E}[\varphi(X)] = \int \varphi(x)\,\mathrm{P}_X(\mathrm{d}x)$.*

Proof. Assume that φ is simple: $\varphi = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$ for some $n \in \mathbb{N}^*$, $\alpha_k \in [0, +\infty]$, $A_k \in \mathcal{F}$. We have:

$$\mathbb{E}[\varphi(X)] = \sum_{k=1}^n \alpha_k \mathbb{P}(X \in A_k) = \sum_{k=1}^n \alpha_k \mathrm{P}_X(A_k) = \int \varphi(x)\,\mathrm{P}_X(\mathrm{d}x).$$

Then use the monotone convergence theorem to get $\mathbb{E}[\varphi(X)] = \int \varphi(x)\,\mathrm{P}_X(\mathrm{d}x)$ when φ is measurable and $[0, +\infty]$-valued. Use the definition of $\mathbb{E}[\varphi(X)]$ and $\int \varphi(x)\,\mathrm{P}_X(\mathrm{d}x)$, when they are well defined, to conclude when φ is real-valued and measurable. □

Obviously, if X and Y have the same distribution, then $\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)]$ for all real-valued function φ such that $\mathbb{E}[\varphi(X)]$ is well defined, in particular if φ is bounded.

Remark 1.57. We give a closed formula for the expectation of discrete random variable. Let X be a random variable taking values in a measurable space $(E, \mathcal{E})$. We say that X is a discrete random variable if $\{x\} \in \mathcal{E}$ for all $x \in E$ and $\mathbb{P}(X \in \Delta_X) = 1$, where $\Delta_X = \{x \in E;\, \mathbb{P}(X = x) > 0\}$ is the discrete support of the distribution of X. Notice that Δ_X is at most countable and thus belongs to $\mathcal{E}$.

If X is a discrete random variable and φ a $[0, +\infty]$-valued function defined on E, then we have:

$$\mathbb{E}[\varphi(X)] = \sum_{x \in \Delta_X} \varphi(x)\mathbb{P}(X = x). \tag{1.8}$$

Equation (1.8) also holds for φ a real-valued measurable function as soon as $\mathbb{E}[\varphi(X)]$ is well defined. ◊

Remark 1.58. Let $B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$. By considering the probability measure $\frac{1}{\mathbb{P}(B)}\,\mathbf{1}_B\,\mathbb{P} : A \mapsto \mathbb{P}(A \cap B)/\mathbb{P}(B)$, see Corollary 1.44, we can define the expectation conditionally on B by, for all real-valued random variable Y such that $\mathbb{E}[Y]$ is well defined:

$$\mathbb{E}[Y|B] = \frac{\mathbb{E}[Y\mathbf{1}_B]}{\mathbb{P}(B)}. \tag{1.9}$$

If furthermore $\mathbb{P}(B) < 1$, then we easily get that $\mathbb{E}[Y] = \mathbb{P}(B)\mathbb{E}[Y|B] + \mathbb{P}(B^c)\mathbb{E}[Y|B^c]$. ◊

A real-valued random variable X is square-integrable if it belongs to $L^2(\mathbb{P})$. Since $2\,|x| \leq 1 + |x|^2$, we deduce from the monotonicity property of the expectation that if $X \in L^2(\mathbb{P})$ then $X \in L^1(\mathbb{P})$, that is X is integrable. This means that $L^2(\mathbb{P}) \subset L^1(\mathbb{P})$.

For $X = (X_1, \ldots, X_d)$ and $\mathbb{R}^d$-valued random variable, we say that $\mathbb{E}[X]$ is well defined (resp. X_i is integrable, resp. square integrable) if $\mathbb{E}[X_i]$ is well defined (resp. X_i is integrable, resp. square integrable) for all $i \in [\![1, d]\!]$, and we set $\mathbb{E}[X] = (\mathbb{E}[X_1], \ldots, \mathbb{E}[X_d])$.

We recall an $\mathbb{R}$-valued function φ defined on $\mathbb{R}^d$ is convex if $\varphi(qx + (1-q)y) \leq q\varphi(x) + (1-q)\varphi(y)$ for all $x, y \in \mathbb{R}^d$ and $q \in (0,1)$. The function φ is strictly convex if this convex inequality is strict for all $x \neq y$. Let $\langle \cdot, \cdot \rangle$ denote the scalar product of $\mathbb{R}^d$. Then, it is well known that if φ is an $\mathbb{R}$-valued convex function defined on $\mathbb{R}^d$, then it is continuous[6] and there exists[7] a sequence $((a_n, b_n), n \in \mathbb{N})$ with $a_n \in \mathbb{R}^d$ and $b_n \in \mathbb{R}$ such that for all $x \in \mathbb{R}^d$:

$$\varphi(x) = \sup_{n \in \mathbb{N}} \left(b_n + \langle a_n, x \rangle \right). \tag{1.10}$$

We give further inequalities which complete Proposition 1.49. We recall that a $\mathbb{R}$-valued convex function defined on $\mathbb{R}^d$ is continuous (and thus measurable).

Proposition 1.59.

- ***Tchebychev inequality.*** *Let X be real-valued random variable. Let $a > 0$. We have:*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[X^2]}{a^2}.$$

- ***Jensen inequality.*** *Let X be an $\mathbb{R}^d$-valued integrable random variable. Let φ be a $\mathbb{R}$-valued convex function defined on $\mathbb{R}^d$. We have that $\mathbb{E}[\varphi(X)]$ is well defined and:*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]. \tag{1.11}$$

  Furthermore, if φ is strictly convex, the inequality in (1.11) is an equality if and only if X is a.s. constant.

Remark 1.60. If X is a real-valued integrable random variable, we deduce from Cauchy-Schwarz inequality or Jensen inequality that $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$. $\Diamond$

Proof. Since $\mathbf{1}_{\{|X| \geq a\}} \leq X^2/a^2$, we deduce the Tchebychev inequality from the monotonicity property of the expectation and Example 1.55.

Let φ be a real-valued convex function. Using (1.10), we get $\varphi(X) \geq b_0 + \langle a_0, X \rangle$ and thus $\varphi(X) \geq -|b_0| - |a_0||X|$. Since X is integrable, we deduce that $\mathbb{E}[\varphi(X)^-] < +\infty$, and thus $\mathbb{E}[\varphi(X)]$ is well defined. Then, using the monotonicity of the expectation, we get that for all $n \in \mathbb{N}$, $\mathbb{E}[\varphi(X)] \geq b_n + \langle a_n, \mathbb{E}[X] \rangle$. Taking the supremum over all $n \in \mathbb{N}$ and using the characterization (1.10), we get (1.11).

[6] It is enough to prove the continuity at 0 and without loss of generality, we can assume that $\varphi(0) = 0$. Since φ is finite on the 2^d vertices of the cube $[-1,1]^d$, it is bounded from above by a finite constant, say M. Using the convex inequality, we deduce that φ is bounded on $[-1,1]^d$ by M. Let $\alpha \in (0,1)$ and $y \in [-\alpha, \alpha]^d$. Using the convex inequality with $x = y/\alpha$, $y = 0$ and $q = \alpha$, we get that $\varphi(y) \leq \alpha\varphi(y/\alpha) \leq \alpha M$. Using the convex inequality with $x = y$, $y = -y/\alpha$ and $q = 1/(1+\alpha)$, we also get that $0 \leq \varphi(y)/(1+\alpha) + M\alpha/(1+\alpha)$. This gives that $|\varphi(y)| \leq \alpha M$. Thus φ is continuous at 0.

[7] This is a consequence of the separation theorem for convex sets. See for example Proposition B.1.2.1 in J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Springer-Verlag, 2001.

To complete the proof, we shall check that if X is not equal a.s. to a constant and if φ is strictly convex, then the inequality in (1.11) is strict. For simplicity, we consider the case $d = 1$ as the case $d \geq 2$ can be proved similarly. Set $B = \{X \leq \mathbb{E}[X]\}$. Since X is non-constant, we deduce that $\mathbb{P}(B) \in (0, 1)$ and that $\mathbb{E}[X|B] < \mathbb{E}[X|B^c]$. Recall that $\mathbb{E}[X] = \mathbb{P}(B)\mathbb{E}[X|B] + \mathbb{P}(B^c)\mathbb{E}[X|B^c]$. We get that:

$$\begin{aligned}\varphi(\mathbb{E}[X]) &< \mathbb{P}(B)\varphi(\mathbb{E}[X|B]) + \mathbb{P}(B^c)\varphi(\mathbb{E}[X|B^c])\\ &\leq \mathbb{P}(B)\mathbb{E}[\varphi(X)|B] + \mathbb{P}(B^c)\mathbb{E}[\varphi(X)|B^c] = \mathbb{E}[\varphi(X)],\end{aligned}$$

where we used the strict convexity of φ and that $\mathbb{E}[X|B] \neq \mathbb{E}[X|B^c]$ for the first inequality and Jensen inequality for the second. This proves that the inequality in (1.11) is strict. □

We end this section with the covariance and variance. Let X, Y be two real-valued square-integrable random variables. Thanks to Cauchy-Schwarz inequality, XY is integrable. The covariance of X and Y, $\mathrm{Cov}(X, Y)$, and the variance of X, $\mathrm{Var}(X)$, are defined by:

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \quad \text{and} \quad \mathrm{Var}(X) = \mathrm{Cov}(X, X).$$

By linearity, we also get that:

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] \quad \text{and} \quad \mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y). \tag{1.12}$$

The covariance is a bilinear form on $L^2(\mathbb{P})$ and for $a, b \in \mathbb{R}$, we get:

$$\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X).$$

Using Lemma 1.38 with $f = (X - \mathbb{E}[X])^2$, we get that $\mathrm{Var}(X) = 0$ implies X is a.s. constant.

The covariance can be defined for random vectors as follows.

Definition 1.61. *Let $X = (X_1, \ldots, X_d)$ and $Y = (Y_1, \ldots, Y_p)$ be respectively two $\mathbb{R}^d$-valued and $\mathbb{R}^p$-valued square-integrable random variables with $d, p \in \mathbb{N}^*$. The covariance matrix of X and Y, $\mathrm{Cov}(X, Y)$, is a $d \times p$ matrix defined by:*

$$\mathrm{Cov}(X, Y) = (\mathrm{Cov}(X_i, Y_j),\ i \in [\![1, d]\!] \text{ and } j \in [\![1, p]\!]).$$

1.2.6 Independence

Recall the independence of σ-fields given in Definition 1.11 and of random variables given in Definition 1.31.

Proposition 1.62. *Let $((E_i, \mathcal{E}_i), i \in I)$ be a collection of measurable spaces and $(X_i, i \in I)$ a random variable taking values in the product space $\prod_{i\in I} E_i$ endowed with the product σ-field. The random variables $(X_i, i \in I)$ are independent if and only if for all finite subset $J \subset I$, for all bounded real-valued measurable function f_j defined on E_j for $j \in J$, we have:*

$$\mathbb{E}\left[\prod_{j\in J} f_j(X_j)\right] = \prod_{j=1}^{n} \mathbb{E}[f_j(X_j)]. \tag{1.13}$$

Proof. If (1.13) is true, then taking $f_j = \mathbf{1}_{A_j}$ with $A_j \in \mathcal{E}_j$, we deduce from Definition 1.31 that the random variables $(X_i, i \in I)$ are independent.

If $(X_i, i \in I)$ are independent, then Definitions 1.31 implies that (1.13) holds for indicator functions. By linearity, we get that (1.13) holds also for simple functions. Use monotone convergence theorem and then linearity to deduce that (1.13) holds for bounded real-valued measurable functions. $\square$

Bibliography

[1] O. Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[2] J. Neveu. *Bases mathématiques du calcul des probabilités.* Masson et Cie, Éditeurs, Paris, 1970.

Chapter 2

Conditional expectation

Let X be a square integrable real-valued random variable. The constant c which minimizes $\mathbb{E}[(X-c)^2]$ is the expectation of X. Indeed, we have, with $m = \mathbb{E}[X]$:

$$\mathbb{E}[(X-c)^2] = \mathbb{E}[(X-m)^2 + (m-c)^2 + 2(X-m)(m-c)] = \mathrm{Var}(X) + (m-c)^2.$$

In some sense, the expectation of X is the best approximation of X by a constant (with a quadratic criterion).

More generally, the conditional expectation of X given another random variable Y will be defined as the best approximation of X by a function of Y. In Section 2.1, we define the conditional expectation of a square integrable random variable as a projection. In Section 2.2, we extend the conditional expectation to random variables whose expectations are well defined. In Section 2.3, we provide explicit formulas for discrete and continuous random variables.

We shall consider that all the random variables of this chapter are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Recall that the normed vector space $L^2 = L^2(\Omega, \mathcal{F}, \mathbb{P})$ denote the set of (equivalent classes of) square integrable real-valued random variables.

2.1 Projection in the L^2 space

The bilinear form $\langle \cdot, \cdot \rangle_{L^2}$ on L^2 defined by $\langle X, Y \rangle_{L^2} = \mathbb{E}[XY]$ is the scalar product corresponding to the norm $\|\cdot\|_2$. The space L^2 with the product scalar $\langle \cdot, \cdot \rangle_{L^2}$ is an Hilbert space, as it is complete, thanks to Proposition 1.50. Notice that square-integrable real-valued random variables which are independent and centered are orthogonal for the scalar product $\langle \cdot, \cdot \rangle_{L^2}$.

We shall consider the following results on projection in Hilbert spaces.

Theorem 2.1. *Let H be a closed vector sub-space of L^2 and $X \in L^2$.*

(i) (Existence.) There exists a real-valued random variable $X_H \in H$, called the orthogonal projection of X on H, such that $\mathbb{E}[(X - X_H)^2] = \inf\{\mathbb{E}[(X-Y)^2];\, Y \in H\}$. And, for all $Y \in H$, we have $\mathbb{E}[XY] = \mathbb{E}[X_H Y]$.

(ii) (Uniqueness.) Let $Z \in H$ such that $\mathbb{E}[(X-Z)^2] = \inf\{\mathbb{E}[(X-Y)^2];\, Y \in H\}$ or such that $\mathbb{E}[ZY] = \mathbb{E}[XY]$ for all $Y \in H$. Then, we have that a.s. $Z = X_H$.

Proof. We set $a = \inf\{\mathbb{E}[(X-Y)^2]; Y \in H\}$. The following median formula is clear:

$$\mathbb{E}[(Z'-Y')^2] + \mathbb{E}[(Z'+Y')^2] = 2\mathbb{E}[Z'^2] + 2\mathbb{E}[Y'^2] \quad \text{for all } Y', Z' \in L^2.$$

Let $(X_n, n \in \mathbb{N})$ be a sequence of H such that $\lim_{n\to+\infty} \mathbb{E}[(X-X_n)^2] = a$. Using the median formula with $Z' = X_n - X$ and $Y' = X_m - X$, we get:

$$\mathbb{E}[(X_n - X_m)^2] = 2\mathbb{E}[(X-X_n)^2] + 2\mathbb{E}[(X-X_m)^2] - 4\mathbb{E}[(X-I)^2],$$

with $I = (X_n + X_m)/2 \in H$. As $\mathbb{E}[(X-I)^2] \geq a$, we deduce that the sequence $(X_n, n \in \mathbb{N})$ is a Cauchy sequence in L^2 and thus converge in L^2, say towards X_H. In particular, we have $\mathbb{E}[(X-X_H)^2] = a$. Since H is closed, we get that the limit X_H belongs to H.

Let $Z \in H$ be such that $\mathbb{E}[(X-Z)^2] = a$. For $Y \in H$, the function $t \mapsto \mathbb{E}[(X-Z+tY)^2] = a + 2t\mathbb{E}[(X-Z)Y] + t^2\mathbb{E}[Y^2]$ is minimal for $t = 0$. This implies that its derivative at $t = 0$ is zero, that is $\mathbb{E}[(X-Z)Y] = 0$. In particular, we have $\mathbb{E}[(X-X_H)Y] = 0$. This ends the proof of (i).

On the one hand, let $Z \in H$ be such that $\mathbb{E}[(X-Z)^2] = a$. We deduce from the previous arguments that for all $Y \in H$:

$$\mathbb{E}[(X_H - Z)Y] = \mathbb{E}[(X-Z)Y] - \mathbb{E}[(X-X_H)Y] = 0.$$

Taking $Y = (X_H - Z)$, gives that $\mathbb{E}[(X_H - Z)^2] = 0$ and thus a.s. $Z = X_H$, see Lemma 1.38.

On the other hand, if there exists $Z \in H$ such that $\mathbb{E}[ZY] = \mathbb{E}[XY]$ for all $Y \in H$, arguing as above, we also deduce that a.s. $Z = X_H$. □

According to the remarks at the beginning of this chapter, we see that if X is a real-valued square-integrable random variable, then $\mathbb{E}[X]$ can be seen as the orthogonal projection of X on the vector space of the constant random variables.

2.2 Conditional expectation

Let $\mathcal{H} \subset \mathcal{F}$ be a σ-field. We recall that a random variable Y (which is by definition $\mathcal{F}$-measurable) is $\mathcal{H}$-measurable if $\sigma(Y)$, the σ-field generated by Y, is a subset of $\mathcal{H}$. The expectation of X conditionally on $\mathcal{H}$ corresponds intuitively to the best "approximation" of X by an $\mathcal{H}$-measurable random variable.

Notice that if X is a real-valued random variable such that $\mathbb{E}[X]$ is well defined, then $\mathbb{E}[X\mathbf{1}_A]$ is also well defined for any $A \in \mathcal{F}$.

Definition 2.2. *Let X be a real-valued random variable such that $\mathbb{E}[X]$ is well defined. We say that a real-valued $\mathcal{H}$-measurable random variable Z, such that $\mathbb{E}[Z]$ is well defined, is the expectation of X conditionally on $\mathcal{H}$ if:*

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A] \quad \textit{for all } A \in \mathcal{H}. \tag{2.1}$$

The next lemma asserts that, if the expectation of X conditionally on $\mathcal{H}$ exists then it is unique up to an a.s. equality. It will be denoted by $\mathbb{E}[X|\mathcal{H}]$.

Lemma 2.3 (Uniqueness of the conditional expectation)**.** *Let Z and Z' be real-valued random variables, $\mathcal{H}$-measurable with $\mathbb{E}[Z]$ and $\mathbb{E}[Z']$ well defined, and such that $\mathbb{E}[Z\mathbf{1}_A] = \mathbb{E}[Z'\mathbf{1}_A]$ for all $A \in \mathcal{H}$. Then, we get that a.s. $Z = Z'$.*

Proof. Let $n \in \mathbb{N}^*$ and consider $A = \{n \geq Z > Z' \geq -n\}$ which belongs to $\mathcal{H}$. By linearity, we deduce from the hypothesis that $\mathbb{E}[(Z - Z')\mathbf{1}_{\{n \geq Z > Z' \geq -n\}}] = 0$. Lemma 1.38 implies that $\mathbb{P}(n \geq Z > Z' \geq -n) = 0$ and thus $\mathbb{P}(+\infty > Z > Z' > -\infty) = 0$ by monotone convergence. Considering $A = \{Z = +\infty, n \geq Z'\}$, $A = \{Z \geq n, Z' = -\infty\}$ and $A = \{Z = +\infty, Z' = -\infty\}$ leads similarly to $\mathbb{P}(Z > Z', Z = +\infty \text{ or } Z' = -\infty) = 0$. So we get $\mathbb{P}(Z > Z') = 0$. By symmetry, we deduce that a.s. $Z = Z'$. □

We use the orthogonal projection theorem on Hilbert spaces, to define the conditional expectation for square-integrable real-valued random variables.

Proposition 2.4. *If $X \in L^2$, then $\mathbb{E}[X|\mathcal{H}]$ is the orthogonal projection defined in Proposition 2.1, of X on the vector space H of all square-integrable $\mathcal{H}$-measurable random variables.*

Proof. The set H corresponds to the space $L^2(\Omega, \mathcal{H}, \mathbb{P})$. It is closed thanks to Proposition 1.50. The set H is thus a closed vector subspace of L^2. Property (i) (with $Y = \mathbf{1}_A$) from Theorem 2.1 implies then that (2.1) holds and thus that the orthogonal projection of $X \in L^2$ on H is the expectation of X conditionally on $\mathcal{H}$. □

Notice that for $A \in \mathcal{F}$, we have $\mathbf{1}_A \in L^2$, and we shall use the notation:

$$\mathbb{P}(A|\mathcal{H}) = \mathbb{E}\left[\mathbf{1}_A \middle| \mathcal{H}\right]. \tag{2.2}$$

We have the following properties.

Proposition 2.5. *Let X and Y be real-valued square-integrable random variables.*

(i) Positivity. If a.s. $X \geq 0$ then we have that a.s. $\mathbb{E}[X|\mathcal{H}] \geq 0$.

(ii) Linearity. For $a, b \in \mathbb{R}$, we have that a.s. $\mathbb{E}[aX + bY|\mathcal{H}] = a\mathbb{E}[X|\mathcal{H}] + b\mathbb{E}[Y|\mathcal{H}]$.

(iii) Monotone convergence. Let $(X_n, n \in \mathbb{N})$ be a sequence of real-valued square integrable random variables such that for all $n \in \mathbb{N}$ a.s. $0 \leq X_n \leq X_{n+1}$. Then, we have that a.s.:

$$\lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{H}] = \mathbb{E}\left[\lim_{n\to+\infty} X_n \middle| \mathcal{H}\right].$$

Proof. Let X be a square-integrable a.s. non-negative random variable. We set $A = \{\mathbb{E}[X|\mathcal{H}] < 0\}$. We have:

$$0 \geq \mathbb{E}[\mathbb{E}[X|\mathcal{H}]\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A] \geq 0,$$

where we used that $A \in \mathcal{H}$ and (2.1) for the equality. We deduce that $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]\mathbf{1}_A] = 0$ and thus that a.s. $\mathbb{E}[X|\mathcal{H}] \geq 0$ according to Lemma 1.38.

The linearity property is a consequence of the linearity property of the expectation, (2.1) and Lemma 2.3.

Let $(X_n, n \in \mathbb{N})$ be a sequence of real-valued square-integrable random variables such that for all $n \in \mathbb{N}$ a.s. $0 \leq X_n \leq X_{n+1}$. We deduce from the linearity and positivity properties of the conditional expectation that for all $n \in \mathbb{N}$ a.s. $0 \leq \mathbb{E}[X_n|\mathcal{H}] \leq \mathbb{E}[X_{n+1}|\mathcal{H}]$. The random-variable $Z = \lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{H}]$ is $\mathcal{H}$-measurable according to Corollary 1.42 and a.s. non-negative. The monotone convergence theorem implies that for all $A \in \mathcal{H}$:

$$\mathbb{E}\left[Z\mathbf{1}_A\right] = \lim_{n\to+\infty} \mathbb{E}\left[\mathbb{E}[X_n|\mathcal{H}]\mathbf{1}_A\right] = \lim_{n\to+\infty} \mathbb{E}\left[X_n\mathbf{1}_A\right] = \mathbb{E}\left[\lim_{n\to+\infty} X_n\mathbf{1}_A\right].$$

Deduce from (2.1) and Lemma 2.3 that a.s. $Z = \mathbb{E}[\lim_{n\to+\infty} X_n|\mathcal{H}]$. This ends the proof. □

We extend the definition of conditional expectations to random variables whose expectation is well defined.

Proposition 2.6. *Let X be a real-valued random variable such that $\mathbb{E}[X]$ is well defined. Then its expectation conditionally on $\mathcal{H}$, $\mathbb{E}[X|\mathcal{H}]$, exists and is unique up to an a.s. equality. Furthermore, the expectation of $\mathbb{E}[X|\mathcal{H}]$ is well defined and is equal to $\mathbb{E}[X]$:*

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]\right] = \mathbb{E}[X]. \tag{2.3}$$

If X is non-negative a.s. (resp. integrable), so is $\mathbb{E}[X|\mathcal{H}]$.

Proof. Consider first the case where X is a.s. non-negative. The random variable X is the a.s. limit of a sequence of positive square-integrable random variables. Property (iii) from Proposition 2.5 implies that $\mathbb{E}[X|\mathcal{H}]$ exists. It is unique thanks to Lemma 2.3. It is a.s. non-negative as limit of non-negative random variables. Taking $A = \Omega$ in (2.1), we get (2.3).

We now consider the general case. Recall that $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$. From the previous argument the expectations of $\mathbb{E}[X^+|\mathcal{H}]$ and $\mathbb{E}[X^-|\mathcal{H}]$ are well defined and respectively equal to $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$. Since one of those two expectations is finite, we deduce that a.s. $\mathbb{E}[X^+|\mathcal{H}]$ if finite or a.s. $\mathbb{E}[X^-|\mathcal{H}]$ is finite. Then use (2.1) and Lemma 2.3 to deduce that $\mathbb{E}[X^+|\mathcal{H}] - \mathbb{E}[X^-|\mathcal{H}]$ is equal to $\mathbb{E}[X|\mathcal{H}]$, the expectation of X conditionally on $\mathcal{H}$. Since $\mathbb{E}[X|\mathcal{H}]$ is the difference of two non-negative random variables, one of them being integrable, we deduce that the expectation of $\mathbb{E}[X|\mathcal{H}]$ is well defined and use (2.1) with $A = \Omega$ to get (2.3). Eventually, if X is integrable, so are $\mathbb{E}[X^+|\mathcal{H}]$ and $\mathbb{E}[X^-|\mathcal{H}]$ thanks to (2.3) for non-negative random variables. This implies that $\mathbb{E}[X|\mathcal{H}]$ is integrable. □

We summarize in the next proposition the properties of the conditional expectation directly inherited from the properties of the expectation.

Proposition 2.7. *We have the following properties.*

(i) Positivity. If X is a real-valued random variable such that a.s. $X \geq 0$, then a.s. $\mathbb{E}[X|\mathcal{H}] \geq 0$.

(ii) Linearity. For a, b in $\mathbb{R}$ (resp. in $[0, +\infty)$), X, Y real-valued random-variables with X and Y integrable (resp. with $\mathbb{E}[X^+ + Y^+]$ or $\mathbb{E}[X^- + Y^-]$ finite), we have $\mathbb{E}[aX + bY|\mathcal{H}] = a\mathbb{E}[X|\mathcal{H}] + b\mathbb{E}[Y|\mathcal{H}]$.

(iii) Monotony. For X, Y real-valued random variables such that a.s. $X \leq Y$ and $\mathbb{E}[X]$ as well as $\mathbb{E}[Y]$ are well defined, we have $\mathbb{E}[X|\mathcal{H}] \leq \mathbb{E}[Y|\mathcal{H}]$.

(iv) *Monotone convergence. Let* $(X_n, n \in \mathbb{N})$ *be real-valued random variables such that for all* $n \in \mathbb{N}$ *a.s.* $0 \leq X_n \leq X_{n+1}$. *Then we have that a.s.:*

$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{H}] = \mathbb{E}\left[\lim_{n\to\infty} X_n\middle|\mathcal{H}\right].$$

(v) *Fatou Lemma. Let* $(X_n, n \in \mathbb{N})$ *be real-valued random variables such that for all* $n \in \mathbb{N}$ *a.s.* $0 \leq X_n$. *Then we have that a.s.:*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\middle|\mathcal{H}\right] \leq \liminf_{n\to\infty} \mathbb{E}\left[X_n|\mathcal{H}\right].$$

(vi) *Dominated convergence (Lebesgue). Let* X, Y, $(X_n, n \in \mathbb{N})$ *be real-valued random variables such that a.s.* $\lim_{n\to\infty} X_n = X$, *for all* $n \in \mathbb{N}$ *a.s.* $|X_n| \leq Y$ *and* $\mathbb{E}[Y] < +\infty$. *Then we have that a.s.:*

$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{H}] = \mathbb{E}\left[\lim_{n\to\infty} X_n\middle|\mathcal{H}\right].$$

(vii) *The Tchebychev, Hölder, Cauchy-Schwarz, Minkowski and Jensen inequalities from Propositions 1.49 and 1.59 holds with the expectation replaced by the conditional expectation.*

For example, we state Jensen inequality from property (vii) above. Let φ be a $\mathbb{R}$-valued convex function defined on $\mathbb{R}^d$. Let X be an integrable $\mathbb{R}^d$-valued random variable. Then, $\mathbb{E}[\varphi(X)|\mathcal{H}]$ is well defined and a.s.:

$$\varphi(\mathbb{E}\left[X|\,\mathcal{H}\right]) \leq \mathbb{E}\left[\varphi(X)|\,\mathcal{H}\right]. \tag{2.4}$$

Furthermore, if φ is strictly convex, the inequality in (2.4) is an equality if and only if X is a.s. equal to an $\mathcal{H}$-measurable random variable.

Proof. The positivity property comes from Proposition 2.6. The linearity property comes from the linearity of the expectation, (2.1) and Lemma 2.3. The monotony property is a consequence of the positivity and linearity properties. The proof of the monotone convergence theorem is based on the same arguments as in the proof of Proposition 2.5. Fatou Lemma and the dominated convergence theorem are consequences of the monotone convergence theorem, see proof of Lemma 1.45 and of Theorem 1.46. The proofs of the inequalities are similar to the proofs of Propositions 1.49 and 1.59. (Be careful when characterizing the equality in (2.4) when φ is strictly convex.) □

Using the monotone or dominated convergence theorems, it is easy to prove the following Corollary which generalizes (2.1).

Corollary 2.8. *Let* X *and* Y *be two real-valued random variables such that* $\mathbb{E}[X]$ *and* $\mathbb{E}[XY]$ *are well defined, and* Y *is* $\mathcal{H}$*-measurable. Then* $\mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]Y\right]$ *is well defined and we have:*

$$\mathbb{E}[XY] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]Y\right]. \tag{2.5}$$

Recall Definitions 1.11 and 1.30 on independence. We complete the properties of the conditional expectation.

Proposition 2.9. *Let X be a real-valued random variable such that $\mathbb{E}[X]$ is well defined.*

(i) If X is $\mathcal{H}$-measurable, then we have that a.s. $\mathbb{E}[X|\mathcal{H}] = X$.

(ii) If X is independent of $\mathcal{H}$, then we have that a.s. $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$.

(iii) If Y is a real-valued $\mathcal{H}$-measurable random variable such that $\mathbb{E}[XY]$ is well defined, then we have that a.s. $\mathbb{E}[YX|\mathcal{H}] = Y\mathbb{E}[X|\mathcal{H}]$.

(iv) If $\mathcal{G} \subset \mathcal{H}$ is a σ-field, then we have that a.s. $\mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}\right] = \mathbb{E}[X|\mathcal{G}]$.

(v) If $\mathcal{G} \subset \mathcal{F}$ is a σ-field independent of $\mathcal{H}$ and independent of X (that is $\mathcal{G}$ is independent of $\mathcal{H} \vee \sigma(X)$), then we have that a.s. $\mathbb{E}[X|\mathcal{G} \vee \mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$.

Proof. If X is $\mathcal{H}$-measurable, then we can choose $Z = X$ in (2.1) and use Lemma 2.3 to get property (i). If X is independent of $\mathcal{H}$, then for all $A \in \mathcal{H}$, we have $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X]\mathbf{1}_A]$, and we can choose $Z = \mathbb{E}[X]$ in (2.1) and use Lemma 2.3 to get property (ii). If Y is a real-valued $\mathcal{H}$-measurable random variable such that $\mathbb{E}[XY]$ is well defined, then $\mathbb{E}[XY\mathbf{1}_A]$ is also well defined for $A \in \mathcal{H}$, and according to (2.5), we have $\mathbb{E}[XY\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]Y\mathbf{1}_A]$. Then, we can choose $Z = Y\mathbb{E}[X|\mathcal{H}]$ in (2.1), with X replaced by $X\mathbf{1}_A$, and use Lemma 2.3 to get property (iii).

We prove property (iv). Let $A \in \mathcal{G} \subset \mathcal{H}$. We have:

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A\right] = \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]\mathbf{1}_A\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}\right]\mathbf{1}_A\right],$$

where we used (2.1) with $\mathcal{H}$ replaced by $\mathcal{G}$ for the first equality, (2.1) for the second and (2.1) with $\mathcal{H}$ replaced by $\mathcal{G}$ and X by $\mathbb{E}[X|\mathcal{H}]$ for the last. Then we deduce property (iv) from Definition 2.2 and Lemma 2.3.

We prove property (v) first when X is integrable. For $A \in \mathcal{G}$ and $B \in \mathcal{H}$, we have:

$$\mathbb{E}\left[\mathbf{1}_{A\cap B}X\right] = \mathbb{E}\left[\mathbf{1}_A\mathbf{1}_B X\right] = \mathbb{E}[\mathbf{1}_A]\mathbb{E}[\mathbf{1}_B X] = \mathbb{E}[\mathbf{1}_A]\mathbb{E}[\mathbf{1}_B\mathbb{E}[X|\mathcal{H}]] = \mathbb{E}[\mathbf{1}_A\mathbf{1}_B\mathbb{E}[X|\mathcal{H}]],$$

where we used that $\mathbf{1}_A$ is independent of $\mathcal{H}\vee\sigma(X)$ in the second equality and independent of $\mathcal{H}$ in the last. Using the dominated convergence theorem, we get that $\mathcal{A} = \{A \in \mathcal{F},\, \mathbb{E}\left[\mathbf{1}_A X\right] = \mathbb{E}[\mathbf{1}_A\mathbb{E}[X|\mathcal{H}]]\}$ is a monotone class. It contains $\mathcal{C} = \{A \cap B;\ A \in \mathcal{G}, B \in \mathcal{H}\}$ which is stable by finite intersection. The monotone class theorem implies that $\mathcal{A}$ contains $\sigma(\mathcal{C})$ and thus $\mathcal{G} \vee \mathcal{H}$. Then we deduce property (v) from Definition 2.2 and Lemma 2.3. Use the monotone convergence theorem to extend the result to non-negative random variable and use that $\mathbb{E}[X|\mathcal{H}'] = \mathbb{E}[X^+|\mathcal{H}'] - \mathbb{E}[X^-|\mathcal{H}']$ for any σ-field $\mathcal{H}' \subset \mathcal{F}$ when $\mathbb{E}[X]$ is well defined, to extend the result to any real random variable X such that $\mathbb{E}[X]$ is well defined. □

We extend the definition of conditional expectation to $\mathbb{R}^d$-valued random variables.

Definition 2.10. *Let $d \in \mathbb{N}^*$. Let $X = (X_1, \ldots, X_d)$ be an $\mathbb{R}^d$-valued random variable such that $\mathbb{E}[X]$ is well defined. The conditional expectation of X conditionally on $\mathcal{H}$, denoted by $\mathbb{E}[X|\mathcal{H}]$, is given by $(\mathbb{E}[X_1|\mathcal{H}], \ldots, \mathbb{E}[X_d|\mathcal{H}])$.*

2.3 Conditional expectation with resp. to a random variable

Let V be a random variable taking values in a measurable space $(E, \mathcal{E})$. Recall that $\sigma(V)$ denote the σ-field generated by V. Let X be a real-valued random variable. We write $\mathbb{E}[X|V]$ for $\mathbb{E}[X|\sigma(V)]$. The next result states that $\mathbb{E}[X|V]$ is a measurable function of V. It is a direct consequence of Proposition 1.25.

Corollary 2.11. *Let V be a random variable taking values in a measurable space $(E, \mathcal{E})$ and X a real-valued random variable such that $\mathbb{E}[X]$ is well defined. There exists a real-valued measurable function g defined on E such that a.s. $\mathbb{E}[X|V] = g(V)$.*

In the next two paragraphs we give an explicit formula for g when V is a discrete random variable and when $X = \varphi(Y, V)$ with Y some random variable taking values in a measurable space $(S, \mathcal{S})$ such that (Y, V) has a density with respect to some product measure on $S \times E$.

Recall (2.2) for the notation $\mathbb{P}(A|\,\mathcal{H})$ for $A \in \mathcal{F}$; and we shall write $\mathbb{P}(A|V)$ for $\mathbb{P}(A|\,\sigma(V))$.

2.3.1 The discrete case

The following corollary provides an explicit formula for the expectation conditionally on a discrete random variable. Recall the definition of a discrete random variable in Remark 1.57 and of the expectation conditionally on an event in Remark 1.58.

Corollary 2.12. *Let $(E, \mathcal{E})$ be a measurable space and V be a discrete E-valued random variable. Let X be a real-valued random variable such that $\mathbb{E}[X]$ is well defined. Then, we have that a.s. $\mathbb{E}[X|V] = g(V)$ with:*

$$g(v) = \frac{\mathbb{E}[X\mathbf{1}_{\{V=v\}}]}{\mathbb{P}(V = v)} = \mathbb{E}[X|V = v] \quad \text{if } \mathbb{P}(V = v) > 0, \quad \text{and} \quad g(v) = 0 \quad \text{otherwise.} \tag{2.6}$$

Proof. According to Corollary 2.11, we have $\mathbb{E}[X|V] = g(V)$ for some real-valued measurable function g. We deduce from (2.1) with $A = \{V = v\}$ that $\mathbb{E}[X\mathbf{1}_{\{V=v\}}] = g(v)\mathbb{P}(V = v)$. If $\mathbb{P}(V = v) > 0$, we get:

$$g(v) = \frac{\mathbb{E}[X\mathbf{1}_{\{V=v\}}]}{\mathbb{P}(V = v)} = \mathbb{E}[X|V = v].$$

The value of $\mathbb{E}[X|V = v]$ when $\mathbb{P}(V = v) = 0$ is unimportant, and can be set equal to 0. □

Remark 2.13. Let $(E, \mathcal{E})$ be a measurable space and V be a discrete E-valued random variable with discrete support $\Delta_V = \{v \in E, \mathbb{P}(V = v) > 0\}$. For $v \in \Delta_V$, denote by $\mathbb{P}_v$ the probability measure on $(\Omega, \mathcal{F})$ defined by $\mathbb{P}_v(A) = \mathbb{P}(A|V = v)$ for $A \in \mathcal{F}$. The law of X conditionally on $\{V = v\}$, denoted by $\mathrm{P}_{X|v}$ is the image of the probability measure $\mathbb{P}_v$ by X, and we define the law of X conditionally on V as the collection of probability measure $\mathrm{P}_{X|V} = (\mathrm{P}_{X|v}, v \in \Delta_V)$. An illustration is given in the next example. ◊

Example 2.14. Let $(X_i, i \in [\![1, n]\!])$ be independent Bernoulli random variables with the same parameter $p \in (0, 1)$. We set $S_n = \sum_{i=1}^n X_i$, which has a binomial distribution with parameter (n, p). We shall compute the law of X_1 conditionally on S_n. We get for $k \in [\![1, n]\!]$:

$$\mathbb{P}(X_1 = 1|S_n = k) = \frac{\mathbb{P}(X_1 = 1, S_n = k)}{\mathbb{P}(S_n = k)} = \frac{\mathbb{P}(X_1 = 1)\mathbb{P}(X_2 + \cdots + X_n = k - 1)}{\mathbb{P}(S_n = k)} = \frac{k}{n},$$

where we used independence for X_1 and $(X_2, \ldots, X_n)$ for the second equality and that $X_2 + \cdots + X_n$ has a binomial distribution with parameter $(n-1, p)$ for the last. For $k = 0$, we get directly that $\mathbb{P}(X_1 = 1 | S_n = k) = 0$. We deduce that X_1 conditionally on $\{S_n = k\}$ is a Bernoulli random variable with parameter k/n for all $k \in [\![0, n]\!]$. We shall say that, conditionally on S_n, X_1 has the Bernoulli distribution with parameter S_n/n.

Using Corollary 2.12, we get that $\mathbb{E}[X_1 | S_n] = S_n/n$, which could have been obtained directly as the expectation of a Bernoulli random variable is given by its parameter. $\triangle$

2.3.2 The density case

Let Y be a random variable taking values in $(S, \mathcal{S})$ such that (Y, V) has a density with respect to some product measure on $S \times E$. See Fubini Theorem 1.53 for the definition of product measure. More precisely, we assume the probability distribution of (Y, V) is given by $f_{(Y,V)}(y, v)\, \mu(\mathrm{d}y)\nu(\mathrm{d}v)$, where μ and ν are σ-finite measures respectively on $(S, \mathcal{S})$ and $(E, \mathcal{E})$ and $f_{Y,V}$ is a $[0, +\infty]$-valued measurable function such that $\int f_{(Y,V)}\, \mathrm{d}\, \mu \otimes \nu = 1$. In this setting, we give a closed formula for $\mathbb{E}[X|V]$ when $X = \varphi(Y, V)$, with φ a real-valued measurable function defined on $S \times E$ endowed with the product σ-field.

According to Fubini theorem, V has probability distribution $f_V \nu$ with density (with respect to the measure ν) given by $f_V(v) = \int f_{(Y,V)}(y, v)\, \mu(\mathrm{d}y)$ and Y has probability distribution $f_Y \mu$ with density (with respect to the measure μ) given by $f_Y(y) = \int f_{(Y,V)}(y, v)\, \nu(\mathrm{d}v)$. We now define the law of Y conditionally on V.

Definition 2.15. *The probability distribution of Y conditionally on $\{V = v\}$, with $v \in E$ such that $f_V(v) \in (0 + \infty)$, is defined by $f_{Y|V}(y|v)\, \mu(\mathrm{d}y)$ with its density $f_{Y|V}$ (with respect to the measure μ) given by:*

$$f_{Y|V}(y|v) = \frac{f_{(Y,V)}(y, v)}{f_V(v)}, \quad y \in S.$$

By convention, we set $f_{Y|V}(y|v) = 0$ if $f_V(v) \notin (0, +\infty)$.

Thanks to Fubini theorem, we get that, for v such that $f_V(v) \in (0, +\infty)$, the function $y \mapsto f_{Y|V}(y|v)$ is a density as it is non-negative and $\int f_{Y|V}(y|v)\, \mu(\mathrm{d}y) = 1$.

We now give the expectation of $X = \varphi(Y, V)$, for some function φ, conditionally on V.

Proposition 2.16. *Let $(E, \mathcal{E}, \nu)$ and $(S, \mathcal{S}, \mu)$ be measured space such that ν and μ are σ-finite. Let (Y, V) be an $S \times E$-valued random variable with density $(y, v) \mapsto f_{(Y,V)}(y, v)$ with respect to the product measure $\mu(\mathrm{d}y)\nu(\mathrm{d}v)$. Let φ be a real-valued measurable function defined on $S \times E$ and set $X = \varphi(Y, V)$. Assume that $\mathbb{E}[X]$ is well defined. Then we have that a.s. $\mathbb{E}[X|V] = g(V)$, with:*

$$g(v) = \int \varphi(y, v) f_{Y|V}(y|v)\, \mu(\mathrm{d}y). \tag{2.7}$$

Proof. Let $A \in \sigma(V)$. The function $\mathbf{1}_A$ is $\sigma(V)$-measurable, and thus, thanks to Proposition 1.25, there exists a measurable function h such that $\mathbf{1}_A = h(V)$. Since f_V is a density, we get that $\int \mathbf{1}_{\{f_V \notin (0,+\infty)\}} \, f_V \, \mathrm{d}\nu = 0$. We have:

$$\begin{aligned}
\mathbb{E}[X\mathbf{1}_A] &= \mathbb{E}[\varphi(Y,V)h(V)] \\
&= \int \varphi(y,v)h(v)f_{(Y,V)}(y,v)\ \mu(\mathrm{d}y)\nu(\mathrm{d}v) \\
&= \int \varphi(y,v)h(v)f_{(Y,V)}(y,v)\mathbf{1}_{\{f_V(v)\in(0,+\infty)\}}\ \mu(\mathrm{d}y)\nu(\mathrm{d}v) \\
&= \int h(v)\left(\int \varphi(y,v)f_{Y|V}(y|v)\ \mu(\mathrm{d}y)\right) f_V(v)\mathbf{1}_{\{f_V(v)\in(0,+\infty)\}}\ \nu(\mathrm{d}v) \\
&= \int h(v)g(v)f_V(v)\ \nu(\mathrm{d}v) \\
&= \mathbb{E}[g(V)h(V)] = \mathbb{E}[g(V)\mathbf{1}_A],
\end{aligned}$$

where we used that $\mathrm{d}\,\mu\otimes\nu$-a.e.:

$$f_{(Y,V)}(y,v) = f_{(Y,V)}(y,v)\mathbf{1}_{\{f_V(v)\in(0,+\infty)\}}$$

as $\int \mathbf{1}_{\{f_V\notin(0,+\infty)\}} f_{(Y,V)}\,\mathrm{d}\,\mu\otimes\nu = \int \mathbf{1}_{\{f_V\notin(0,+\infty)\}} f_V\,\mathrm{d}\nu = 0$ for the third equality, the definition of $f_{Y|V}$ and Fubini theorem for the fourth and the definition of g given by (2.7) for the fith. Using (2.1) and Lemma 2.3, we deduce that a.s. $g(V) = \mathbb{E}[X|V]$. □

2.3.3 Elements on the conditional distribution

We shall present in this section some elementary notions on the conditional distribution.

Let $(E,\mathcal{E},\nu)$ be a measured space and $(S,\mathcal{S})$ a measurable space. A probability kernel κ is a $[0,1]$-valued function defined on $E\times\mathcal{S}$ such that: for all $v\in E$, the map $A\mapsto\kappa(v,A)$ is a measure on $(S,\mathcal{S})$; for all $A\in\mathcal{S}$, the map $v\mapsto\kappa(v,A)$ is measurable; and $\nu(\mathrm{d}v)$-a.e. $\kappa(v,S)=1$. It is left to the reader to prove that for any $[0,+\infty]$-valued measurable function φ defined on $S\times E$, the map $v\mapsto\int\varphi(y,v)\,\kappa(v,\mathrm{d}y)$ is measurable.

Definition 2.17. *Let (Y,V) be an $S\times E$-valued random variable, such that the distribution of V has a density*[1] *with respect to the measure ν. The probability kernel κ is the conditional distribution of Y given V if a.s.:*

$$\mathbb{P}(Y\in A|V) = \kappa(V,A) \quad \textit{for all } A\in\mathcal{S}.$$

If the probability kernel κ is the conditional distribution of Y given V, then arguing as in the proof of Fubini's Theorem 1.53, we get that for any $[0,+\infty]$-valued measurable function φ defined on $S\times E$ a.s.:

$$\mathbb{E}[\varphi(Y,V)|V] = g(V) \quad \text{with} \quad g(v) = \int \varphi(y,v)\,\kappa(v,\mathrm{d}y).$$

The existence of a probability kernel[2] allows to give a representation of the conditional expectation which holds simultaneously for all nice functions φ (but on a set of 0 probability

[1] If one takes $\nu = \mathrm{P}_V$, then the density is constant equal to 1.

[2] The existence of the conditional distribution of Y, taking values in S, given V can be proven under some topological property of the space $(S,\mathcal{S})$. See Theorem 5.3 in O. Kallenberg. *Foundations of modern probability.* Springer-Verlag, 2002.

for V). When V is a discrete random variable, Remark 2.13 states that the kernel κ given by $\kappa(v, \mathrm{d}x) = \mathbf{1}_{\{v \in \Delta_V\}} \mathrm{P}_{X|v}(\mathrm{d}x)$ is, with $\nu = \mathrm{P}_V$, the conditional distribution of X given V.

Example 2.18. In Example 2.14, with $P = S/n$, the conditional distribution of X_1 given P is the Bernoulli distribution with parameter P. This corresponds to the kernel $\kappa(p, \mathrm{d}x) = (1-p)\delta_0(\mathrm{d}x) + p\delta_1(\mathrm{d}x)$. (Notice one only needs to consider $p \in [0, 1]$.)

In Exercise 8.20, the conditional distribution of Y given V is the uniform distribution on $[0, V]$. This corresponds to the kernel $\kappa(v, \mathrm{d}y) = v^{-1}\mathbf{1}_{[0,v]}(y)\,\lambda(\mathrm{d}y)$. (Notice one only needs to consider $v \in (0, +\infty)$.) △

Chapter 3

Discrete Markov chains

A Markov chain is a sequence of random variables $X = (X_n, n \in \mathbb{N})$ which represents the dynamical evolution (in discrete time) of a stochastic system: X_n represents the state of the system at time n. The fundamental property of a Markov chain is that the evolution after time n of the system depends on the previous states only through the state of the system at time n. In other words, conditionally on X_n, $(X_0, \ldots, X_n)$ and $(X_{n+k}, k \in \mathbb{N})$ are independent. Markov chains appears naturally in a large variety of domain: networks, population genetics, mathematical finance, stock management, stochastic optimization algorithms, simulations, ... We shall be interested in the asymptotic behavior of X for large times. In what follows, we assume that the state space is at most countable.

We give in Section 3.1 the definition and the first properties of the Markov chains. Then, we consider invariant measures in Section 3.2. We characterize the states of the Markov chain and introduce the notion of irreducible chain in Section 3.3. Intuitively, an irreducible chain has a positive probability starting from one state to go in one or many steps to any other state. The ergodic theorems from Section 3.4 give the asymptotic behavior of an irreducible Markov chain for large time. They are among the most interesting results on Markov chains. Their proof is postponed to Section 3.4.3. In Section 3.5, we present and analyze some well known applications of Markov chains. We refer to [7, 3] for a recent and very detailed presentation of Markov chains.

We shall consider that all the random variables of this chapter are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In all this chapter we shall consider a discrete state space[1] E (not reduced to one state) with the σ-field $\mathcal{E} = \mathcal{P}(E)$.

3.1 Definition and properties

Let $X = (X_n, n \in \mathbb{N})$ be a sequence of E-valued random variables, which will be seen as a process, X_n being the state of the process at time n. We represent the information available at time $n \in \mathbb{N}$ by a σ-field $\mathcal{F}_n$, which is non-decreasing with n.

[1] The set E is discrete if E is at most countable, all $x \in E$ are isolated, that is all subsets of E are open and closed. For example, the set $\mathbb{N}$ with the Euclidean distance is a discrete set, while the set $\{0\} \cup \{1/k,\, k \in \mathbb{N}^*\}$ with the Euclidean distance is not a discrete set as the set $\{0\}$ is not open.

Definition 3.1. *A filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$ (with respect to the measurable space $(\Omega, \mathcal{F})$) is a sequence of σ-fields such that $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ for all $n \in \mathbb{N}$. A E-valued process $X = (X_n, n \in \mathbb{N})$ is $\mathbb{F}$-adapted if X_n is $\mathcal{F}_n$-measurable for all $n \in \mathbb{N}$.*

In the setting of stochastic process, one usually (but not always) chooses the natural filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$ which is generated by X: for all $n \in \mathbb{N}$, $\mathcal{F}_n$ is the σ-field generated by $(X_0, \ldots, X_n)$ and the $\mathbb{P}$-null sets. This obviously implies that X is $\mathbb{F}$-adapted.

A Markov chain is a process such that, conditionally on the process at time n, the past before time n and the evolution of the process after time n are independent.

Definition 3.2. *The process $X = (X_n, n \in \mathbb{N})$ is a Markov chain with respect to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$ if it is adapted and it has the Markov property: for all $n \in \mathbb{N}$, conditionally on X_n, $\mathcal{F}_n$ and $(X_k,\, k \geq n)$ are independent, that is for all $A \in \mathcal{F}_n$ and $B \in \sigma(X_k,\, k \geq n)$, $\mathbb{P}(A \cap B|\, X_n) = \mathbb{P}(A|\, X_n)\mathbb{P}(B|\, X_n)$.*

In the previous definition, we shall omit to mention the filtration when it is the natural filtration. Since X is adapted to $\mathbb{F}$, if X is a Markov chain with respect to $\mathbb{F}$, it is also a Markov chain with respect to its natural filtration.

We give equivalent definitions for being a Markov chain.

Proposition 3.3. *Let $X = (X_n, n \in \mathbb{N})$ be a E-valued process adapted to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$. The following properties are equivalent.*

(i) X is Markov chain.

(ii) For all $n \in \mathbb{N}$ and $B \in \sigma(X_k,\, k \geq n)$, we have a.s. $\mathbb{P}(B|\, \mathcal{F}_n) = \mathbb{P}(B|\, X_n)$.

(iii) For all $n \in \mathbb{N}$ and $y \in E$, we have a.s. $\mathbb{P}(X_{n+1} = y|\, \mathcal{F}_n) = \mathbb{P}(X_{n+1} = y|\, X_n)$.

Proof. That property (i) implies property (ii) is a direct consequence of Exercise 8.18 (with $\mathcal{A} = \mathcal{F}_n$, $\mathcal{B} = \sigma(X_k,\, k \geq n)$ and $\mathcal{H} = \sigma(X_n)$). Let us check that property (ii) implies property (i). Assume property (ii). Let $A \in \mathcal{F}_n$ and $B \in \sigma(X_k,\, k \geq n)$. A.s. we have, using property (ii) for the second equality:

$$\mathbb{P}(A \cap B|\, X_n) = \mathbb{E}\left[\mathbf{1}_A \mathbb{E}\left[\mathbf{1}_B|\, \mathcal{F}_n\right] |\, X_n\right] = \mathbb{E}\left[\mathbf{1}_A \mathbb{E}\left[\mathbf{1}_B|\, X_n\right] |\, X_n\right] = \mathbb{P}(A|\, X_n)\mathbb{P}(B|\, X_n).$$

This gives property (i).

Taking $B = \{X_{n+1} = y\}$ in property (ii) gives property (iii). We now assume property (iii) holds, and we prove property (ii). As $\sigma(X_k,\, k \geq n)$ is generated by the events $B_k = \{X_n = y_0, \ldots, X_{n+k} = y_k\}$ where $k \in \mathbb{N}$ and $y_0, \ldots, y_k \in E$, we deduce from the monotone class theorem, and more precisely Corollary 1.14, that, to prove (ii), it is enough to prove that a.s.

$$\mathbb{P}(B_k|\, \mathcal{F}_n) = \mathbb{P}(B_k|\, X_n). \tag{3.1}$$

We shall prove this by induction. Notice (3.1) is true for $k = 1$ thanks to (iii). Assume that

(3.1) is true for $k \geq 1$. Then, we have a.s.:

$$\begin{aligned}
\mathbb{P}(B_{k+1}|\,\mathcal{F}_n) = \mathbb{E}\left[\mathbf{1}_{B_{k+1}}|\,\mathcal{F}_n\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{X_{n+k+1}=y_{k+1}\}}\mathbf{1}_{B_k}|\,\mathcal{F}_{n+k}\right]|\,\mathcal{F}_n\right] \\
&= \mathbb{E}\left[\mathbb{P}(X_{n+k+1}=y_{k+1}|\,\mathcal{F}_{n+k})\mathbf{1}_{B_k}|\,\mathcal{F}_n\right] \\
&= \mathbb{E}\left[\mathbb{P}(X_{n+k+1}=y_{k+1}|\,X_{n+k})\mathbf{1}_{B_k}|\,\mathcal{F}_n\right] \\
&= \mathbb{P}(X_{n+k+1}=y_{k+1}|\,X_{n+k}=y_k)\,\mathbb{P}(B_k|\,\mathcal{F}_n) \\
&= \mathbb{P}(X_{n+k+1}=y_{k+1}|\,X_{n+k}=y_k)\,\mathbb{P}(B_k|\,X_n),
\end{aligned}$$

where we used that $B_k \in \mathcal{F}_{n+k}$ for the second equality, property (iii) for the third, and that $X_{n+k} = y_{n+k}$ on B_k and, see Corollary 2.12, that a.s. $\mathbb{P}(X_{n+k+1} = y_{k+1}|\,X_{n+k})\mathbf{1}_{\{X_{n+k}=y_k\}} = \mathbb{P}(X_{n+k+1} = y_{k+1}|\,X_{n+k} = y_k)\mathbf{1}_{\{X_{n+k}=y_k\}}$ for the fourth and the induction for the last. In particular, we deduce that $\mathbb{P}(B_{k+1}|\,\mathcal{F}_n)$ is $\sigma(X_n)$-measurable. This readily implies that $\mathbb{P}(B_{k+1}|\,\mathcal{F}_n) = \mathbb{P}(B_{k+1}|\,X_n)$ and thus (3.1) is true for k replaced by $k+1$. This ends the proof of property (ii). □

As an immediate consequence of this proposition, using property (iv) of Proposition 2.9 and that $\sigma(X_0, \ldots, X_n) \subset \mathcal{F}_n$, we deduce that for a Markov chain $X = (X_n, n \in \mathbb{N})$:

$$\text{a.s.}\quad \mathbb{P}(X_{n+1} = y \mid \mathcal{F}_n) = \mathbb{P}(X_{n+1} = y \mid X_0, \ldots, X_n) = \mathbb{P}(X_{n+1} = y \mid X_n). \tag{3.2}$$

Unless specified otherwise, we shall consider $\mathbb{F}$ is the natural filtration of X.

Example 3.4. We present the example of the simple random walk, which has been (and is still) thoroughly studied, see [8, 6]. We take $E = \mathbb{Z}$. Let $p \in (0,1)$ and $U = (U_n, n \in \mathbb{N}^*)$ be independent random variables taking values in $\{-1, 1\}$ with the same distribution $\mathbb{P}(U_n = 1) = 1 - \mathbb{P}(U_n = -1) = p$. Let X_0 be a $\mathbb{Z}$-valued random variable independent of U and set for $n \in \mathbb{N}^*$:

$$X_n = X_0 + \sum_{k=1}^{n} U_k.$$

By construction we get property (iii) from Proposition 3.3 holds as $\mathbb{P}(X_{n+1} = y|\,\mathcal{F}_n) = \mathbb{P}(U_{n+1} = y - X_n|\,\mathcal{F}_n) = \varphi(y - X_n)$ with $\varphi(z) = \mathbb{P}(U_{n+1} = z)$, since U_{n+1} is independent of $\mathcal{F}_n$ and thanks to (8.1) with $Y = U_{n+1}$, $V = X_n$ and $\mathcal{H} = \mathcal{F}_n$. Thus the process X is a Markov chain. △

Motivated by this example, we have the following lemma whose proof is similar and left to the reader.

Lemma 3.5. *Let $(S, \mathcal{S})$ be a measurable space. Let $U = (U_n, n \in \mathbb{N}^*)$ be a sequence of independent S-valued random variables. Let X_0 be a E-valued random variable independent of U. Let f be a measurable function defined on $E \times S$ taking values in E. The stochastic dynamical system $X = (X_n, n \in \mathbb{N})$ defined by $X_{n+1} = f(X_n, U_{n+1})$ for $n \in \mathbb{N}$ is a Markov chain.*

The sequence U in Lemma 3.5 is called the sequence of innovations. In what follows, we link the Markov chains with the matrix formalism.

Definition 3.6. *A matrix $P = (P(x,y), x, y \in E)$ on E is stochastic if: $P(x,y) \geq 0$ for all $x, y \in E$, and $\sum_{y \in E} P(x,y) = 1$ for all $x \in E$.*

In view of (3.2), it is natural to focus on the transition probability $\mathbb{P}(X_{n+1} = y \mid X_n)$.

Definition 3.7. *A Markov chain X on E has transition matrices $(P_n, n \in \mathbb{N}^*)$ if $(P_n, n \in \mathbb{N}^*)$ is a sequence of stochastic matrices on E and for all $y \in E$ a.s.:*

$$\mathbb{P}(X_{n+1} = y \mid X_n) = P_{n+1}(X_n, y). \tag{3.3}$$

The Markov chain is called homogeneous when the sequence $(P_n, n \in \mathbb{N}^*)$ is constant, and its common value, say P, is then called the[2] *transition matrix of X.*

The transition matrix of the simple random walk described in Example 3.4 is given by $P(x, y) = 0$ if $|x - y| \neq 1$, $P(x, x + 1) = p$ and $P(x, x - 1) = 1 - p$ for $x, y \in \mathbb{Z}$.

Unless specified otherwise, we shall consider homogeneous Markov chains.

The next proposition states that the transition matrix and the initial distribution characterize the distribution of the Markov chain.

Proposition 3.8. *The distribution of a (homogeneous) Markov chain $X = (X_n, n \in \mathbb{N})$ is characterized by its transition matrix, P, and the initial probability distribution, μ_0, of X_0. Moreover, we have for all $n \in \mathbb{N}^*$, $x_0, \ldots, x_n \in E$:*

$$\mathbb{P}(X_0 = x_0, \ldots, X_n = x_n) = \mu_0(x_0) \prod_{k=1}^{n} P(x_{k-1}, x_k). \tag{3.4}$$

In order to stress the dependence of the distribution of the Markov chain X on the probability distribution μ_0 of X_0, we may write $\mathbb{P}_{\mu_0}$ and $\mathbb{E}_{\mu_0}$. When μ_0 is simply the Dirac mass at x (that is $\mathbb{P}(X_0 = x) = 1$), then we simply write $\mathbb{P}_x$ and $\mathbb{E}_x$ and say the Markov chain is started at x.

Proof. We have that for $k \in \mathbb{N}^*$, $x_0, \ldots, x_{k+1} \in E$, with $B_k = \{X_0 = x_0, \ldots, X_k = x_k\}$, that:

$$\begin{aligned}
\mathbb{P}(X_{k+1} = x_{k+1}, B_k) &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{X_{k+1}=x_{k+1}\}} \mathbf{1}_{B_k} \mid \mathcal{F}_k\right]\right] \\
&= \mathbb{E}\left[\mathbb{P}(X_{k+1} = x_{k+1} \mid \mathcal{F}_k)\, \mathbf{1}_{B_k}\right] \\
&= \mathbb{E}\left[P(X_k, x_{k+1}) \mathbf{1}_{B_k}\right] \\
&= P(x_k, x_{k+1}) \mathbb{P}(B_k),
\end{aligned}$$

where we used that $B_k \in \mathcal{F}_k$ for the second equality, (3.2) for the third, that $X_k = x_k$ on B_k for the last. We then deduce that (3.4) holds by induction.

Use that $\{(x_0, \ldots, x_n)\}$ for $x_0, \ldots, x_n \in E$ generates the product σ-field on E^{n+1} and Lemma 1.29 to deduce that the left hand side of (3.4) for all $n \in \mathbb{N}$ and $x_0, \ldots, x_n \in E$ characterizes the distribution of X. We then deduce from (3.4) that the distribution of X is characterized by P and μ_0. □

We now give some examples of Markov chains.

Example 3.9. If the process $X = (X_n, n \in \mathbb{N})$ is a sequence of independent random variables with distribution π, then X is a Markov chain with transition matrix P given by $P(x, y) = \pi(y)$ for all $x, y \in E$. △

[2] There is a slight abuse here, as (3.3) might not characterize P. Indeed, if $\mathbb{P}(X_n = x) = 0$ for some $x \in E$ and all $n \in \mathbb{N}^*$, then $P(x, \cdot)$ is not characterized by (3.3). This shall however not be troublesome.

Example 3.10. Let X_n be the number of items in a stock at time n, D_{n+1} the random consumer demand and $q \in \mathbb{N}^*$ the deterministic quantity of items produced between period n and $n+1$. Considering the stock at time $n+1$, we get:

$$X_{n+1} = (X_n + q - D_{n+1})^+.$$

If the demand $D = (D_n, n \in \mathbb{N}^*)$ is a sequence of independent random variables with the same distribution, independent of X_0, then according to Lemma 3.5, the stochastic dynamical system $X = (X_n, n \in \mathbb{N})$ is a Markov chain. Its transition matrix is given by: $P(x, y) = \mathbb{P}(D = k)$ if $y = x+q-k > 0$, and $P(x, 0) = \mathbb{P}(D \geq x+q)$ for $x, y \in \mathbb{N}$. Figure 3.1 represents some simulations of the process X for different probability distributions of the demand. $\triangle$

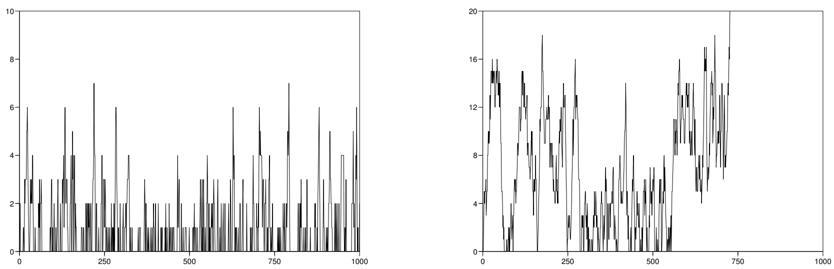

Figure 3.1: Simulations of the the random evolution of a stock with dynamics $X_{n+1} = (X_n + q - D_{n+1})^+$, where $X_0 = 0$, $q = 3$ and the random variables $(D_n, n \in \mathbb{N}^*)$ are independent with Poisson distribution parameter θ ($\theta = 4$ on the left and $\theta = 3$ on the right).

Remark 3.11. Even if a Markov chain is not a stochastic dynamical system, it is distributed as one. Indeed let μ_0 be a probability distribution on E and P a stochastic matrix on E. Let X_0 be a E-valued random variable with distribution μ_0 and $(U_n, n \in \mathbb{N})$ be a sequence of independent random variables, independent of X_0 distributed as $U = (U(x),\, x \in E)$, where $U(x)$ are independent E-valued random variables such that $U(x)$ has distribution $P(x, \cdot)$. Then the stochastic dynamical system $(X_n,\, n \in \mathbb{N})$, defined by $X_{n+1} = U_{n+1}(X_n)$ for $n \in \mathbb{N}$, is a Markov chain on E with initial distribution μ_0 and transition matrix P. $\diamondsuit$

The next corollary is a consequence of the Markov property.

Corollary 3.12. *Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain with respect to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$, taking values in a discrete state space E and with transition matrix P. Let $n \in \mathbb{N}$ and defined the shifted process $\tilde{X} = (\tilde{X}_k = X_{n+k}, k \in \mathbb{N})$. Conditionally on X_n, we have that $\mathcal{F}_n$ and $\tilde{X}$ are independent and that $\tilde{X}$ is a Markov chain with transition matrix P started at X_n, which means that a.s. for all $k \in \mathbb{N}$, all $x_0, \ldots, x_k \in E$:*

$$\begin{aligned}\mathbb{P}(\tilde{X}_0 = x_0, \ldots, \tilde{X}_k = x_k \mid \mathcal{F}_n) &= \mathbb{P}(\tilde{X}_0 = x_0, \ldots, \tilde{X}_k = x_k \mid X_n) \\ &= \mathbf{1}_{\{X_n = x_0\}} \prod_{j=1}^{k} P(x_{j-1}, x_j). \end{aligned} \tag{3.5}$$

Notice that in the previous corollary the initial distribution of the Markov chains X and $\tilde{X}$ are not the same *a priori*.

Proof. By definition of a Markov chain, we have that, conditionally on X_n, $\mathcal{F}_n$ and $\tilde{X}$ are independent. So, we only need to prove that:

$$\mathbb{P}(\tilde{X}_0 = x_0, \ldots, \tilde{X}_k = x_k \mid \mathcal{F}_n) = \mathbf{1}_{\{X_n = x_0\}} \prod_{j=1}^{k} P(x_{j-1}, x_j). \tag{3.6}$$

Set $B_j = \{\tilde{X}_0 = x_0, \ldots, \tilde{X}_j = x_j\} = \{X_n = x_0, \ldots, X_{n+j} = x_j\}$ for $j \in \{0, \ldots, k\}$. Using (3.2) and Definition 3.7 with n replaced by $n + j$, we get for $j \in \{0, \ldots, k-1\}$ that:

$$\mathbb{E}\left[\mathbf{1}_{B_{j+1}} \mid \mathcal{F}_{n+j}\right] = \mathbb{E}\left[\mathbf{1}_{\{X_{n+j+1} = x_{j+1}\}} \mathbf{1}_{B_j} \mid \mathcal{F}_{n+j}\right] = P(X_{n+j}, x_{j+1}) \mathbf{1}_{B_j} = P(x_j, x_{j+1}) \mathbf{1}_{B_j},$$

where we used that $X_{n+j} = x_j$ on B_j for the last equality. This implies that $\mathbb{P}(B_{j+1} | \mathcal{F}_n) = P(x_j, x_{j+1}) \mathbb{P}(B_j \mid \mathcal{F}_n)$. Thus, we deduce that (3.6) holds by induction. Then, conclude using Proposition 3.8 on the characterization of the distribution of a Markov chain. □

In the setting of Markov chains, computing probability distribution or expectation reduce to elementary linear algebra on E. Let P and Q be two matrices defined on E with non-negative entries. We denote by PQ the matrix on E defined by $PQ(x, y) = \sum_{z \in E} P(x, z)Q(z, y)$ for $x, y \in E$. It is easy to check that if P and Q are stochastic, then PQ is also stochastic. We set $P^0 = I_E$ the identity matrix on E and for $k \geq 1$, $P^k = P^{k-1}P$ (or equivalently $P = PP^{k-1}$).

For a line vector $\mu = (\mu(x), x \in E)$ with non-negatives entries, which we shall see as a measure on E, we denote by μP the line vector $(\mu P(y), y \in E)$ defined by $\mu P(y) = \sum_{x \in E} \mu(x)P(x, y)$. For a column vector $f = (f(y), y \in E)$ with real entries, which we shall see as a function defined on E, we denote, by Pf or $P(f)$ the column vector $(Pf(x), x \in E)$ defined by $Pf(x) = \sum_{y \in E} P(x, y)f(y)$. Notice this last quantity, and thus Pf, is well defined as soon as, for all $x \in E$ we have $P(f^+)(x)$ or $P(f^-)(x)$ finite. We also write $\mu f = (\mu, f) = \sum_{x \in E} \mu(x)f(x)$ the integral of the function f with respect to the measure μ, when it is well defined.

We shall consider a measure $\mu = (\mu(x) = \mu(\{x\}), x \in E)$ on E as a line vector with non negative entries. For $A \subset E$, we set $\mu(A) = \sum_{x \in A} \mu(x)$, so that μ is a probability measure if $\sum_{x \in E} \mu(x) = 1$. We shall also consider a real-valued function $f = (f(x), x \in E)$ defined on E as a column vector. The next results give explicit formula to compute (conditional) expectations and distributions.

Proposition 3.13. *Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain with transition matrix P. Denote by μ_n the probability distribution of X_n for $n \in \mathbb{N}$. Let f be a bounded or non-negative function. We have for $n \in \mathbb{N}^*$:*

(i) $\mu_n = \mu_0 P^n$,

(ii) $\mathbb{E}[f(X_n)] = \mu_n f = \mu_0 P^n f$ *and* $\mathbb{E}_x[f(X_n)] = P^n f(x)$ *for* $x \in E$,

(iii) $\mathbb{E}[f(X_n) | \mathcal{F}_{n-1}] = Pf(X_{n-1})$ *a.s.*,

(iv) $\mathbb{E}[f(X_n) | X_0] = P^n f(X_0)$ *a.s.*,

(v) $\mathbb{P}(X_{n+k} = y | \mathcal{F}_n) = P^k(X_n, y)$ *a.s. for all* $k \in \mathbb{N}$, $y \in E$.

Proof. Summing (3.4) over $x_0, \ldots, x_{n-1} \in E$ gives property (i). Property (ii) is a direct consequence of property (i). Using that $\mathbb{P}(X_n = y \mid \mathcal{F}_{n-1}) = P(X_{n-1}, y)$, see (3.2) and (3.3), multiplying by $f(y)$ and summing over $y \in E$ gives property (iii). Iterating (iii) leads to $\mathbb{E}[f(X_n)|\,\mathcal{F}_0] = P^n f(X_0)$, which implies (iv) as a.s. $\mathbb{E}[f(X_n)|\,X_0] = \mathbb{E}[\mathbb{E}[f(X_n)|\,\mathcal{F}_0] \mid X_0]$. Iterating (iii) leads also to $\mathbb{E}[f(X_{n+k})|\,\mathcal{F}_n] = P^k f(X_n)$ a.s., and then take $f = \mathbf{1}_{\{y\}}$. $\square$

Example 3.14. Let $(U_n, n \in \mathbb{N}^*)$ be a sequence of independent Bernoulli random variables with parameter $p \in (0,1)$. Let $p_{\ell,n}$ be the probability to get a sequence of consecutive 1 with length at least ℓ in the sequence $U_1 \ldots U_n$. It is very simple to get a closed formula for $p_{\ell,n}$ using the formalism of Markov chains[3,4]. We consider the Markov chain $X = (X_n, n \in \mathbb{N})$ defined by $X_0 = 0$ and $X_{n+1} = (X_n + 1)\mathbf{1}_{\{U_{n+1}=1, X_n<\ell\}} + \ell\mathbf{1}_{\{X_n=\ell\}}$ for $n \in \mathbb{N}$. As soon as we observe a sequence of consecutive 1 with length ℓ, then the process X is constant equal to ℓ. In particular, we have $p_{\ell,n} = \mathbb{P}(X_n = \ell) = P^n(0, \ell)$, where P is the transition matrix of the Markov chain X. The transition matrix is given by $P(x, 0) = 1 - p$ and $P(x, x+1) = p$ for $x \in \{0, \ldots, \ell - 1\}$, $P(\ell, \ell) = 1$ and all the other entries of P are zeros. We give the values of $p_{\ell,n}$ for $n = 100$ and $p = 1/2$ in Figure 3.2. In particular, for $p = 1/2$, we get a probability larger than 1/2 to observe a sequence of 6 consecutive 1 in a sequence of length 100. $\triangle$

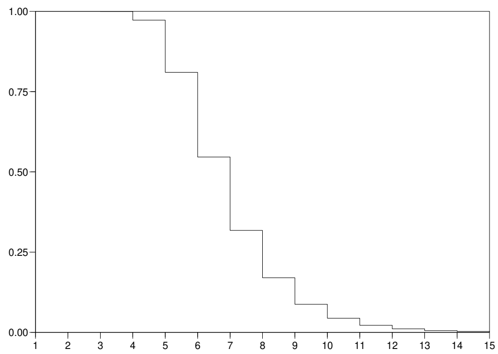

Figure 3.2: Graph of the function $x \mapsto \mathbb{P}(L_n \geq \lfloor x \rfloor)$, with L_n the maximal length of the sequences of consecutive 1 in a sequence of length $n = 100$ of independent Bernoulli random variables with parameter $p = 1/2$.

3.2 Invariant probability measures, reversibility

Invariant probability measures appear naturally in the asymptotic study of Markov chains at large times.

[3] L. Guibas and A. Odlyzko. String overlaps, pattern matching, and nontransitive games. *J. Combin. Theory Ser. A*, vol. 30(2), pp. 183-208, 1981.

[4] J. Fu and V. Koutras. Distribution theory of runs: a Markov chain approach. *J. Amer. Statist. Assoc.*, vol. 89(427), pp. 1050-1058, 1994.

Definition 3.15. *A probability measure π is invariant for a stochastic matrix P if $\pi = \pi P$. It is also called a stationary probability measure.*

Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain with transition matrix P with starting probability measure $\mu_0 = \pi$ an invariant probability measure for P. Denote by μ_n the probability distribution of X_n. We have $\mu_1 = \pi P = \pi$ and by recurrence we get $\mu_n = \pi$ for all $n \in \mathbb{N}^*$. This means that X_n is distributed as X_0: the distribution of X_n is stationary, that is constant in time.

Remark 3.16. Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain with transition matrix P with starting probability measure $\mu_0 = \pi$ an invariant probability measure for P. For simplicity, let us assume further that $\pi(x) > 0$ for all $x \in E$. For $x, y \in E$, we set:

$$Q(x, y) = \frac{\pi(y)P(y, x)}{\pi(x)}. \tag{3.7}$$

Since π is an invariant probability measure, we have $\sum_{y\in E} Q(x, y) = 1$ for all $x \in E$. Thus the matrix Q is stochastic. Notice that π is also an invariant probability measure for Q. For $x, y \in E$, $n \in \mathbb{N}$, we have:

$$\mathbb{P}_\pi(X_n = y|X_{n+1} = x) = \frac{\mathbb{P}_\pi(X_n = y, X_{n+1} = x)}{\mathbb{P}_\pi(X_{n+1} = x)} = Q(x, y).$$

More generally, it is easy to check that for all $n \in \mathbb{N}$, $x_0, \ldots, x_n \in E$:

$$\mathbb{P}_\pi(X_n = x_0, \ldots, X_0 = x_n) = \pi(x_0) \prod_{k=1}^{n} Q(x_{k-1}, x_k).$$

In other words $(X_n, X_{n-1}, \ldots, X_0)$ is distributed under $\mathbb{P}_\pi$ as the first n steps of a Markov chains with transition matrix Q with initial distribution π. Intuitively, the time reversal of the process X under π is a Markov chain with transition matrix Q. $\Diamond$

There is an important particular case where a probability measure π is invariant for a stochastic matrix.

Definition 3.17. *A stochastic matrix P is reversible with respect to a probability measure π if for all $x, y \in E$:*

$$\pi(x)P(x, y) = \pi(y)P(y, x). \tag{3.8}$$

A Markov chain X is reversible with respect to a probability measure π if its transition matrix is reversible with respect to π.

Summing (3.8) over $x \in E$, we deduce the following lemma.

Lemma 3.18. *If a stochastic matrix P is reversible with respect to a probability measure π, then π is an invariant probability measure for P.*

See examples of the Ehrenfest urn model and the Metropolis-Hastings algorithm in Section 3.5 for reversible Markov chains.

Remark 3.19. If P in Remark 3.16 is also reversible with respect to the probability measure π, then we get $P = Q$. Therefore, under $\mathbb{P}_\pi$, we get that $(X_0, \dots, X_{n-1}, X_n)$ and $(X_n, X_{n-1}, \dots, X_0)$ have the same distribution. We give a stronger statement in the next Remark. $\lozenge$

Remark 3.20. Let P be a stochastic matrix on E reversible with respect to a probability measure π. The following construction is inspired by Remark 3.11. Let $(U_n, n \in \mathbb{Z}^*)$ be a sequence of independent random variables distributed as $U = (U(x), x \in E)$, where the E-valued random variables are independent and $U(x)$ is distributed as $P(x, \cdot)$. Let X_0 be a E-valued random variable independent of $(U_n, n \in \mathbb{Z}^*)$ with distribution π. For $n \in \mathbb{N}^*$, set $X_{n+1} = U_{n+1}(X_n)$ and $X_{-(n+1)} = U_{-(n+1)}(X_{-n})$. Then the process $X = (X_n, n \in \mathbb{Z})$ can be seen as a Markov chain with time index $\mathbb{Z}$ instead of $\mathbb{N}$ in Definition 3.2 (the proof of this fact is left to the reader). We deduce from Remark 3.16 that $\tilde{X} = (\tilde{X}_n = X_{-n}, n \in \mathbb{Z})$ is then also a Markov chain with time index $\mathbb{Z}$. It is called the time reversal process of X. One can easily check that its transition matrix is P, so that X and $\tilde{X}$ have the same distribution. $\lozenge$

3.3 Irreducibility, recurrence, transience, periodicity

Let P be a stochastic matrix on E and $X = (X_n, n \in \mathbb{N})$ be a Markov chain with transition matrix P. Recall E is a finite or countable discrete space with at least two elements.

3.3.1 Communicating classes

In order to study the longtime behavior of the Markov chain X, we shall decompose the state space E in subsets on which the study of X will be easier.

We introduce some definitions. A state y is accessible from a state x, which we shall write $x \to y$, if $P^n(x, y) > 0$ for some $n \in \mathbb{N}$, or equivalently $\mathbb{P}_x(X_n = y \text{ for some } n \in \mathbb{N}) > 0$. Since $P^0 = I_E$, the identity matrix on E, we get that $x \to x$. The states x and y communicate, which we shall write as $x \leftrightarrow y$ if they are accessible from each other (that is $x \to y$ and $y \to x$). It is clear that "to communicate with" is an equivalence relation, and we denote by C_x the equivalent class of x. The communicating classes form a partition of the state space E. Notice the communicating classes are completely determined by the zero of the transition matrix P. We say the Markov chain X is irreducible if all states communicate with each other, that is E is a (and the only one) communicating class.

A communicating class C is called closed if for all $x \in C$ we have that $x \to y$ implies $y \in C$ (that is $x \leftrightarrow y$), and open otherwise. Intuitively, when a Markov chain reach a closed communicating class, it stays therein. A state $x \in E$ is called an absorbing state if $C_x = \{x\}$ and C_x is closed. Equivalently, the state x is an absorbing state if and only if $P(x, x) = 1$ and thus $\mathbb{P}_x(X_n = x \text{ for all } n \in \mathbb{N}) = 1$. In particular a Markov chain with an absorbing state is not irreducible.

Example 3.21. In Example 3.4, the simple random walk is an irreducible Markov chain with state space $\mathbb{Z}$ (that is $\mathbb{Z}$ is a closed communicating class).

In Example 3.14, the state ℓ is an absorbing state, and $\{0, \cdots, \ell - 1\}$ is an open communicating class.

The Markov chain in the Ehrenfest's urn model, see Section 3.5, is irreducible. The Markov chain of the Wright-Fischer model, see Section 3.5, has two absorbing states 0 and N and one open communicating class $\{1, \ldots, N-1\}$. △

3.3.2 Recurrence and transience

We use the convention $\inf \emptyset = +\infty$. We define the (first) return time of $x \in E$ for the Markov chain X by:

$$T^x = \inf\{n \geq 1;\, X_n = x\}.$$

Definition 3.22. *Let X be a Markov chain on E. The state $x \in E$ is transient if $\mathbb{P}_x(T^x = \infty) > 0$, and recurrent (or persistent) otherwise. The Markov chain is transient (resp. recurrent) if all the states are transient (resp. recurrent).*

We set $N^x = \sum_{n\in\mathbb{N}} \mathbf{1}_{\{X_n=x\}}$ the number of visits of the state x. The next proposition gives a characterization for transience and recurrence.

Proposition 3.23. *Let X be a Markov chain on E with transition matrix P.*

(i) Let $x \in E$ be recurrent. Then we have $\mathbb{P}_x(N^x = \infty) = 1$ and $\sum_{n\in\mathbb{N}} P^n(x,x) = +\infty$.

(ii) Let $x \in E$ be transient. Then we have $\mathbb{P}_x(N^x < \infty) = 1$, $\sum_{n\in\mathbb{N}} P^n(x,x) < +\infty$ and N^x has under $\mathbb{P}_x$ a geometric distribution with parameter $\mathbb{P}_x(T^x = \infty)$. And for all probability measure ν on E, we have $\mathbb{P}_\nu(N^x < \infty) = 1$. Furthermore, if π is an invariant measure for P, then $\pi(x) = 0$.

(iii) The elements of the same communicating class are either all transient or all recurrent.

(iv) The elements of an open communicating class are transient.

To have a complete picture, in view of property (iv) above, we shall study closed communicating classes (see Remark 3.25 below for a first result in this direction). For this reason, we shall consider Markov chains started in a closed communicating class. This amounts to study irreducible Markov chains, as a Markov chain started in a closed communicating class remains in it.

Proof. We set $p = \mathbb{P}_x(T^x = \infty) = \mathbb{P}_x(N^x = 1)$. Notice that $\{T^x < \infty\} = \{N^x > 1\}$ under $\mathbb{P}_x$. By decomposing according to the possible values of T^x, we get for $n \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{P}_x(N^x > n+1) &= \sum_{r\in\mathbb{N}^*} \mathbb{P}_x(N^x > n+1, T^x = r) \\
&= \sum_{r\in\mathbb{N}^*} \mathbb{P}_x\left(T^x = r, X_r = x, \sum_{\ell\in\mathbb{N}} \mathbf{1}_{\{X_{r+\ell}=x\}} > n\right) \\
&= \sum_{r\in\mathbb{N}^*} \mathbb{P}_x\left(T^x = r, X_r = x\right) \mathbb{P}_x\left(\sum_{\ell\in\mathbb{N}} \mathbf{1}_{\{X_\ell=x\}} > n\right) \\
&= \mathbb{P}_x(T_x < \infty)\mathbb{P}_x(N^x > n) \qquad (3.9) \\
&= (1-p)\mathbb{P}_x(N^x > n),
\end{aligned}$$

where we used the Markov property at time r for the third equality. Using that $\mathbb{P}_x(N^x > 0) = 1$, we deduce that $\mathbb{P}_x(N^x > n) = (1-p)^n$ for $n \in \mathbb{N}$. This gives that N^x has under $\mathbb{P}_x$ a geometric distribution with parameter $p \in [0,1]$. Notice also that $\mathbb{E}_x[N^x] = \sum_{n\in\mathbb{N}} \mathbb{P}_x(X_n = x) = \sum_{n\in\mathbb{N}} P^n(x,x)$, which is finite if and only if $p > 0$. Thus, if x is transient, then $p > 0$ and we get $\mathbb{P}_x(N_x < \infty) = 1$ and $\mathbb{E}_x[N^x]$ is finite. And, if x is recurrent, then $p = 0$ and we get $\mathbb{P}_x(N_x < \infty) = 0$ and $\mathbb{E}_x[N^x]$ is infinite. This proves property (i) and the first part of property (ii).

We prove the second part of property (ii). Let ν be a probability measure on E. As x is transient, by decomposing according to the values of T^x and using the Markov chain property for the first equality, we get:

$$\mathbb{P}_\nu(N^x = +\infty) = \sum_{n\in\mathbb{N}^*} \mathbb{P}_\nu(T^x = n)\mathbb{P}_x(N^x = +\infty) = 0,$$

that is $\mathbb{P}_\nu(N^x < \infty) = 1$. Let π be an invariant measure. Use $\mathbb{P}_\pi(N^x < \infty) = 1$ to get that:

$$\mathbb{P}_\pi\text{-a.s.} \quad \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{X_k = x\}} = \lim_{n\to\infty} \frac{1}{n} N^x = 0.$$

Since $\frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k=x\}}$ is bounded by 1, we deduce that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \mathbb{P}_\pi(X_k = x) = 0$ by dominated convergence. As π is invariant, we get that $\mathbb{P}_\pi(X_k = x) = \pi P^k(x) = \pi(x)$. We deduce that $\pi(x) = 0$. This finishes the proof of property (ii).

We prove property (iii). Let x, y be two elements of the same communicating class. In particular, there exists $n_1, n_2 \in \mathbb{N}$ such that $P^{n_1}(y,x) > 0$ and $P^{n_2}(x,y) > 0$. We deduce that for all $n \in \mathbb{N}$:

$$P^{n+n_1+n_2}(y,y) \geq P^{n_1}(y,x)P^n(x,x)P^{n_2}(x,y), \tag{3.10}$$

$$P^{n+n_1+n_2}(x,x) \geq P^{n_2}(x,y)P^n(y,y)P^{n_1}(y,x). \tag{3.11}$$

This implies that the sums $\sum_{n\in\mathbb{N}} P^n(x,x)$ and $\sum_{n\in\mathbb{N}} P^n(y,y)$ are both either converging or diverging. Thanks to properties (i) and (ii), we get that either x and y are both transient or both recurrent. This gives (iii).

We now prove property (iv). If C is an open communicating class, then there exist $x \in C$ and $y \notin C$ such that $P(x,y) > 0$. Since x is not accessible from y, we get $\mathbb{P}_y(T^x = \infty) = 1$. Using the Markov property, we get that $\mathbb{P}_x(T^x = \infty) \geq P(x,y)\mathbb{P}_y(T^x = \infty) > 0$. This gives that x is transient. □

According to property (iii) from Proposition 3.23, we get that an irreducible Markov chain is either transient or recurrent. And, in the former case the probability of $\{N^x < \infty\}$ is equal to 1 for all choice of the initial distribution. The next lemma asserts that for an irreducible recurrent Markov chain, the probability of $\{N^x < \infty\}$ is strictly less than 1 for all choice of the initial distribution.

Lemma 3.24. *Let X be an irreducible Markov chain on E. If X is transient, then $\mathbb{P}(N^x < \infty) = 1$ for all $x \in E$. If X is recurrent, then $\mathbb{P}(N^x = \infty) = 1$ for all $x \in E$.*

Proof. For the transient case, see property (ii) of Proposition 3.23. We assume that X is recurrent. Let $x \in E$. By decomposing according to the values of T^x and using the Markov property for the first equality and property (i) of Proposition 3.23 for the second, we get:

$$\mathbb{P}(N^x < \infty) = \mathbb{P}(T^x = \infty) + \sum_{n \in \mathbb{N}} \mathbb{P}(T^x = n)\mathbb{P}_x(N^x < \infty) = \mathbb{P}(T^x = \infty). \tag{3.12}$$

To conclude, we shall prove that $\mathbb{P}(T^x < \infty) = 1$. We get that for $m \in \mathbb{N}^*$:

$$1 = \mathbb{P}_x(X_n = x \text{ for some } n \geq m+1) = \sum_{y \in E} \mathbb{P}_x(X_m = y)\mathbb{P}_y(T^x < \infty),$$

where for the first equality we used that $\mathbb{P}_x(N^x = \infty) = 1$, and for the second the Markov property at time m and that $\mathbb{P}_y(X_n = x \text{ for some } n \geq 1) = \mathbb{P}_y(T^x < \infty)$. As $\sum_{y \in E} \mathbb{P}_x(X_m = y) = 1$ and $\mathbb{P}_y(T^x < \infty) \leq 1$, we deduce that $\mathbb{P}_y(T^x < \infty) = 1$ for all $y \in E$ such that $\mathbb{P}_x(X_m = y) > 0$. Since X is irreducible, for all $y \in E$, there exists $m \in \mathbb{N}^*$ such that $\mathbb{P}_x(X_m = y) > 0$. We deduce that $\mathbb{P}_y(T^x < \infty) = 1$ for all $y \in E$ and thus $\mathbb{P}(T^x < \infty) = 1$. Then use (3.12) to get $\mathbb{P}(N^x < \infty) = 0$. □

Remark 3.25. Let X be an irreducible Markov chain on a finite state space E. Since $\sum_{x \in E} N^x = \infty$ and E is finite, we deduce that $\mathbb{P}(N^x = \infty \text{ for some } x \in E) = 1$. This implies that $\mathbb{P}(N^x = \infty) > 0$ for some $x \in E$. We deduce from Lemma 3.24 that X is recurrent. Thus, all elements of a finite closed communicating class are recurrent. ◊

3.3.3 Periodicity

In Example 3.4 of the simple random walk $X = (X_n, n \in \mathbb{N})$, we notice that if X_0 is even (resp. odd), then X_{2n+1} is odd (resp. even) and X_{2n} is even (resp. odd) for $n \in \mathbb{N}$. Therefore the state space $\mathbb{Z}$ can be written as disjoint union of two sub-sets: the even integers, $2\mathbb{Z}$, and the odd integers, $2\mathbb{Z}+1$. And, a.s. the the Markov chain jumps from one sub-set to the other one. From the Lemma 3.28 below, we see that X has period 2 in this example.

Definition 3.26. *Let X be a Markov chain on E with transition matrix P. The period d of a state $x \in E$ is the greatest common divisor (GCD) of the set $\{n \in \mathbb{N}^*; P^n(x,x) > 0\}$, with the convention that $d = \infty$ if this set is empty. The state is aperiodic if $d = 1$.*

Notice that the set $\{n \in \mathbb{N}^*; P^n(x,x) > 0\}$ is empty if and only if $\mathbb{P}_x(T^x = \infty) = 1$, and that this also implies that $\{x\}$ is an open communicating class.

Proposition 3.27. *Let X be a Markov chain on E with transition matrix P. We have the following properties.*

(i) If $x \in E$ has a finite period d, then there exists $n_0 \in \mathbb{N}$ such that $P^{nd}(x,x) > 0$ for all $n \geq n_0$.

(ii) The elements of the same communicating class have the same period.

In view of (ii) above, we get that if X is irreducible, then all the states have the same finite period. For this reason, we shall say that an irreducible Markov chain is aperiodic (resp. has period d) if one of the states is aperiodic (resp. has period d).

Proof. We first consider the case $d = 1$. Let $x \in E$ be aperiodic. We consider the non-empty set $I = \{n \in \mathbb{N}^*; P^n(x,x) > 0\}$. Since $P^{n+m}(x,x) \geq P^n(x,x)P^m(x,x)$, we deduce that I is stable by addition. By hypothesis, there exist $n_1, \ldots, n_K \in I$ which are relatively prime. According to Bézout's lemma, there exist $a_1, \ldots, a_K \in \mathbb{Z}$ such that $\sum_{k=1}^K a_k n_k = 1$. We set $n_+ = \sum_{k=1; a_k>0}^K a_k n_k$ and $n_- = \sum_{k=1; a_k<0}^K |a_k| \, n_k$. If $n_- = 0$, then we deduce that $1 \in I$ and so (i) is proved with $n_0 = 1$. We assume now that $n_- \geq 1$. We get that $n_+, n_- \in I$ and $n_+ - n_- = 1$. Let $n \geq n_-^2$. Considering the Euclidean division of n by n_-, we get there exist integers $r \in \{0, \ldots, n_- - 1\}$ and $q \geq n_-$ such that:

$$n = qn_- + r = qn_- + r(n_+ - n_-) = (q - r)n_- + rn_+.$$

Since $q - r \geq 0$ and I is stable by addition, we get that $n \in I$. This proves property (i) with $n_0 = n_-^2$.

For $d \geq 2$ finite, consider $Q = P^d$. It is easy to check that x is then aperiodic when considering the Markov chain with transition matrix Q. Thus, there exists $n_0 \geq 1$, such that for $Q^n(x,x) > 0$ for all $n \geq n_0$, that is $P^{nd}(x,x) > 0$ for all $n \geq n_0$. This proves property (i).

Property (ii) is a direct consequence of property (i), (3.10) and (3.11). □

We give a natural interpretation of the period.

Lemma 3.28. *Let $X = (X_n, n \in \mathbb{N})$ be an irreducible Markov chain on E with period d. Then, there exists a partition $(E_i,\, i \in [\![0, d-1]\!])$ of E such that, with the convention $E_d = E_0$:*

$$\mathbb{P}_x(X_1 \in E_{i+1}) = 1 \quad \text{for all } i \in [\![0, d-1]\!] \text{ and } x \in E_i. \tag{3.13}$$

Proof. Since X is irreducible, we get that the period d is finite. Let $x_0 \in E$. Consider the sets $E_i = \{x \in E; \text{there exists } n \in \mathbb{N} \text{ such that } P^{nd+i}(x_0, x) > 0\}$ for $i \in [\![0, d-1]\!]$. Since X is irreducible, for $x \in E$ there exists $m \in \mathbb{N}$ such that $P^m(x_0, x) > 0$. This gives that $x \in E_i$ with $i = m \mod (d)$. We deduce that $E = \bigcup_{i=1}^{d-1} E_i$.

If $x \in E_i \bigcap E_j$, then using that $P^k(x, x_0) > 0$ for some $k \in \mathbb{N}$, we get there exists $n, m \in \mathbb{N}$ such that $P^{nd+i+k}(x_0, x_0) > 0$ and $P^{md+j+k}(x_0, x_0) > 0$. By definition of the period, we deduce that $i = j \mod (d)$. This implies that $E_i \bigcap E_j = \emptyset$ if $i \neq j$ and $i, j \in [\![0, d-1]\!]$.

To conclude, notice that if $x \in E_i$, that is $P^{nd+i}(x_0, x) > 0$ for some $n \in \mathbb{N}$, and $z \in E$ such that $P(x, z) > 0$, then we get that $P^{nd+i+1}(x_0, z) > 0$ and thus $z \in E_{i+1}$. This readily implies (3.13). Since $x_0 \in E_0$, we get that E_0 is non empty. Using (3.13), we get by recurrence that E_i for $i \in [\![0, d-1]\!]$) is non empty. Thus, $(E_i,\, i \in [\![0, d-1]\!])$ is a partition of E. □

The next lemma will be used in Section 3.4.3.

Lemma 3.29. *Let $X = (X_n, n \in \mathbb{N})$ and $Y = (Y_n, n \in \mathbb{N})$ be two independent Markov chains with respective discrete state spaces E and F. Then, the process $Z = ((X_n, Y_n), n \in \mathbb{N})$ is a Markov chain with state space $E \times F$. If π (resp. ν) is an invariant probability measure for X (resp. Y), then $\pi \otimes \nu$ is an invariant probability measure for Z. If X and Y are irreducible and furthermore X or Y is aperiodic, then Z is irreducible on $E \times F$.*

Proof. Let P and Q be the transition matrix of X and Y. Using the independence of X and Y, it is easy to prove that Z is a Markov chain with transition matrix R given by $R(z, z') = P(x, x')Q(y, y')$ with $z = (x, y), z' = (x', y') \in E \times F$.

If π (resp. ν) is an invariant measure for X (resp. Y), then we have for $z = (x, y) \in E \times F$:

$$(\pi\otimes\nu)R(z) = \sum_{x'\in E, y'\in F} \pi(x')\nu(y')R((x', y'), (x, y)) = \sum_{x'\in E, y'\in F} \pi(x')\nu(y')P(x', x)Q(y', y) = \pi\otimes\nu(z).$$

Therefore the probability measure $\pi \otimes \nu$ is invariant for Z.

Let us assume that X is aperiodic and irreducible and that Y is irreducible. Let $z = (x, y), z' = (x', y') \in E \times F$. Since X and Y are irreducible, there exists $n_1, n_2, n_3 \in \mathbb{N}^*$ such that $P^{n_1}(x, x') > 0$, $Q^{n_2}(y, y') > 0$ and $Q^{n_3}(y', y') > 0$. Property (i) of Proposition 3.27 gives that $P^{kn_3+n_2-n_1}(x', x') > 0$ for $k \in \mathbb{N}^*$ large enough. Thus, we get for k large enough:

$$\begin{aligned} R^{kn_3+n_2}(z, z') &= P^{kn_3+n_2}(x, x')Q^{kn_3+n_2}(y, y') \\ &\geq P^{n_1}(x, x')P^{kn_3+n_2-n_1}(x', x')Q^{n_2}(y, y')Q^{n_3}(y', y')^k > 0. \end{aligned}$$

We deduce that Z is irreducible. We get the same result if Y is aperiodic instead of X. □

3.4 Asymptotic theorems

3.4.1 Main results

Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain on a discrete state space E. We recall the first return time of x is given by $T^x = \inf\{n \geq 1; X_n = x\}$. Since $T^x \geq 1$, we get that $\mathbb{E}_x[T^x] \geq 1$. We set for $x \in E$:

$$\pi(x) = \frac{1}{\mathbb{E}_x[T^x]} \in [0, 1]. \tag{3.14}$$

For an irreducible transient Markov chain, we recall that $\mathbb{P}_x(T^x = +\infty) > 0$ and thus $\mathbb{E}_x[T^x] = +\infty$ for all $x \in E$, so that $\pi = 0$.

Definition 3.30. *A recurrent state $x \in E$ is null recurrent if $\pi(x) = 0$ and positive recurrent if $\pi(x) > 0$. The Markov chain is null (resp. positive) recurrent if all the states are null (resp. positive) recurrent.*

We shall consider asymptotic events whose probability depends only on the transition matrix and not on the initial distribution of the Markov chain. This motivates the following definition. An event $A \in \sigma(X)$ is said to be almost sure (a.s.) for a Markov chain $X = (X_n, n \in \mathbb{N})$ if $\mathbb{P}_x(A) = 1$ for all starting state $x \in E$ of X, or equivalently $\mathbb{P}_{\mu_0}(A) = 1$ for all initial distribution μ_0 of X_0.

The next two fundamental theorems on the asymptotics of irreducible Markov chain will be proved in Section 3.4.3.

Theorem 3.31. *Let $X = (X_n, n \in \mathbb{N})$ be an irreducible Markov chain on E. Let π be given by (3.14).*

(i) The Markov chain X is either transient or null recurrent or positive recurrent.

(ii) *If the Markov chain is transient or null recurrent, then there is no invariant probability measure. Furthermore, we have $\pi = 0$.*

(iii) *For all $x \in E$, we have:*

$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_k=x\}} \xrightarrow[n\to\infty]{\text{a.s.}} \pi(x). \tag{3.15}$$

The next result is specifically on irreducible positive recurrent Markov chain. The definition of the convergence in distribution of sequence of random variables and some of its characterization are given in Section 7.2.1.

Theorem 3.32 (Ergodic theorem). *Let $X = (X_n, n \in \mathbb{N})$ be an irreducible positive recurrent Markov chain on E.*

(i) *The measure π defined by (3.14) is the unique invariant probability of X. (And we have $\pi(x) > 0$ for all $x \in E$.)*

(ii) *For all real-valued function f defined on E such that (π, f) is well defined, we have:*

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow[n\to\infty]{\text{a.s.}} (\pi, f). \tag{3.16}$$

(iii) *If X is aperiodic, then we have the convergence in distribution $X_n \xrightarrow[n\to\infty]{(d)} \pi$ and:*

$$\lim_{n\to\infty} \sum_{y\in E} |P^n(x,y) - \pi(y)| = 0 \quad \text{for all } x \in E. \tag{3.17}$$

In particular for an irreducible positive recurrent Markov chain, the empirical mean or time average converges a.s. to the spatial average with respect to the invariant probability measure. In the aperiodic case, we also get that the asymptotic behavior of the Markov chain is given by the stationary regime. We give the following easy to remember corollary.

Corollary 3.33. *An irreducible Markov chain $X = (X_n, n \in \mathbb{N})$ on a finite state space is positive recurrent: π defined by (3.14) is its unique invariant probability measure, $\pi(x) > 0$ for all $x \in E$ and (3.16) holds for all $\mathbb{R}$-valued function f defined on E. If furthermore X is aperiodic, then the sequence $(X_n, n \in \mathbb{N})$ converges in distribution towards π.*

Proof. Summing (3.15) over $x \in E$, we get that $\sum_{x\in E} \pi(x) = 1$. Thus the Markov chain is positive recurrent according to Theorems 3.31, properties (i)-(ii), and 3.32, property (i). The remaining part of the corollary is a direct consequence of Theorem 3.32. □

The convergences of the empirical means, see (3.16), for irreducible positive recurrent Markov chains is a generalization of the strong law of large number recalled in Section 7.2.2. Indeed, if $X = (X_n, n \in \mathbb{N})$ is a sequence of independent random variables taking values in E with the same distribution π, then, X is a Markov chain with transition matrix P whose lines are all equal to π (that is $P(x, y) = \pi(y)$ for all $x, y \in E$). Notice then that P is reversible with respect to π. Assume for simplicity that $\pi(x) > 0$ for all $x \in E$ so that X is irreducible with invariant probability π. Then (3.16) corresponds exactly to the strong law of large numbers. By the way, the initial motivation of the introduction of Markov chains by Markov[5] in 1906 was to extend the law of large number and the central limit theorem (CLT) to sequences of dependent random variables.

Eventually notice that the limits in (3.16) or in (iii) of Theorem 3.31 does not involve the initial distribution of the Markov chain. Forgetting the initial condition is an important property of the Markov chains.

3.4.2 Complement on the asymptotic results

We shall state without proof some results on the CLT for irreducible positive recurrent Markov chain and on invariant measures for irreducible null recurrent Markov chain.

On the CLT in the positive recurrent case

Similarly to the CLT for sequences of independent random variables with the same distribution, see Section 7.2.2, it is possible to provide the fluctuations associated to (3.16) under reasonable assumptions. Let $X = (X_n, n \in \mathbb{N})$ be an irreducible positive recurrent Markov chain on E with transition matrix P and invariant probability measure π. Set $I_n(f) = \frac{1}{n}\sum_{k=1}^n f(X_k)$ for $n \in \mathbb{N}^*$ and f a real-valued function defined on E such that (π, f^2) is finite. Thanks to Theorem 3.32, we have that a.s. $\lim_{n\to\infty} I_n(f) = (\pi, f)$. Without loss of generality, we assume that:

$$(\pi, f) = 0.$$

As in the CLT for independent random variables, we expect the convergence in distribution of $(\sqrt{n}\, I_n(f), n \in \mathbb{N}^*)$ towards a centered Gaussian random variable. With this idea in mind, it is natural to consider the variance of $\sqrt{n}\, I_n(f)$:

$$\begin{aligned}
\operatorname{Var}\left(\sqrt{n}\, I_n(f)\right) &= \frac{1}{n}\sum_{k,\ell=1}^{n} \operatorname{Cov}(f(X_k), f(X_\ell)) \\
&= \frac{1}{n}\sum_{k=1}^{n}\left(\operatorname{Var}(f(X_k)) + 2\sum_{j=1}^{n-k}\operatorname{Cov}(f(X_k), f(X_{k+j}))\right).
\end{aligned}$$

It is legitimate to expect that the variance of the limit Gaussian random variable is the limit of $\operatorname{Var}(\sqrt{n}\, I_n(f))$ and as the mean in time correspond intuitively to the average under the invariant probability measure, this would be, as $(\pi, f) = 0$:

$$\sigma(f)^2 = \mathbb{E}_\pi\left[f^2(X_0)\right] + 2\mathbb{E}_\pi\Big[\sum_{j\in\mathbb{N}^*} f(X_0)f(X_j)\Big] = \Big(\pi, f^2\Big) + 2\Big(\pi, \sum_{j\in\mathbb{N}^*} fP^j f\Big). \tag{3.18}$$

To be precise, we state Theorems II.4.1 and II.4.3 from [1]. For $x \in E$, set:

$$H_n(x) = \sum_{y\in E} |P^n(x,y) - \pi(y)|.$$

We recall that according to (3.17), if X is aperiodic then we have the ergodicity property $\lim_{n\to\infty} H_n(x) = 0$ for all $x \in E$.

[5] G. Basharin, A. Langville, V. Naumov. The life and work of A.A. Markov. *Linear Algebra and its Applications*, vol. 386, pp. 3-26, 2004

Theorem. *Let X be an irreducible positive recurrent and aperiodic Markov chain with invariant probability measure π. Let f be a real-valued function defined on E such that $(\pi, f^2) < +\infty$ and $(\pi, f) = 0$. If one of the two following conditions is satisfied:*

(i) f is bounded and $\sum_{n\in\mathbb{N}^*}(\pi, H_n) < +\infty$ (ergodicity of degree 2);

(ii) $\lim_{n\to\infty}\sup_{x\in E} H_n(x) = 0$ (uniform ergodicity);

Then, $\sigma(f)^2$ given by (3.18) is finite and non-negative, and:

$$\sqrt{n}\, I_n(f) \xrightarrow[n\to\infty]{(d)} \mathcal{N}\left(0, \sigma(f)^2\right).$$

Usually the variance $\sigma(f)^2$ is positive, but for some particular Markov chain and particular function f, it may be null. Concerning the hypothesis (i) and (ii) in the previous theorem, we also mention that uniform ergodicity implies there exists $c > 1$ such that $\sup_{x\in E} H_n(x) \leq c^{-n}$ for large n, which in turns implies the ergodicity of degree 2. Notice that if the state space E is finite, then an irreducible aperiodic Markov chain is uniformly ergodic.

Based on the excursion approach developed in Section 3.4.3, it is also possible to give an alternative result for the CLT of Markov chains, see Theorems 17.2.2, 17.4.4 and 17.5.3 in [7]. For f a real-valued function defined on E and $x \in E$, we set, when it is well defined:

$$S_x(f) = \sum_{k=1}^{T^x} f(X_k).$$

Theorem. *Let X be an irreducible positive recurrent Markov chain with invariant probability measure π. Let f be a real-valued function defined on E such that (π, f) is well defined with $(\pi, f) = 0$. Let $x \in E$ such that $\mathbb{E}_x[S_x(1)^2] = \mathbb{E}_x[(T^x)^2] < +\infty$ and $\mathbb{E}_x[S_x(|f|)^2] < +\infty$ (so that $S_x(f)$ is a.s. well defined). Set*

$$\sigma'(f)^2 = \pi(x)\mathbb{E}_x\left[S_x(f)^2\right]. \tag{3.19}$$

Then, we have that:

$$\sqrt{n}\, I_n(f) \xrightarrow[n\to\infty]{(d)} \mathcal{N}\left(0, \sigma'(f)^2\right).$$

Furthermore (3.19) holds for all $x \in E$.

An other approach is based on the Poisson equation. Assume $(\pi, |f|)$ is finite. We say that a $\mathbb{R}$-valued function $\hat{f}$ is a solution to the Poisson equation if $P\hat{f}$ is well defined and:

$$\hat{f} - P\hat{f} = f - (\pi, f). \tag{3.20}$$

Theorem. *Let X be an irreducible positive recurrent Markov chain with invariant probability measure π. Let f be a real-valued function defined on E such that $(\pi, |f|) < +\infty$ and $(\pi, f) = 0$. Assume there exists a solution $\hat{f}$ to the Poisson equation such that $(\pi, \hat{f}^2) < +\infty$. Set*

$$\sigma''(f)^2 = \left(\pi, \hat{f}^2 - (P\hat{f})^2\right). \tag{3.21}$$

Then we have that:

$$\sqrt{n}\, I_n(f) \xrightarrow[n\to\infty]{(d)} \mathcal{N}\left(0, \sigma''(f)^2\right).$$

Of course, the asymptotic variances given by (3.18), (3.19) and (3.21) coincide when the hypothesis of the three previous theorem hold. This is in particular the case if E is finite (even if X is periodic).

More on the null recurrent case

It is possible to have more precise ergodic results for irreducible null recurrent Markov chains, but with less natural probabilistic interpretation.

Let ν be a measure on E such that $\nu \neq 0$ and $\nu(x) < +\infty$ for all $x \in E$. We say that ν is an invariant measure for a stochastic matrix P if $\nu P = \nu$. This generalizes Definition 3.15. It can be proved that if X is an irreducible positive recurrent Markov chain, then the only invariant measures are $\lambda\pi$ where π is the invariant probability measure and $\lambda > 0$.

Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain. For $x \in E$, we define the measure ν_x by:

$$\nu_x(y) = \mathbb{E}_x\left[\sum_{k=1}^{T^x} \mathbf{1}_{\{X_k=y\}}\right] \quad \text{for } y \in E.$$

Notice the measure ν_x is infinite as $(\nu_x, \mathbf{1}) = \mathbb{E}_x[T^x] = +\infty$. According to [2, 1, 4], we have the following results.

Theorem. *Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain with transition matrix P. If x is recurrent then ν_x is an invariant measure for P.*

If furthermore X is irreducible null recurrent, then we get the following results:

(i) The measure ν_x is the only invariant measure (up to a positive multiplicative constant) and $\nu_x(y) > 0$ for all $y \in E$. And for all $y, z \in E$, we have $\nu_y(z) = \nu_x(z)/\nu_x(y)$.

(ii) For all $\mathbb{R}$-valued functions f, g defined on E such that (ν, f) is well defined and g is non-negative with $0 < (\nu, g) < +\infty$, we have:

$$\frac{\sum_{k=1}^n f(X_k)}{\sum_{k=1}^n g(X_k)} \xrightarrow[n\to\infty]{a.s.} \frac{(\nu, f)}{(\nu, g)}.$$

(iii) We have $\lim_{n\to+\infty} \mathbb{P}(X_n = y) = 0$ for all $y \in E$.

For irreducible transient Markov chain, there is no simple answer on the existence or uniqueness of invariant measure, see the two exercises below; furthermore notice that the sum $\sum_{k=1}^n \mathbf{1}_{\{X_k=x\}}$ is constant for large n and that $\lim_{n\to\infty} \mathbb{P}(X_n = x) = 0$ for all $x \in E$.

3.4.3 Proof of the asymptotic theorems

Let $X = (X_n, n \geq 0)$ be an **irreducible** Markov chain on a discrete state space E, with transition matrix P. Recall the measure π defined by (3.14). The next lemma insures that if there exists an invariant probability measure, then it has to be π.

Lemma 3.34. *Let X be an irreducible Markov chain. If (3.15) holds and ν is an invariant probability measure, then we have $\nu = \pi$.*

Proof. Assume that ν is an invariant probability measure. Since the left hand-side member of (3.15) is bounded by 1, using dominated convergence and taking the expectation in (3.15) with ν as initial distribution of X, we get that for all $x \in E$:

$$\frac{1}{n}\sum_{k=1}^n \nu P^k(x) \xrightarrow[n\to\infty]{} \pi(x).$$

Since ν is invariant, we get $\nu P^k = \nu$. We deduce that $\nu(x) = \pi(x)$ for all $x \in E$. □

The next results is on transient Markov chains.

Lemma 3.35. *Let X be an irreducible transient Markov chain. We have: $\pi = 0$, (3.15) holds and X has no invariant probability measure.*

Proof. Property (ii) of Proposition 3.23 implies that $\pi = 0$ and that $\sum_{k\in\mathbb{N}} \mathbf{1}_{\{X_k=x\}} = N^x$ is a.s. finite. We deduce that (3.15) holds. Then use that $\pi = 0$ and Lemma 3.34 to deduce that X has no invariant probability measure. □

From now on we assume that X is **irreducible** and **recurrent**.

Let $x \in E$ be fixed. Lemma 3.24 gives that a.s. the number of visit of x is a.s. infinite. We can thus define a.s. the successive return times to x. By convention, we write $T_0^x = 0$ and for $n \in \mathbb{N}$:

$$T_{n+1}^x = \inf\{k > T_n^x;\, X_k = x\}.$$

We define the successive excursions $(Y_n, n \in \mathbb{N}^*)$ out of the state x as follows:

$$Y_n = (T_n^x - T_{n-1}^x, X_{T_{n-1}^x}, X_{T_{n-1}^x+1}, \dots, X_{T_n^x}). \tag{3.22}$$

The random variable Y_n describes the n-th excursion out for the state x. Notice that x is the end of the excursion, that is $X_{T_n^x} = x$, and for $n \geq 2$ it is also the starting point of the excursion as $X_{T_{n-1}^x} = x$. So Y_n takes values in the discrete space $E^{\text{traj}} = \cup_{k\in\mathbb{N}^*}\{k\}\times E^k \times \{x\}$. The next lemma is the key ingredient to prove the asymptotic results for recurrent Markov chains.

Lemma 3.36. *Let X be an irreducible recurrent Markov chain. The random variables $(Y_n, n \in \mathbb{N}^*)$ defined by (3.22) are independent. And the random variables $(Y_n, n \geq 2)$ are all distributed as Y_1 under $\mathbb{P}_x$.*

Proof. For $y = (r, x_0, \dots, x_r) \in E^{\text{traj}}$, we set $t_y = r$ the length of the excursion and we recall that the end point of the excursion is equal to x: $x_r = x$. We shall first prove that for all $n \in \mathbb{N}^*$, $y_1, \dots, y_n \in E^{\text{traj}}$, we have:

$$\mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n) = \mathbb{P}(Y_1 = y_1) \prod_{k=2}^{n} \mathbb{P}_x(Y_1 = y_k). \tag{3.23}$$

For $n = 1$ and $y_1 \in E^{\text{traj}}$, Equation (3.23) holds trivially. Let $n \geq 2$ and $y_1, \dots, y_n \in E^{\text{traj}}$. On the event $\{Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}\}$, the time $s = \sum_{k=1}^{n-1} t_{y_k}$ is the end of the $n-1$-th excursion, and at this time we have $X_s = x$ as all the excursions end at state x. Using the Markov property at time s and that $X_s = x$ on $\{Y_1 = y_1, \dots, Y_{n-1} = y_{n-1}\}$, we get that:

$$\mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n) = \mathbb{P}(Y_1 = y_1, \dots, Y_{n-1} = y_{n-1})\mathbb{P}_x(Y_1 = y_n).$$

Then, we get (3.23) by induction. Use Definition 1.31 and (3.23) for any $n \in \mathbb{N}^*$ and $y_1, \dots, y_n \in E^{\text{traj}}$ to conclude. □

We will now prove (3.15) for irreducible recurrent Markov chains. This and Lemma 3.35 will give property (iii) from Theorem 3.31.

Proposition 3.37. *Let X be an irreducible recurrent Markov chain. Then (3.15) holds.*

Proof. Let $x \in E$ be fixed. Since $T_n^x = T_1^x + \sum_{k=2}^n (T_k^x - T_{k-1}^x)$, with T_1^x a.s. finite, and $(T_k^x - T_{k-1}^x, n \geq 2)$ are, according to Lemma 3.36, independent positive random variables distributed as T^x under $\mathbb{P}_x$, we deduce from the law of large number, see Theorem 7.15, that:

$$\frac{T_n^x}{n} \xrightarrow[n\to\infty]{\textbf{a.s.}} \mathbb{E}_x[T^x]. \tag{3.24}$$

We define the number of visit of x from time 1 to $n \in \mathbb{N}^*$:

$$N_n^x = \sum_{k=1}^n \mathbf{1}_{\{X_k = x\}}. \tag{3.25}$$

By construction, we have:

$$T_{N_n^x}^x \leq n < T_{N_n^x+1}^x. \tag{3.26}$$

This gives $\dfrac{N_n^x}{N_n^x+1}\dfrac{N_n^x+1}{T_{N_n^x+1}^x} \leq \dfrac{N_n^x}{n} \leq \dfrac{N_n^x}{T_{N_n^x}^x}$. Since x is recurrent, we get that a.s. $\lim_{n\to\infty} N_n^x = +\infty$. We deduce from (3.24) that a.s. $\lim_{n\to\infty} N_n^x/n = 1/\mathbb{E}_x[T^x] = \pi(x)$. □

Next lemma and property (iii) of Proposition 3.23 give property (i) of Theorem 3.31.

Lemma 3.38. *Let X be an irreducible recurrent Markov chain. Then, it is either null recurrent or positive recurrent.*

Proof. Let $x \in E$. Notice the left hand-side of (3.15) is bounded by 1. Integrating (3.15) with respect to $\mathbb{P}_x$, we get by dominated convergence that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n P^k(x,x) = \pi(x)$. Since X is irreducible, we deduce from (3.11), that if the above limit is zero for a given x, it is zero for all $x \in E$. This implies that either $\pi = 0$ or $\pi(x) > 0$ for all $x \in E$. □

The proof of the next lemma is a direct consequence of Lemma 3.34 and the fact that $\pi = 0$ for irreducible null recurrent Markov chains.

Lemma 3.39. *Let X be an irreducible null recurrent Markov chain. Then, there is no invariant probability measure.*

Lemmas 3.35 and 3.39 imply property (ii) of Theorem 3.31. This ends the proof of Theorem 3.31.

The end of this section is devoted to the proof of Theorem 3.32. From now on we assume that X is **irreducible** and **positive recurrent**.

Proposition 3.40. *Let X be an irreducible positive recurrent Markov chain. Then, the measure π defined in (3.14) is a probability measure. For all real-valued function f defined on E such that (π, f) is well defined, we have (3.16).*

Proof. Let $x \in E$. We keep notations from the proof of Lemma 3.36. Let f be a finite non-negative function defined on E. We set for $y = (r, x_0, \dots, x_r) \in E^{\text{traj}}$:

$$F(y) = \sum_{k=1}^r f(x_k).$$

According to Lemma 3.36, the random variables $(F(Y_n), n \geq 2)$ are independent non-negative and distributed as $F(Y_1)$ under $\mathbb{P}_x$. As $F(Y_1)$ is finite, we deduce from the law of large number, see Theorem 7.15, that a.s. $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n F(Y_k) = \mathbb{E}_x[F(Y_1)]$. Since $\sum_{i=1}^{T_n^x} f(X_i) = \sum_{k=1}^n F(Y_k)$, we deduce from (3.24) that:

$$\frac{1}{T_n^x}\sum_{i=1}^{T_n^x} f(X_i) = \frac{n}{T_n^x}\,\frac{1}{n}\sum_{k=1}^n F(Y_k) \xrightarrow[n\to\infty]{\text{a.s.}} \pi(x)\,\mathbb{E}_x[F(Y_1)].$$

Recall that $T^x_{N_n^x} \leq n < T^x_{N_n^x+1}$ from (3.26). Since f is non-negative, we deduce that:

$$\frac{T^x_{N_n^x}}{T^x_{N_n^x+1}}\,\frac{1}{T^x_{N_n^x}}\sum_{i=1}^{T^x_{N_n^x}} f(X_i) \leq \frac{1}{n}\sum_{i=1}^n f(X_i) \leq \frac{T^x_{N_n^x+1}}{T^x_{N_n^x}}\,\frac{1}{T^x_{N_n^x+1}}\sum_{i=1}^{T^x_{N_n^x+1}} f(X_i).$$

Since a.s. $\lim_{n\to\infty} N_n^x = +\infty$, $\lim_{n\to\infty} T_n^x = +\infty$ and $\lim_{n\to+\infty} T_n^x/T_{n+1}^x = 1$, see (3.24), we deduce that:

$$\frac{1}{n}\sum_{i=1}^n f(X_i) \xrightarrow[n\to\infty]{\text{a.s.}} \pi(x)\mathbb{E}_x[F(Y_1)]. \tag{3.27}$$

Taking $f = \mathbf{1}_{\{y\}}$ in the equation above, we deduce from (3.15) that:

$$\pi(y) = \pi(x)\mathbb{E}_x\Big[\sum_{k=1}^{T^x} \mathbf{1}_{\{X_k=y\}}\Big]. \tag{3.28}$$

Summing over $y \in E$, we get by monotone convergence that $\sum_{y\in E}\pi(y) = \pi(x)\mathbb{E}_x[T^x] = 1$. This gives that π is a probability measure. By monotone convergence, we deduce from (3.28) that:

$$\pi(x)\mathbb{E}_x[F(Y_1)] = \sum_{y\in E} f(y)\pi(x)\mathbb{E}_x\Big[\sum_{k=1}^{T^x}\mathbf{1}_{\{X_k=y\}}\Big] = \sum_{y\in E} f(y)\pi(y) = (\pi, f).$$

Using then (3.27), we deduce that (3.16) holds when f is finite and non-negative. If f is non-negative but not finite, the result is immediate as $N_x = \infty$ a.s. for all $x \in E$ and $(\pi, f) = +\infty$. If f is real-valued such that (π, f) is well defined, then considering (3.16) with f replaced by f^+ and f^-, and making the difference of the two limits, we get (3.16). □

Next Proposition and Proposition 3.40 give properties (i) and (ii) of Theorem 3.32.

Proposition 3.41. *Let X be an irreducible positive recurrent Markov chain. Then, the measure π defined in (3.14) is the unique invariant probability measure.*

Proof. According to Proposition 3.40, the measure π is a probability measure. We now check it is invariant. Let μ be the distribution of X_0. We set:

$$\bar{\mu}_n(x) = \frac{1}{n}\sum_{k=1}^n \mu P^k(x).$$

By dominated convergence, taking the expectation in (3.15) with respect to $\mathbb{P}_\mu$, we get $\lim_{n\to\infty}\bar{\mu}_n(x) = \pi(x)$ for all $x \in E$. Similarly, using (3.16) with f bounded, we get that $\lim_{n\to\infty}(\bar{\mu}_n, f) = (\pi, f)$.

Let $y \in E$ be fixed and $f(\cdot) = P(\cdot, y)$. We notice that $\lim_{n\to\infty}(\bar{\mu}_n, f) = (\pi, f) = \pi P(y)$ and that $(\bar{\mu}_n, f) = \bar{\mu}_n P(y) = \frac{n+1}{n}\bar{\mu}_{n+1}(y) - \frac{1}{n}\,\mu P(y)$. Letting n goes to infinity in those equalities, we get that $\pi P(y) = \pi(y)$. Since y is arbitrary, we deduce that π is invariant. By Lemma 3.34, this is the unique invariant probability measure. □

The next proposition and Lemma 7.14 give property (iii) from Theorem 3.32. Its proof relies on a coupling argument.

Proposition 3.42. *An irreducible positive recurrent aperiodic Markov chain converges in distribution towards its unique invariant probability measure.*

Proof. Let $X = (X_n, n \in \mathbb{N})$ be an irreducible positive recurrent aperiodic Markov chain. Recall that π defined in (3.14) is its unique invariant probability measure. Let $Y = (Y_n, n \in \mathbb{N})$ be a Markov chain independent of X with the same transition matrix and initial distribution π. Thanks to Lemma 3.29, the Markov chain $Z = ((X_n, Y_n), n \in \mathbb{N})$ is irreducible and it has $\pi \otimes \pi$ has invariant probability measure. This gives that Z is positive recurrent.

Let $x \in E$ and consider $T = \inf\{n \geq 1;\, X_n = Y_n = x\}$ the return time of Z to (x, x). For $y \in E$, we have:

$$\mathbb{P}(X_n = y) = \mathbb{P}(X_n = y, T \leq n) + \mathbb{P}(X_n = y, T > n) \leq \mathbb{P}(X_n = y, T \leq n) + \mathbb{P}(T > n).$$

Decomposing according to the events $\{T = k\}$ for $k \in \mathbb{N}^*$, and using that $X_k = x = Y_k$ on $\{T = k\}$, that X and Y have the same transition matrix, as well as the Markov property at time k, we get that $\mathbb{P}(X_n = y, T \leq n) = \mathbb{P}(Y_n = y, T \leq n)$. Thus, we obtain:

$$\mathbb{P}(X_n = y) \leq \mathbb{P}(Y_n = y, T \leq n) + \mathbb{P}(T > n) \leq \mathbb{P}(Y_n = y) + \mathbb{P}(T > n).$$

By symmetry we can replace (X_n, Y_n) in the previous inequality by (Y_n, X_n) and deduce that:

$$|\mathbb{P}(X_n = y) - \mathbb{P}(Y_n = y)| \leq \mathbb{P}(T > n).$$

Since Z is recurrent, we get that a.s. T is finite. Using that $\mathbb{P}(Y_n = y) = \pi(y)$, as π is invariant and the initial distribution of Y, we deduce that $\lim_{n\to\infty} |\mathbb{P}(X_n = y) - \pi(y)| = 0$ for all $y \in E$. Then, use Lemma 7.14 to conclude. □

3.5 Examples and applications

In this section, we give some well known examples of Markov chains.

Random walk on $\mathbb{Z}^d$

Let $d \in \mathbb{N}^*$. Let U be a $\mathbb{Z}^d$-valued random variables with probability distribution $p = (p(x) = \mathbb{P}(U = x), x \in \mathbb{Z}^d)$. Let $(U_n, n \in \mathbb{N}^*)$ be a sequence of independent random variables distributed as U, and X_0 a $\mathbb{Z}^d$-valued independent random variable. We consider the random walk $X = (X_n, n \in \mathbb{N})$ with increments distributed as U defined by:

$$X_n = X_0 + \sum_{k=1}^{n} U_k \quad \text{for } n \in \mathbb{N}^*.$$

The transition matrix P of X is given by $P(x,y) = p(y-x)$. We assume that X is irreducible (equivalently the smallest additive sub-group of $\mathbb{Z}^d$ which contains the support $\{x \in \mathbb{Z}^d;\, p(x) > 0\}$ is $\mathbb{Z}^d$). Because $\sum_{x\in\mathbb{Z}^d} P(x,y) = 1$, we deduce that the counting measure on $\mathbb{Z}^d$ is invariant. (According to Section 3.4.2, this implies that irreducible random walks are transient or null recurrent.) We refer to [8, 6] for a detailed account on random walks.

The simple symmetric random walk corresponds to U being uniform on the set of cardinal $2d$: $\{x \in \mathbb{Z}^d;\, |x| = 1\}$, with $|x|$ denoting the Euclidean norm on $\mathbb{R}^d$. It is irreducible with period 2 (as $P^2(x,x) > 0$ and by parity $P^{2n+1}(x,x) = 0$ for all $n \in \mathbb{N}$).

We summarize the main results on transience/recurrence for random walks, see [8] Theorem 8.1.

Theorem. *Let X be an irreducible random walk on $\mathbb{Z}^d$ with increments distributed as U. We have the following results:*

(i) If $d = 1$, $U \in L^1$ and $\mathbb{E}[U] = 0$, then X is null recurrent.

(ii) If $d = 2$, $U \in L^2$ and $\mathbb{E}[U] = 0$, then X is null recurrent.

(iii) If $d = 3$, then X is transient.

Metropolis-Hastings algorithm

Let π be a given probability distribution on E such that $\pi(x) > 0$ for all $x \in E$. The aim of the Metropolis-Hastings[6] algorithm is to simulate a random variable with distribution (asymptotically close to) π.

We consider a stochastic matrix Q on E which is irreducible (that is for all $x, y \in E$, there exists $n \in \mathbb{N}^*$ such that $Q^n(x,y) > 0$) and such that for all $x, y \in E$, if $Q(x,y) = 0$ then $Q(y,x) = 0$. The matrix Q is called the selection matrix.

We say a function $\rho = (\rho(x,y);\, x, y \in E$ such that $Q(x,y) > 0)$ taking values in $(0,1]$ is an accepting probability function if for $x, y \in E$ such that $Q(x,y) > 0$, we have:

$$\rho(x,y)\pi(x)Q(x,y) = \rho(y,x)\pi(y)Q(y,x). \tag{3.29}$$

An example of an accepting probability function is given by:

$$\rho(x,y) = \gamma\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right) \quad \text{for } x, y \in E \text{ such that } Q(x,y) > 0, \tag{3.30}$$

where γ is a function defined on $(0,+\infty)$ taking values in $(0,1]$ satisfying $\gamma(u) = u\gamma(1/u)$ for $u > 0$. A common choice for γ is $\gamma(u) = \min(1,u)$ (Metropolis algorithm) or $\gamma(u) = u/(1+u)$ (Boltzmann or Barker algorithm).

We now describe the Metropolis-Hastings algorithm. Let X_0 be a random variable on E with distribution μ_0. At step $n+1$, the random variables $X_0, \ldots, X_n$ are defined, and we explain how to generate X_{n+1}. First consider a random variable Y_{n+1} with distribution $Q(X_n, \cdot)$. With probability $\rho(X_n, Y_{n+1})$, we accept the transition and set $X_{n+1} = Y_{n+1}$. If the

[6] W. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, vol. 57, pp.97-109, 1970.

transition is rejected, we set $X_{n+1} = X_n$. By construction $X = (X_n, n \in \mathbb{N})$ is a stochastic dynamical system and thus a Markov chain. Its transition matrix is given by, for $x, y \in E$:

$$P(x,y) = \begin{cases} Q(x,y)\rho(x,y) & \text{if } x \neq y, \\ 1 - \sum_{z \neq x} P(x,z) & \text{if } x = y. \end{cases}$$

Since $\rho > 0$, we have that $Q(x,y) > 0$ implies $P(x,y) > 0$ and, for $x \neq y$, that $Q(x,y) > 0$ is equivalent to $P(x,y) > 0$. We deduce that X is irreducible as Q is irreducible. Condition (3.29) implies that X is reversible with respect to the probability π. Thus, the Markov chain is irreducible recurrent positive with invariant probability π. Let f be a real-valued function f defined on E, such that (π, f) is well defined. An approximation of (π, f), is according to Theorem 3.31, given by $\frac{1}{n}\sum_{k=1}^{n} f(X_k)$ for n large. The drawback of this approach is that it does not come with a confidence interval of (π, f). If furthermore either Q is aperiodic or there exists $x, y \in E$ such that $Q(x,y) > 0$ and $\rho(x,y) < 1$ so that $P(x,x) > 0$, then the Markov chain X is aperiodic. In this case, Theorem 3.32 implies then that X converges in distribution towards π.

It may happens that π is known up to a normalizing constant. This is the case of the so called Boltzmann or Gibbs measure in statistical physics for example, where E is the state space of a system, and the probability for the system to be in configuration $x \in E$ is $\pi(x) = Z_T^{-1} \exp(-H(x)/T)$, where $H(x)$ is the energy of the system in configuration x, T the temperature and Z_T the normalizing constant. It is usually very difficult to compute an approximation of Z_T.

When using the accepting probability function given by (3.30), then only the ratio $\pi(x)/\pi(y)$ is needed to be computed to simulate X. In particular, the simulation does not rely on the value of Z_T.

Wright Fisher model

The Wright-Fisher model for population evolution has been introduced by Fisher in 1922[7] and Wright[8] in 1931. Consider a population of constant size N with individuals with one time unit of lifetime and which reproduce at each unit of time. We assume the reproduction is random, and there is no mating (each individual can have children). More formally, if $Y_i^n \in [\![1, N]\!]$ is the parent of individual i at generation $n \in \mathbb{N}^*$ alive at generation $n-1$, then the random variables $(Y_i^{n+1}, i \in [\![1, N]\!], n \in \mathbb{N})$ are independent uniformly distributed on $[\![1, N]\!]$. Intuitively, each child chooses independently and uniformly its parent.

We assume that individuals may be either of type A or type a, and that a child inherit the type from its parent. Let X_n be the number of the individuals at time n of type A. By construction $X = (X_n, n \in \mathbb{N})$ is a Markov chain on $E_N = [\![0, N]\!]$. Conditionally on X_n, each child at generation $n+1$ has probability X_n/N to be of type A. Thus the distribution of X_{n+1} has conditionally on X_n a binomial distribution with parameter $(X_n/N, N)$. The transition matrix P_N is thus given by:

$$P_N(i,j) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j} \quad \text{for } i, j \in E_N.$$

[7] R. A. Fisher. On the dominance ratio. *Proc. Roy. Soc. Edinburgh*, vol. 42, pp. 321-341, 1922.

[8] S. Wright. Evolution in Mendelian populations. *Genetics*, vol. 16, pp.97-159, 1931.

Notice that 0 and N are absorbing state, and that $\{1, \dots, N-1\}$ is an open communicating class. The quantity of interest in this model is the extinction time of the diversity (that is the entry time of $\{0, N\}$):

$$\tau_N = \inf\{n \geq 0; X_n \in \{0, N\}\},$$

with the convention $\inf \emptyset = \infty$. Using martingale techniques developed in Chapter 4, one can easily prove the following result.

Lemma 3.43. *A.s. the extinction time τ_N is finite and $\mathbb{P}(X_{\tau_N} = N | X_0) = X_0/N$.*

It is interesting to study the mean extinction time $\mathbf{t}_N = (\mathbf{t}_N(i); i \in E_N)$ defined by $\mathbf{t}_N(i) = \mathbb{E}_i[\tau_N]$. We have $\mathbf{t}_N(0) = \mathbf{t}_N(N) = 0$ and for $i \in \{1, \dots, N-1\}$:

$$\begin{aligned}
\mathbf{t}_N(i) &= \sum_{j \in E_N} \mathbb{E}_i[\tau_N \mathbf{1}_{\{X_1 = j\}}] \\
&= 1 + \sum_{j \in E_N} \mathbb{E}_i \left[\inf\{n \geq 0;\ X_{n+1} \in \{0, N\}\} \, \mathbf{1}_{\{X_1 = j\}} \right] \\
&= 1 + \sum_{j \in E_N} \mathbb{E}_j[\tau_N] \mathbb{P}_i(X_1 = j) \\
&= 1 + P_N \mathbf{t}_N(i),
\end{aligned}$$

where we used the Markov property at time 1 for the third equality. As 0 and N are absorbing state, we have $\mathbf{t}_N(i) = P\mathbf{t}_N(i) = 0$ for $i \in \{0, N\}$. Let e_0 (resp. e_N) denote the element of $\mathbb{R}^{N+1}$ with all entries equal to 0 but for the first (resp. last) which is equal to 1, and $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^{N+1}$. We have:

$$\mathbf{t}_N = P_N \mathbf{t}_N + \mathbf{1} - e_0 - e_N.$$

So to compute $\mathbf{t}_N$, one has to solve a linear system. For large N, we have the following result[9] for $x \in [0, 1]$:

$$\frac{1}{N} \mathbb{E}_{\lfloor Nx \rfloor}[\tau_N] \xrightarrow[n \to \infty]{} -2 \left(x \log(x) + (1-x) \log(1-x) \right).$$

where $\lfloor z \rfloor$ is the integer part of $z \in \mathbb{R}$. We give an illustration of this approximation in Figure 3.3.

Ehrenfest urn model

The Ehrenfest[10] model has been introduced in 1907 to describe some "paradoxes" in statistical physics. We consider N particles in two identical containers. A each discrete time, we take one particle at random and move it to the other container. Let X_n denote the number of particles in the first container at time n, X_0 being the initial configuration. The sequence $X = (X_n, n \in \mathbb{N})$ represents the evolution of the system. The equilibrium states should concentrate about half of the particles in each container. In this model one container being empty is possible, but almost unobserved. We shall explain this situation using results on

[9] W. J. Ewens. *Mathematical population genetics.* Springer-Verlag, second edition, 2004.

[10] T. Ehrenfest and P. Ehrenfest. *The conceptual foundations of the statistical approach in mechanics.* Cornell Univ. Press, 1959.

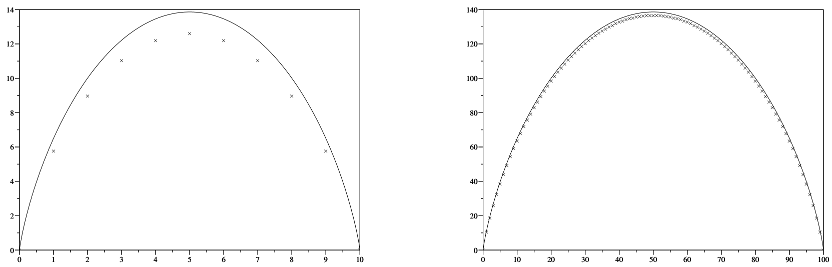

Figure 3.3: Mean extinction time of the diversity ($k \mapsto \mathbb{E}_k[\tau_N]$) and its continuous limit, $Nx \mapsto -2N\,(x\log(x) + (1-x)\log(1-x))$, for $N = 10$ (left) and $N = 100$ (right).

Markov chains[11]. By construction, as all the particles play the same role, the process X is a Markov chain on $E = [\![0, N]\!]$ with transition matrix P given by $P(k, \ell) = 0$ if $|k - \ell| \neq 1$, $P(k, k+1) = (N-k)/N$ and $P(k, k-1) = k/N$ for $k, \ell \in E$. We deduce the Markov chain X is irreducible. Notice that X is reversible with respect to the binomial distribution $\pi_N = (\pi_N(k), k \in E)$, where $\pi_N(k) = 2^{-N}\binom{N}{k}$ for $k \in E$. To see this, it is enough to check that $\pi_N(k)P(k, k+1) = \pi_N(k+1)P(k+1, k)$ for all $k \in [\![0, N-1]\!]$. For $k \in [\![0, N-1]\!]$, we have:

$$\pi_N(k)P(k, k+1) = 2^{-N}\binom{N}{k}\frac{N-k}{N} = 2^{-N}\binom{N}{k+1}\frac{k+1}{N} = \pi_N(k+1)P(k+1, k).$$

According to Lemma 3.18 and Theorem 3.32, we deduce that π_N is the unique invariant probability measure of X. Let $a > 0$ and define the interval $I_{a,N} = [(N \pm a\sqrt{N})/2]$. We also get that the empirical mean time $n^{-1}\sum_{k=1}^n \mathbf{1}_{\{X_k \in I_{a,N}\}}$ spent by the system in the interval $I_{a,N}$ converges a.s., as n goes to infinity, towards $\pi_N\,(I_{a,N})$. Thanks to the CLT, we have that $\pi_N\,(I_{a,N})$ converges, as N goes to infinity, towards $\mathbb{P}(G \in [-a, a])$ where $G \sim \mathcal{N}(0, 1)$ is a standard Gaussian random variable. For a larger than some units (say 2 or 3), this latter probability is close to 1. This implies that it is unlikely to observe values away from $N/2$ by some units time $\sqrt{N}$. Using large deviation theory for the Bernoulli distribution with parameter $1/2$, we get that for $\varepsilon \in (0, 1)$:

$$\frac{2}{N}\log(\pi_N([0, N(1-\varepsilon)/2])) \xrightarrow[N\to\infty]{} -(1+\varepsilon)\log(1+\varepsilon) - (1-\varepsilon)\log(1-\varepsilon).$$

Thus the probability to observe the values from the $N/2$ further by some small units time N decrease exponentially fast towards 0 as N goes to infinity.

For $k, \ell \in E$, let $\mathbf{t}_{k,\ell} = \mathbb{E}_k[T^\ell]$ be the mean of the return time to ℓ starting from k. Set $N_0 = \lfloor N/2 \rfloor$. Using (3.14) and Stirling formula, we get:

$$\mathbf{t}_{N_0,N_0} = \frac{1}{\pi_N(N_0)} \sim \sqrt{\pi N/2} \quad \text{and} \quad \mathbf{t}_{0,0} = \frac{1}{\pi_N(0)} = 2^{-N}.$$

Notice that $\mathbf{t}_{0,0}$ and $\mathbf{t}_{N_0,N_0}$ are not of the same order.

[11] S. Karlin and J. McGregor. Ehrenfest urn models. *J. Appl. Probab*, vol. 2, pp. 352-376, 1965

We are now interested in the mean of the return time from 0 to N_0 and from 0 to N_0. Let $\ell \geq 2$. By decomposing with respect to X_1, we have $\mathbf{t}_{\ell-1,\ell} = 1 + ((\ell-1)/N)\mathbf{t}_{\ell-2,\ell}$ and for $k \in [\![0, \ell-2]\!]$:

$$\mathbf{t}_{k,\ell} = 1 + \frac{k}{N}\mathbf{t}_{k-1,\ell} + \frac{N-k}{N}\mathbf{t}_{k+1,\ell}.$$

Then, using some lengthy computations, we get by recurrence that for $0 \leq k < \ell \leq N$:

$$\mathbf{t}_{k,\ell} = \frac{N}{2}\int_0^1 \frac{du}{u}\,(1-u)^{N-\ell}(1+u)^k\left[(1+u)^{\ell-k} - (1-u)^{\ell-k}\right].$$

We then deduce that:

$$\mathbf{t}_{0,N_0} \sim \frac{N}{4}\log(N) \quad \text{and} \quad \mathbf{t}_{N_0,0} \sim 2^N.$$

This is another indication that one sees mostly the process around N_0 than around 0. Notice the mean time to reach an equilibrium starting from 0 is about $N\log(N)/4$. In fact, one can show a cut-off phenomenon[12]: starting from any initial distribution, one needs at most about $N\log(N)/4$ steps to be close to the invariant measure.

Queuing and stock models

Queuing theory goes back to A. Erlang[13] in 1909 whose work was motivated by telephone exchanges. Since then, this domain has known a huge amount of work. We shall consider a toy example in discrete time. We consider Y_n the size of the queue at the end of the service of the n-th client, with initial state Y_0. We have $Y_{n+1} = (Y_n - 1 + V_{n+1})^+$, where V_{n+1} is the number of clients who arrived during the service of the n-th client. The random variables $(V_n, n \in \mathbb{N}^*)$ are assumed to be independent with the same distribution and independent of Y_0, so that $(Y_n, n \in \mathbb{N})$ is a Markov chain on $\mathbb{N}$. More generally, we can consider the Markov chain $X = (X_n, n \in \mathbb{N})$ on $\mathbb{N}$ defined as a stochastic dynamical system, for $n \in \mathbb{N}$:

$$X_{n+1} = (X_n + U_{n+1})^+, \tag{3.31}$$

where the innovation $(U_n, n \in \mathbb{N}^*)$ is a sequence of $\mathbb{Z}$-valued independent random variables with the same distribution and independent of X_0. Notice this Markov chains is also a model for the evolution of a stock, with U_n being the delivery minus the consummation at time n.

The next lemma gives some criterion for the transience or recurrence for X.

Lemma. *Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain defined by (3.31). We assume that $U_1 \in L^1$, $\mathbb{P}(U_1 > 0) > 0$, $\mathbb{P}(U_1 < 0) > 0$ and X is irreducible.*

1. *If $\mathbb{E}[U_1] > 0$, then X is transient.*
2. *If $\mathbb{E}[U_1] = 0$ and $U_1 \in L^2$, then X is null recurrent.*
3. *If $\mathbb{E}[U_1] < 0$ and U_1 has exponential moments, then X is positive recurrent.*

[12] G.-Y. Chen and L. Saloff-Coste. The L^2-cutoff for reversible Markov processes. *J. Funct. Analysis*, vol. 258, pp. 2246-2315, 2010.

[13] A. Erlang. The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik B*, vol. 20, pp. 33-39, 1909.

Bibliography

[1] X. Chen. *Limit theorems for functionals of ergodic Markov chains with general state space.* Mem. Amer. Math. Soc., 1999.

[2] K. L. Chung. *Markov chains with stationary transition probabilities.* Second edition. Die Grundlehren der mathematischen Wissenschaften, Band 104. Springer-Verlag New York, 1967.

[3] R. Douc, E. Moulines, P. Priouret, and P. Soulier. *Markov chains.* Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2018.

[4] R. Durrett. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge, fourth edition, 2010.

[5] W. Feller. *An introduction to probability theory and its applications. Vol. I.* Third edition. John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[6] G. F. Lawler and V. Limic. *Random walk: a modern introduction*, volume 123 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2010.

[7] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.

[8] F. Spitzer. *Principles of random walk.* Springer-Verlag, second edition, 1976.

Chapter 4

Martingales

In all this chapter, we consider $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space and $\mathbb{F} = (\mathcal{F}_n,\, n \in \mathbb{N})$ a filtration. We also set $\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$. We say a process $H = (H_n, n \in \mathbb{N})$ is i) $\mathbb{F}$-adapted if H_n is $\mathcal{F}_n$-measurable for all $n \in \mathbb{N}$; ii) integrable if H_n is integrable for all $n \in \mathbb{N}$; iii) bounded if $\sup_{n\in\mathbb{N}} |H_n|$ is a.s. bounded. Those definitions are extended in an obvious way to processes indexed by $\overline{\mathbb{N}}$ instead of $\mathbb{N}$. We say a process $H = (H_n, n \in \mathbb{N}^*)$ is $\mathbb{F}$-predictable if H_n is $\mathcal{F}_{n-1}$-measurable for all $n \in \mathbb{N}^*$.

In Section 4.1, we define random times called stopping times and their associated σ-field. That allows to extend the Markov property of Markov chains to the stopping times, which is the so-called strong Markov property. Section 4.2 is devoted to the definition and first properties of the martingales (and super-martingales and sub-martingales). Martingales are a powerful tool to study processes in particular because of the maximal inequalities, see Section 4.3, and the convergence results, see Sections 4.4 and 4.5.

The presentation of this chapter follows closely [1], see also [2] for numerous applications.

4.1 Stopping times

Stopping times are random times which play an important role in Markov process theory and martingale theory.

Definition 4.1. *An $\overline{\mathbb{N}}$-valued random variable τ is an $\mathbb{F}$-stopping time if $\{\tau \leq n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.*

From the definition above, notice that if τ is an $\mathbb{F}$-stopping time, then $\{\tau = \infty\} = \bigcap_{n\in\mathbb{N}}\{\tau \leq n\}^c$ belongs to $\mathcal{F}_\infty$.

When there is no ambiguity on the filtration $\mathbb{F}$, we shall write stopping time instead of $\mathbb{F}$-stopping time. It is clear that the integers are stopping time.

Example 4.2. For the simple random walk $X = (X_n, n \in \mathbb{N})$, see Example 3.4, and $\mathbb{F}$ the natural filtration of X, it is easy to check that the return time to 0, $T^0 = \inf\{n \geq 1;\, X_n = 0\}$, with the convention that $\inf \emptyset = +\infty$, is a stopping time. It is also easy to check that $T^0 - 1$ is not a stopping time. $\triangle$

In the next lemma, we give equivalent characterization of stopping times.

Lemma 4.3. *Let τ be a $\overline{\mathbb{N}}$-valued random variable.*

(i) τ is a stopping time if and only if $\{\tau > n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.

(ii) τ is a stopping time if and only if $\{\tau = n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.

Proof. Use that $\{\tau > n\} = \{\tau \leq n\}^c$ to get (i). Use that $\{\tau = n\} = \{\tau \leq n\} \bigcap \{\tau \leq n-1\}^c$ and that $\{\tau \leq n\} = \bigcup_{k=0}^{n} \{\tau = k\}$ to get (ii). □

We give in the following proposition some properties of stopping times.

Proposition 4.4. *Let $(\tau_n, n \in \mathbb{N})$ be a sequence of stopping times. The random variables $\sup_{n\in\mathbb{N}} \tau_n$, $\inf_{n\in\mathbb{N}} \tau_n$, $\limsup_{n\to\infty} \tau_n$ and $\liminf_{n\to\infty} \tau_n$ are stopping times.*

Proof. We have that $\{\sup_{k\in\mathbb{N}} \tau_k \leq n\} = \bigcap_{k\in\mathbb{N}} \{\tau_k \leq n\}$ belongs to $\mathcal{F}_n$ for all $n \in \mathbb{N}$ as τ_k are stopping time for $k \in \mathbb{N}$. This proves that $\sup_{k\in\mathbb{N}} \tau_k$ is a stopping time. Similarly, use that $\{\inf_{k\in\mathbb{N}} \tau_k \leq n\} = \bigcup_{k\in\mathbb{N}} \{\tau_k \leq n\}$ to deduce that $\inf_{k\in\mathbb{N}} \tau_k$ is a stopping time.

Since stopping time are $\overline{\mathbb{N}}$-valued random variables, we get that $\{\limsup_{k\to\infty} \tau_k \leq n\} = \bigcup_{m\in\mathbb{N}} \bigcap_{k\geq m} \{\tau_k \leq n\}$ for $n \in \mathbb{N}$. This last event belongs to $\mathcal{F}_n$ as τ_k are stopping times for $k \in \mathbb{N}$. We deduce that $\limsup_{k\to\infty} \tau_k$ is a stopping time. Similarly, use that $\{\liminf_{k\to\infty} \tau_k \leq n\} = \bigcap_{m\in\mathbb{N}} \bigcup_{k\geq m} \{\tau_k \leq n\}$ for $n \in \mathbb{N}$ to deduce that $\liminf_{k\to\infty} \tau_k$ is a stopping time. □

It is left to the reader to check that the σ-field $\mathcal{F}_\tau$ in the next definition is indeed a σ-field and a subset of $\mathcal{F}_\infty$.

Definition 4.5. *Let τ be a $\mathbb{F}$-stopping time. The σ-field $\mathcal{F}_\tau$ of the events which are prior to a stopping time τ is defined by:*

$$\mathcal{F}_\tau = \{B \in \mathcal{F}_\infty;\, B \cap \{\tau = n\} \in \mathcal{F}_n \quad \text{for all} \quad n \in \mathbb{N}\}.$$

Clearly, we have that τ is $\mathcal{F}_\tau$-measurable.

Remark 4.6. Consider $X = (X_n, n \in \mathbb{N})$ a Markov chain on a discrete state space E with its natural filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$. Recall the return time to $x \in E$ defined by $T^x = \inf\{n \geq 1;\, X_n = x\}$ and the excursion $Y_1 = (T^x, X_0, \ldots, X_{T^x})$ defined in section 3.4.3. It is easy to check that T^x is an $\mathbb{F}$-stopping time and that $\mathcal{F}_{T^x}$ is equal to $\sigma(Y_1)$. Roughly speaking the σ-field $\mathcal{F}_{T^x}$ contains all the information on X prior to T^x. ◊

We give an elementary characterization of the $\mathcal{F}_\tau$-measurable random variables.

Lemma 4.7. *Let Y be a $\mathcal{F}_\infty$-measurable real-valued random variable and τ a stopping time.*

(i) The random variable Y is $\mathcal{F}_\tau$-measurable if and only if $Y\mathbf{1}_{\{\tau=n\}}$ is $\mathcal{F}_n$-measurable for all $n \in \mathbb{N}$.

(ii) If $\mathbb{E}[Y]$ is well defined, then we have that a.s.:

$$\mathbb{E}[Y|\,\mathcal{F}_\tau] = \sum_{n\in\overline{\mathbb{N}}} \mathbf{1}_{\{\tau=n\}}\, \mathbb{E}[Y|\,\mathcal{F}_n]. \tag{4.1}$$

Proof. We prove (i). Set $Y_n = Y\mathbf{1}_{\{\tau=n\}}$. We first assume that Y is $\mathcal{F}_\tau$-measurable and we prove that Y_n is $\mathcal{F}_n$-measurable for all $n \in \mathbb{N}$. If $Y = \mathbf{1}_B$ with $B \in \mathcal{F}_\infty$, we clearly get that

Y_n is $\mathcal{F}_n$-measurable for all $n \in \mathbb{N}$ by definition of $\mathcal{F}_\tau$. It is then easy to extend this result to any $\mathcal{F}_\infty$-measurable random variable which takes finitely different values in $\overline{\mathbb{R}}$, and then to extend to any $\mathcal{F}_\infty$-measurable real-valued random variable Y by considering any sequence of random variables $(Y^k, k \in \mathbb{N})$ which converges to Y and such that Y^k is $\mathcal{F}_\tau$-measurable and takes finitely many values in $\overline{\mathbb{R}}$ (for example take $Y^k = 2^{-k}\lfloor 2^k Y \rfloor \mathbf{1}_{\{|Y|\leq k\}} + Y\mathbf{1}_{\{|Y|=+\infty\}}$).

We now assume that Y_n is $\mathcal{F}_n$-measurable for all $n \in \mathbb{N}$ and we prove that Y is $\mathcal{F}_\tau$-measurable. Let $A \in \mathcal{B}(\overline{\mathbb{R}})$ and set $B = Y^{-1}(A)$. Notice that B belongs to $\mathcal{F}_\infty$ as Y is $\mathcal{F}_\infty$-measurable. First assume that $0 \not\in A$. In this case, we get $B \cap \{\tau = n\} = Y_n^{-1}(A)$ and thus $B \cap \{\tau = n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$. This gives $B \in \mathcal{F}_\tau$. If $0 \in A$, then uses that $B = Y^{-1}(A) = \left(Y^{-1}(A^c)\right)^c$ to also get that $B \in \mathcal{F}_\tau$. This implies that Y is $\mathcal{F}_\tau$-measurable. This ends the proof of (i).

We now prove (ii). Assume first that $Y \geq 0$ and set:

$$Z = \sum_{n\in\overline{\mathbb{N}}} \mathbb{E}[Y|\,\mathcal{F}_n]\mathbf{1}_{\{\tau=n\}}.$$

Since Y is $\mathcal{F}_\infty$-measurable, we also get that $Y\mathbf{1}_{\{\tau=\infty\}}$ is $\mathcal{F}_\infty$-measurable. Thus, we deduce from (i) that Z is $\mathcal{F}_\tau$-measurable. For $B \in \mathcal{F}_\tau$, we have:

$$\mathbb{E}\left[Z\mathbf{1}_B\right] = \sum_{n\in\overline{\mathbb{N}}} \mathbb{E}\left[\mathbb{E}[Y|\,\mathcal{F}_n]\mathbf{1}_{\{\tau=n\}\cap B}\right] = \sum_{n\in\overline{\mathbb{N}}} \mathbb{E}\left[Y\mathbf{1}_{\{\tau=n\}\cap B}\right] = \mathbb{E}\left[Y\mathbf{1}_B\right],$$

where we used monotone convergence for the first equality, the fact that $\{\tau = n\} \cap B$ belongs to $\mathcal{F}_n$ and (2.1) for the second and monotone convergence for the last. As Z is $\mathcal{F}_\tau$-measurable, we deduce from (2.1) that a.s. $Z = \mathbb{E}[Y|\,\mathcal{F}_\tau]$.

Then consider Y a $\mathcal{F}_\infty$-measurable real-valued random variable. Subtracting (4.1) with Y replaced by Y^- to (4.1) with Y replaced by Y^+ gives that (4.1) holds as soon as $\mathbb{E}[Y]$ is well defined. □

Definition 4.8. *Let $X = (X_n, n \in \overline{\mathbb{N}})$ be a $\mathbb{F}$-adapted process and τ a $\mathbb{F}$-stopping time. The random variable X_τ is defined by:*

$$X_\tau = \sum_{n\in\overline{\mathbb{N}}} X_n\, \mathbf{1}_{\{\tau=n\}}.$$

This definition is extended in an obvious way when τ is an a.s. finite stopping time and X a process indexed on $\mathbb{N}$ instead of $\overline{\mathbb{N}}$. By construction the random variable X_τ from Definition 4.8 is $\mathcal{F}_\tau$-measurable. We can now give an extension of the Markov property, see Definition 3.2, when considering random times. Compare the next proposition with Corollary 3.12.

Proposition 4.9 (Strong Markov property). *Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain with respect to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$, taking values in a discrete state space E and with transition matrix P. Let τ be a $\mathbb{F}$-stopping time a.s. finite and define a.s. the shifted process $\tilde{X} = (\tilde{X}_k = X_{\tau+k}, k \in \mathbb{N})$. Conditionally on X_τ, we have that $\mathcal{F}_\tau$ and $\tilde{X}$ are independent and that $\tilde{X}$ is a Markov chain with transition matrix P, which means that a.s. for all $k \in \mathbb{N}$, all $x_0, \ldots, x_k \in E$:*

$$\begin{aligned}\mathbb{P}(\tilde{X}_0 = x_0, \dots, \tilde{X}_k = x_k \mid \mathcal{F}_\tau) &= \mathbb{P}(\tilde{X}_0 = x_0, \dots, \tilde{X}_k = x_k \mid X_\tau) \\ &= \mathbf{1}_{\{X_\tau = x_0\}} \prod_{j=1}^{k} P(x_{j-1}, x_j). \end{aligned} \tag{4.2}$$

Proof. Let $B \in \mathcal{F}_\tau$, $k \in \mathbb{N}$ and $x_0, \dots, x_k \in E$. We first compute:

$$I_n = \mathbb{E}\left[\mathbf{1}_B \mathbf{1}_{\{\tilde{X}_0 = x_0, \dots, \tilde{X}_k = x_k\}} \mid \mathcal{F}_n\right] \mathbf{1}_{\{\tau = n\}}.$$

We have, using that $B \cap \{\tau = n\} \in \mathcal{F}_n$ and the Markov property at time n:

$$I_n = \mathbb{E}\left[\mathbf{1}_{B \cap \{\tau = n\}} \mathbf{1}_{\{X_n = x_0, \dots, X_{n+k} = x_k\}} | \mathcal{F}_n\right] = \mathbf{1}_{B \cap \{\tau = n\}} H(X_n),$$

where for $x \in E$:

$$H(x) = \mathbb{P}_x(X_0 = x_0, \dots, X_k = x_k) = \mathbf{1}_{\{x = x_0\}} \prod_{i=0}^{k-1} P(x_i, x_{i+1}). \tag{4.3}$$

We get from Lemma 4.7 and Definition 4.8 that $\mathbb{E}\left[\mathbf{1}_B \mathbf{1}_{\{\tilde{X}_0 = x_0, \dots, \tilde{X}_k = x_k\}} | \mathcal{F}_\tau\right] = \mathbf{1}_B H(X_\tau)$. Then, taking the expectation conditionally on X_τ, we deduce that:

$$\mathbb{E}\left[\mathbf{1}_B \mathbf{1}_{\{\tilde{X}_0 = x_0, \dots, \tilde{X}_k = x_k\}} | X_\tau\right] = \mathbb{P}(B | X_\tau) H(X_\tau). \tag{4.4}$$

Since this holds for all $B \in \mathcal{F}_\tau$, $k \in \mathbb{N}$ and $x_0, \dots, x_k \in E$, we get that conditionally on X_τ, $\mathcal{F}_\tau$ and $\tilde{X}$ are independent. Take $B = \Omega$ in (4.4) to get $\mathbb{P}(\tilde{X}_0 = x_0, \dots, \tilde{X}_k = x_k | X_\tau) = H(X_\tau)$ and use the definition (4.3) of H to conclude that $\tilde{X}$ is conditionally on X_τ a Markov chain with transition matrix P. Take $B = \Omega$ in the previous computations to get (4.2). □

Using the strong Markov property, it is immediate to get that the excursions of a recurrent irreducible Markov chain out of a given state are independent and, but for the first one, with the same distribution, see the key Lemma 3.36.

We end this section with the following lemma.

Lemma 4.10. *Let τ and τ' be two stopping times.*

(i) The events $\{\tau < \tau'\}$, $\{\tau = \tau'\}$ and $\{\tau \geq \tau'\}$ belongs to $\mathcal{F}_\tau$ and $\mathcal{F}_{\tau'}$.

(ii) If $B \in \mathcal{F}_\tau$, then we have that $B \cap \{\tau \leq \tau'\}$ belongs to $\mathcal{F}_{\tau'}$.

(iii) If $\tau \leq \tau'$, then we have $\mathcal{F}_\tau \subset \mathcal{F}_{\tau'}$.

Proof. We have $\{\tau < \tau'\} \cap \{\tau = n\} = \{\tau = n\} \cap \{\tau' > n\}$ which belongs to $\mathcal{F}_n$ as $\{\tau = n\}$ and $\{\tau' > n\}$ belong already to $\mathcal{F}_n$. Since this holds for all $n \in \mathbb{N}$, we deduce that $\{\tau < \tau'\} \in \mathcal{F}_\tau$. The other results of property (i) can be proved similarly.

Let $B \in \mathcal{F}_\tau$. This implies that $B \cap \{\tau \leq n\}$ belongs to $\mathcal{F}_n$. We deduce that $B \cap \{\tau \leq \tau'\} \cap \{\tau' = n\} = B \cap \{\tau \leq n\} \cap \{\tau' = n\}$ belongs to $\mathcal{F}_n$. Since this holds for $n \in \mathbb{N}$, we get that $B \cap \{\tau \leq \tau'\} \in \mathcal{F}_{\tau'}$. This gives property (ii).

Property (iii) is a direct consequence of property (ii) as $\{\tau \leq \tau'\} = \Omega$. □

Remark 4.11. In some cases, it can be convenient to assume that $\mathcal{F}_0$ contains at least all the $\mathbb{P}$-null sets. Under this condition, if a $\overline{\mathbb{N}}$-valued random variable is a.s. constant, then it is a stopping time. And, more importantly, under this condition, property (iii) of Lemma 4.10 holds if a.s. $\tau \leq \tau'$. ◊

4.2 Martingales and the optional stopping theorem

Definition 4.12. *A real-valued process $M = (M_n, n \in \mathbb{N})$ is called an $\mathbb{F}$-martingale if it is $\mathbb{F}$-adapted, integrable and for all $n \in \mathbb{N}$ a.s.:*

$$\mathbb{E}[M_{n+1} | \mathcal{F}_n] = M_n. \tag{4.5}$$

If (4.5) is replaced by $\mathbb{E}[M_{n+1} | \mathcal{F}_n] \geq M_n$, then M is called an $\mathbb{F}$-sub-martingale.
If (4.5) is replaced by $\mathbb{E}[M_{n+1} | \mathcal{F}_n] \leq M_n$, then M is called an $\mathbb{F}$-super-martingale.

Quoting [1]: "A super-martingale is by definition a sequence of random variables which *decrease in conditional mean*. For a sequence $(M_n, n \in \mathbb{N})$ of non-negative random variables denoting the sequence of values of the fortune of a gambler, the super-martingale condition express the property that at each play the game is unfavorable to the player in conditional mean. On the other hand, a martingale *remains constant in conditional mean* and, for the gambler, corresponds to a game which is on the average fair".

When there is no possible confusion, we omit the filtration; for example we write martingale for $\mathbb{F}$-martingale. See [1], for a theory of super-martingales which are non-negative processes instead of integrable.

Example 4.13 (Random walk in $\mathbb{R}$). Let $(U_n, n \in \mathbb{N}^*)$ be independent integrable real-valued random variables with the same distribution. We consider the random walk $X = (X_n, n \in \mathbb{N})$ defined by $X_0 = 0$ and $X_n = \sum_{k=1}^n U_k$ for $n \in \mathbb{N}^*$. Let $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$ be the natural filtration of the process X. If $\mathbb{E}[U_1] \leq 0$, then X is a super-martingale. If $\mathbb{E}[U_1] = 0$, then X is a martingale.

Assume that U_1 has all its exponential moments (that is $\mathbb{E}[\exp(\lambda U_1)] < +\infty$ for all $\lambda \in \mathbb{R}$), and define $\varphi(\lambda) = \log\left(\mathbb{E}\left[e^{\lambda U_1}\right]\right)$ for $\lambda \in \mathbb{R}$. Let $\lambda \in \mathbb{R}$ be fixed. It is elementary to check that $M^\lambda = (M_n^\lambda, n \in \mathbb{N})$ defined by, for $n \in \mathbb{N}$:

$$M_n^\lambda = e^{\lambda X_n - n\varphi(\lambda)}$$

is a positive martingale. This martingale is called the exponential martingale associated to the random walk X. △

Let $X = (X_n, n \in \mathbb{N})$ and $(H_n, n \in \mathbb{N}^*)$ be two sequences of real-valued random variables which are a.s. finite. We define the discrete stochastic integral of H with respect to X by the process $H \cdot X = (H \cdot X_n, n \in \mathbb{N})$ with $H \cdot X_0 = 0$ and for all $n \in \mathbb{N}^*$:

$$H \cdot X_n = \sum_{k=1}^n H_k \Delta X_k = H \cdot X_{n-1} + H_n \Delta X_n \quad \text{where} \quad \Delta X_k = X_k - X_{k-1}.$$

Lemma 4.14. *Let M be a martingale (resp. super-martingale) and H a bounded real-valued predictable process (resp. and non-negative). Then, the discrete stochastic integral $H \cdot M$ is a martingale (resp. super-martingale).*

Proof. With $M = (M_n, n \in \mathbb{N})$, $H = (H_n, n \in \mathbb{N}^*)$ and $H \cdot M = (H \cdot M_n, n \in \mathbb{N})$, we get that the process $H \cdot M$ is adapted and integrable. Assume that M is a martingale. We have for $n \in \mathbb{N}^*$ a.s.:

$$\mathbb{E}[H \cdot M_n | \mathcal{F}_{n-1}] = H \cdot M_{n-1} + H_n \left(\mathbb{E}[M_n | \mathcal{F}_{n-1}] - M_{n-1}\right) = 0.$$

The conclusion is then straightforward. The case M super-martingale and H non-negative is proved similarly. □

Remark 4.15 (Doob decomposition of a super-martingale). Let $M = (M_n, n \in \mathbb{N})$ be a super-martingale. We set $N_0 = M_0$, $A_0 = 0$ and for $n \in \mathbb{N}$:

$$N_{n+1} = N_n + M_{n+1} - \mathbb{E}[M_{n+1} | \mathcal{F}_n] \quad \text{and} \quad A_{n+1} = A_n + M_n - \mathbb{E}[M_{n+1} | \mathcal{F}_n].$$

By construction the process $N = (N_n \in \mathbb{N})$ is adapted and the process $A = (A_n, n \in \mathbb{N}^*)$ is predictable. It easy to check that N is a martingale and A is non-decreasing and thus non-negative. The decomposition of the super-martingale M as $M_n = N_n - A_n$ with N a martingale and A predictable non-decreasing is called the Doob decomposition of M. ♢

Using Jensen inequality (2.4), we easily derive the next corollary.

Corollary 4.16. *Let $M = (M_n, n \in \mathbb{N})$ be a real-valued $\mathbb{F}$-martingale. Let φ be a $\mathbb{R}$-valued convex function defined on $\mathbb{R}$. Assume that $\varphi(M_n)$ is integrable for all $n \in \mathbb{N}$. Then, the process $(\varphi(M_n), n \in \mathbb{N})$ is a sub-martingale.*

For $x, x' \in \overline{\mathbb{R}}$, we recall that we write $x \wedge x'$ for $\min(x, x')$.

Lemma 4.17. *Let τ be a stopping time and $M = (M_n, n \in \mathbb{N})$ a martingale (resp. super-martingale, sub-martingale). Then, the process $M^\tau = (M_{\tau\wedge n}, n \in \mathbb{N})$ is a martingale (resp. super-martingale, sub-martingale).*

We provide two proofs of this important lemma. The shorter one relies on the stochastic integral. The other one can be generalized when the integrability condition is weakened; it will inspire some computations in Chapter 5.

First proof. Let M be a martingale. The process $H = (H_n, n \in \mathbb{N}^*)$ defined by $H_n = \mathbf{1}_{\{\tau \geq n\}}$ is predictable bounded and non-negative. The discrete stochastic integral of H with respect to M is given by $H \cdot M = (H \cdot M_n, n \in \mathbb{N})$ with:

$$H \cdot M_n = \sum_{k=1}^{n} \mathbf{1}_{\{\tau \geq k\}} (M_k - M_{k-1}) = M_{\tau\wedge n} - M_0.$$

As $H \cdot M$ is a martingale according to Lemma 4.14, we deduce that M^τ is a martingale. The proofs are similar in the super-martingale and sub-martingale cases. □

Second proof. Let M be a martingale. For $n \in \mathbb{N}$, we have:

$$M_{\tau\wedge n} = \sum_{k=0}^{n-1} M_k \mathbf{1}_{\{\tau=k\}} + M_n \mathbf{1}_{\{\tau>n-1\}}.$$

This implies that $M_{\tau\wedge n}$ is integrable and $\mathcal{F}_n$-measurable. For $n \geq 1$, we have:

$$\begin{aligned}\mathbb{E}\left[M_{\tau\wedge n}\middle|\,\mathcal{F}_{n-1}\right] &= \sum_{k=0}^{n-1} M_k \mathbf{1}_{\{\tau=k\}} + \mathbb{E}\left[M_n\middle|\,\mathcal{F}_{n-1}\right]\mathbf{1}_{\{\tau>n-1\}}\\ &= \sum_{k=0}^{n-1} M_k \mathbf{1}_{\{\tau=k\}} + M_{n-1}\mathbf{1}_{\{\tau>n-1\}}\\ &= M_{\tau\wedge(n-1)}.\end{aligned}$$

This implies that M^τ is a martingale. The proofs are similar in the super-martingale and sub-martingale cases. □

We recall that a stopping time τ is bounded if $\mathbb{P}(\tau \leq n_0) = 1$ for some $n_0 \in \mathbb{N}$. If ν is a stopping time, recall the σ-field $\mathcal{F}_\nu$ of the events prior to ν.

Theorem 4.18 (Optional stopping theorem)**.** *Let $M = (M_n, n \in \mathbb{N})$ be a martingale. Let τ and ν be bounded stopping times such that $\nu \leq \tau$. We have a.s.:*

$$\mathbb{E}\left[M_\tau\middle|\,\mathcal{F}_\nu\right] = M_\nu. \tag{4.6}$$

When M is a super-martingale (resp. sub-martingale) the equality in (4.6) is replaced by the inequality $\mathbb{E}\left[M_\tau\middle|\,\mathcal{F}_\nu\right] \leq M_\nu$ (resp. $\mathbb{E}\left[M_\tau\middle|\,\mathcal{F}_\nu\right] \geq M_\nu$).

In particular, if M is a martingale and τ a bounded stopping time, taking the expectation in (4.6) with $\nu = 0$, we get $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$. See Proposition 4.26 for an extension of the optional stopping theorem to unbounded stopping times for closed martingale.

Proof. Let $n_0 \in \mathbb{N}$ be such that a.s. $\tau \leq n_0$. We have according to Lemma 4.7 that a.s.:

$$\mathbb{E}\left[M_\tau\middle|\,\mathcal{F}_\nu\right] = \sum_{n=0}^{n_0} \mathbf{1}_{\{\nu=n\}}\mathbb{E}\left[M_\tau\middle|\,\mathcal{F}_n\right].$$

Since M^τ is a martingale and $\tau \wedge n_0 = \tau$ a.s., we have for $n \leq n_0$ that a.s. $\mathbb{E}\left[M_\tau\middle|\,\mathcal{F}_n\right] = \mathbb{E}\left[M_{\tau\wedge n_0}\middle|\,\mathcal{F}_n\right] = M_{\tau\wedge n}$. Since $\nu \leq \tau$, we deduce that:

$$\mathbb{E}\left[M_\tau\middle|\,\mathcal{F}_\nu\right] = \sum_{n=0}^{n_0} \mathbf{1}_{\{\nu=n\}} M_{\tau\wedge n} = M_\nu.$$

The proofs are similar for super-martingales and sub-martingales. □

See Exercise 8.27 for an application of the martingale theory to simple random walk.

4.3 Maximal inequalities

In this section, we provide inequalities in mean on the path of a martingale using the last value of the path.

Theorem 4.19 (Doob's maximal inequality)**.** *Let $(M_n,\, n \in \mathbb{N})$ be a sub-martingale. Let $a > 0$. Then, we have for $n \in \mathbb{N}$:*

$$a\mathbb{P}\left(\max_{k\in[\![0,n]\!]} M_k \geq a\right) \leq \mathbb{E}\left[M_n \mathbf{1}_{\{\max_{k\in[\![0,n]\!]} M_k \geq a\}}\right] \leq \mathbb{E}\left[M_n^+\right].$$

Proof. Let $n \in \mathbb{N}$. Consider the stopping time $\tau = \inf\{k \in \mathbb{N};\, M_k \geq a\}$, and set $A = \{\max_{k\in[\![0,n]\!]} M_k \geq a\} = \{\tau \leq n\}$. Thanks to the optional stopping theorem, we have $\mathbb{E}[M_n] \geq \mathbb{E}[M_{\tau\wedge n}]$. Since $M_{\tau\wedge n} \geq a\mathbf{1}_A + M_n\mathbf{1}_{A^c}$, we deduce that:

$$\mathbb{E}[M_n] \geq a\mathbb{P}(A) + \mathbb{E}\left[M_n\mathbf{1}_{A^c}\right].$$

This implies that $\mathbb{E}\left[M_n\mathbf{1}_A\right] \geq a\mathbb{P}(A)$. The inequality $\mathbb{E}\left[M_n\mathbf{1}_A\right] \leq \mathbb{E}\left[M_n^+\right]$ is obvious. □

Let $(M_n,\, n \in \mathbb{N})$ be a sequence of real-valued random variables. We define for $n \in \mathbb{N}$:

$$M_n^* = \max_{k\in[\![0,n]\!]} |M_k|. \tag{4.7}$$

We deduce from Corollary 4.16 that if $(M_n,\, n \in \mathbb{N})$ is a martingale, then $(|M_n|,\, n \in \mathbb{N})$ is a sub-martingale and thus, thanks to Theorem 4.19, we have for $a > 0$:

$$a\mathbb{P}\left(M_n^* \geq a\right) \leq \mathbb{E}[|M_n|].$$

Proposition 4.20. *Let $M = (M_n,\, n \in \mathbb{N})$ be a martingale. Assume that $M_n \in L^p$ for some $n \in \mathbb{N}$ and $p > 1$. Then, we have, with $C_p = (p/(p-1))^p$:*

$$\mathbb{E}\left[(M_n^*)^p\right] \leq C_p\mathbb{E}\left[|M_n|^p\right].$$

Proof. We first prove that M_n^* belongs to L^p. We deduce from Corollary 4.16 that $(|M_k|,\, k \in \mathbb{N})$ is a non-negative sub-martingale. We deduce from Jensen inequality that for $0 \leq k \leq n$:

$$\mathbb{E}\left[|M_k|^p\right] \leq \mathbb{E}\left[\mathbb{E}\left[|M_n|\,|\,\mathcal{F}_k\right]^p\right] \leq \mathbb{E}\left[|M_n|^p\right]. \tag{4.8}$$

Since $M_n^* \leq \sum_{k=1}^n |M_k|$, we deduce that M_n^* belongs to L^p.

Thanks to Theorem 4.19 (with M_k replaced by $|M_k|$), we have for all $a > 0$ that $a\mathbb{P}(M_n^* \geq a) \leq \mathbb{E}\left[|M_n|\mathbf{1}_{\{M_n^*\geq a\}}\right]$. Multiplying this inequality by $p(p-1)a^{p-2}$ and integrating over $a > 0$ with respect to the Lebesgue measure gives:

$$\begin{aligned}(p-1)\mathbb{E}\left[(M_n^*)^p\right] &= p(p-1)\int_{a>0} a^{p-1}\mathbb{P}(M_n^* \geq a)\,\mathrm{d}a\\ &\leq p(p-1)\int_{a>0} a^{p-2}\mathbb{E}\left[|M_n|\mathbf{1}_{\{M_n^*\geq a\}}\right]\,\mathrm{d}a\\ &= p\mathbb{E}\left[|M_n|\,(M_n^*)^{p-1}\right].\end{aligned}$$

Using Hölder inequality, we get that $\mathbb{E}\left[|M_n|\,(M_n^*)^{p-1}\right] \leq \mathbb{E}\left[|M_n|^p\right]^{1/p}\mathbb{E}\left[(M_n^*)^p\right]^{(p-1)/p}$. This implies that $(p-1)\mathbb{E}\left[(M_n^*)^p\right]^{1/p} \leq p\mathbb{E}\left[|M_n|^p\right]^{1/p}$. □

4.4 Convergence of martingales

We now state the main result on convergence of martingales whose proof is given at the end of this section.

Theorem 4.21. *Let $M = (M_n, n \in \mathbb{N})$ be a martingale or a sub-martingale or a super-martingale bounded in L^1, that is $\sup_{n\in\mathbb{N}} \mathbb{E}[|M_n|] < +\infty$. Then, the process M converges a.s. to a limit, say M_∞, which is integrable and:*

$$\liminf_{n\to\infty} \mathbb{E}[|M_n|] \geq \mathbb{E}[|M_\infty|]. \tag{4.9}$$

We give in the next corollary direct consequences which are so often used that they deserve to be stated on their own.

Corollary 4.22. *We have the following results.*

(i) Let $M = (M_n, n \in \mathbb{N})$ be a sub-martingale such that $\sup_{n\in\mathbb{N}} \mathbb{E}[M_n^+] < +\infty$. Then, the process M converges a.s. to a limit, say M_∞, which is integrable and (4.9) holds.

(ii) Let $M = (M_n, n \in \mathbb{N})$ be a non-negative martingale or a non-negative super-martingale. Then the process M converges a.s. to a limit, say M_∞, which is integrable, and we have:

$$\lim_{n\to\infty} \mathbb{E}[M_n] \geq \mathbb{E}[M_\infty]. \tag{4.10}$$

Proof. We first prove property (i). As M is a sub-martingale, we have that $\mathbb{E}[M_0] \leq \mathbb{E}[M_n]$ and thus $\mathbb{E}[|M_n|] \leq 2\mathbb{E}[M_n^+] - \mathbb{E}[M_0]$. We deduce the condition $\sup_{n\in\mathbb{N}} \mathbb{E}[M_n^+] < +\infty$ is equivalent to $\sup_{n\in\mathbb{N}} \mathbb{E}[|M_n|] < +\infty$. Then use Theorem 4.21 to conclude.

Let M be a non-negative super-martingale. Considering property (i) with $-M$, we get the a.s. convergence of M towards a limit say M_∞. Then use Fatou's lemma and that the sequence $(\mathbb{E}[M_n], n \in \mathbb{N})$ is non-increasing to get (4.10). □

Remark 4.23. We state without proof the following extension, see [1]. Let $M = (M_n, n \in \mathbb{N})$ be a non-negative non necessarily integrable super-martingale, that is M is adapted, and a.s., for all $n \in \mathbb{N}$, we have $M_n \geq 0$ and $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \leq M_n$. Then, the process M converges a.s. and the limit, say M_∞, satisfies the inequality $\mathbb{E}\left[M_\infty | \mathcal{F}_n\right] \leq M_n$ a.s. for all $n \in \mathbb{N}$. Furthermore, for all stopping times τ and ν such that $\tau \geq \nu$, we have that a.s. $\mathbb{E}\left[M_\tau | \mathcal{F}_\nu\right] \leq M_\nu$. However, Equality (4.6) does not hold in general for positive non necessarily integrable martingale, that is an adapted process $M = (M_n, n \in \mathbb{N})$ such that, for all $n \in \mathbb{N}$, a.s. $M_n \geq 0$ and $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$. ♢

Proof of Theorem 4.21

This proof can be skipped in a first reading. For $a < b \in \mathbb{R}$ and a sequence $x = (x_n, n \in \mathbb{N})$ of elements of $\overline{\mathbb{R}}$, we define the down-crossing and up-crossing times of $[a, b]$ for the sequence x as $\tau_0(x) = 0$ and for all $k \in \mathbb{N}^*$:

$$\nu_k(x) = \inf\{n \geq \tau_{k-1}(x);\, x_n \leq a\} \quad \text{and} \quad \tau_k(x) = \inf\{n \geq \nu_k(x);\, x_n \geq b\},$$

with the convention that $\inf \emptyset = \infty$. We define the number of up-crossings for the sequence x of the interval $[a, b]$ up to time $n \in \mathbb{N}$ as:

$$\beta_{a,b}(x, n) = \sup\{k \in \mathbb{N};\, \tau_k(x) \leq n\}.$$

We shall also consider the total number of up-crossings given by:

$$\beta_{a,b}(x) = \lim_{n\to\infty} \beta_{a,b}(x,n) = \text{Card}\,(\{k \in \mathbb{N};\, \tau_k(x) < \infty\}) \in \overline{\mathbb{N}}.$$

As $a < b$, we have the following implications:

$$\liminf_{n\to\infty} x_n < a < b < \limsup_{n\to\infty} x_n \Longrightarrow \beta_{a,b}(x) = \infty \Longrightarrow \liminf_{n\to\infty} x_n \leq a < b \leq \limsup_{n\to\infty} x_n.$$

We deduce that the sequence x converges in $\overline{\mathbb{R}}$ if and only if $\beta_{a,b}(x) < \infty$ for all $a < b$ with $a, b \in \mathbb{Q}$. Thus, to prove the convergence of the sequence x, it is enough to give a finite upper bounds of $\beta_{a,b}(x)$. Since $x_{\tau_k(x)} - x_{\nu_k(x)} \geq b - a$ when $\tau_k(x) < \infty$ that is $k \leq \beta_{a,b}(x)$, we deduce that:

$$\sum_{k=1}^{\beta_{a,b}(x,n)} (x_{\tau_k(x)} - x_{\nu_k(x)}) \geq (b-a)\beta_{a,b}(x,n). \tag{4.11}$$

Define $H_\ell(x) = \mathbf{1}_{\bigcup_{k\in\mathbb{N}}\{\nu_k(x)<\ell\leq\tau_k(x)\}}$ for $\ell \in \mathbb{N}^*$. Considering the discrete integral $H(x){\cdot}x_n = \sum_{\ell=1}^n H_\ell(x)\Delta x_\ell$, with $\Delta x_\ell = x_\ell - x_{\ell-1}$, we get:

$$H(x){\cdot}x_n \geq \sum_{k=1}^{\beta_{a,b}(x,n)} (x_{\tau_k(x)} - x_{\nu_k(x)}) - (x_n - a)^- \geq (b-a)\beta_{a,b}(x,n) - (x_n - a)^-, \tag{4.12}$$

where for the first inequality we took into account the fact that n may belongs to an up-crossing from a to b, and we used (4.11) for the second.

Up to replacing M by $-M$, we can assume that M is a super-martingale. We now replace x by the super-martingale M. The random variables $\nu_k(M)$, $\tau_k(M)$, for $k \in \mathbb{N}$, are by construction stopping times. This implies that, for $\ell \in \mathbb{N}^*$, the event $\{\nu_k(M) < \ell \leq \tau_k(M)\}$ belongs to $\mathcal{F}_{\ell-1}$. We deduce that the process $H = (H_\ell(M),\, \ell \in \mathbb{N}^*)$ is adapted bounded and non-negative. Thanks to Lemma 4.14 the discrete stochastic integral $(H(M){\cdot}M_n, n \in \mathbb{N})$ is a super-martingale. Since $H(M){\cdot}M_0 = 0$, we get $\mathbb{E}[H(M){\cdot}M_n] \leq 0$. We deduce from (4.12) that:

$$(b-a)\mathbb{E}[\beta_{a,b}(M,n)] \leq \mathbb{E}[(M_n - a)^-] + \mathbb{E}[H(M){\cdot}M_n] \leq \mathbb{E}[|M_n|] + |a|.$$

Letting n goes to infinity in the previous inequality, we get using $\sup_{n\in\mathbb{N}} \mathbb{E}[|M_n|] < +\infty$ and the monotone convergence theorem that $\mathbb{E}[\beta_{a,b}(M)] < +\infty$. This implies that the event $\bigcap_{a<b;\, a,b\in\mathbb{Q}} \{\beta_{a,b}(M) < \infty\}$ has probability 1, that is the super-martingale M a.s. converges to a real-valued random variable, say M_∞.

Using Fatou's lemma, we get:

$$\mathbb{E}[|M_\infty|] = \mathbb{E}[\lim_{n\to\infty} |M_n|] \leq \liminf_{n\to\infty} \mathbb{E}[|M_n|] \leq \sup_{n\in\mathbb{N}} \mathbb{E}[|M_n|] < +\infty.$$

We deduce that M_∞ is integrable and that (4.9) holds.

4.5 More on convergence of martingales

The fact that (4.10) is an equality or not for martingales plays an important role in the applications, which motivate this section. We refer to Section 7.2.3 for the definition of the uniform integrability and some related results.

Theorem 4.24. *Let $M = (M_n, n \in \mathbb{N})$ be a martingale. The next conditions are equivalent.*

(i) *The martingale M converges a.s. and in L^1 to a limit, say M_∞.*

(ii) *There exists a real-valued integrable random variable Z such that $M_n = \mathbb{E}[Z|\mathcal{F}_n]$ a.s. for all $n \in \mathbb{N}$.*

(iii) *The random variables $(M_n, n \in \mathbb{N})$ are uniformly integrable.*

If any of these conditions hold, we have that a.s. for all $n \in \mathbb{N}$:

$$M_n = \mathbb{E}[M_\infty|\mathcal{F}_n]. \tag{4.13}$$

A martingale satisfying the conditions of Theorem 4.24 is called a closed martingale.

Proof. Assume property (i) holds. Denote M_∞ the limit of M. For $m \geq n$, we have $\mathbb{E}[M_m|\mathcal{F}_n] = M_n$ a.s. and thus using Jensen's inequality:

$$\mathbb{E}\Big[|M_n - \mathbb{E}[M_\infty|\mathcal{F}_n]|\Big] = \mathbb{E}\Big[|\mathbb{E}[M_m - M_\infty|\mathcal{F}_n]|\Big] \leq \mathbb{E}\Big[\mathbb{E}[|M_m - M_\infty||\mathcal{F}_n]\Big] = \mathbb{E}\Big[|M_m - M_\infty|\Big].$$

As M converges to M_∞ in L^1, we deduce that the right-hand side of the previous equation goes to 0 as m goes to infinity. We deduce that $\mathbb{E}\Big[|M_n - \mathbb{E}[M_\infty|\mathcal{F}_n]|\Big] = 0$, which gives (4.13) and that (ii) holds with $Z = M_\infty$.

Using Lemma 7.19, we get that property (ii) implies property (iii).

Assume property (iii) holds. Thanks to (b) from property (i) of Proposition 7.18, we get that $\sup_{n\in\mathbb{N}} \mathbb{E}[|M_n|] < +\infty$. Using Theorem 4.21, we deduce that the martingale converges a.s. towards a limit, say M_∞. Since the random variables $(M_n, n \in \mathbb{N})$ are uniformly integrable, we deduce from Proposition 7.21 that the convergence holds also in L^1. Hence, property (i) holds. □

The next Lemma does not hold if we assume that Z is non-negative instead of integrable, see a counter-example page 31 in [1]. Recall that $\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$.

Corollary 4.25. *Let Z be an integrable real-valued random variable. Then the process $(\mathbb{E}[Z|\mathcal{F}_n], n \in \mathbb{N})$ is a closed martingale which converges a.s. and in L^1 towards $\mathbb{E}[Z|\mathcal{F}_\infty]$.*

Proof. Condition (ii) of Theorem 4.24 holds for the martingale $M = (M_n = \mathbb{E}[Z|\mathcal{F}_n], n \in \mathbb{N})$. Since (i) and (ii) of Theorem 4.24 are equivalent, we deduce that M converges a.s. and in L^1 to a real-valued random variable, say M_∞ which is integrable. Using (4.13), we get that for all $A \in \mathcal{F}_n$:

$$\mathbb{E}\left[(Z - M_\infty)\mathbf{1}_A\right] = \mathbb{E}\left[\mathbb{E}\left[(Z - M_\infty)|\mathcal{F}_n\right]\mathbf{1}_A\right] = 0.$$

This implies that the set $\mathcal{A} \subset \mathcal{F}$ of events A such that $\mathbb{E}\left[Z\mathbf{1}_A\right] = \mathbb{E}\left[M_\infty\mathbf{1}_A\right]$ contains $\bigcup_{n\in\mathbb{N}} \mathcal{F}_n$ which is stable by finite intersection. Since Z and M_∞ are integrable, we get, using dominated convergence, that $\mathcal{A}$ is also a λ-system. According to the monotone class theorem, $\mathcal{A}$ contains the σ-field generated by $\bigcup_{n\in\mathbb{N}} \mathcal{F}_n$, that is $\mathcal{F}_\infty$. Then, we deduce from Definition 2.2 and Lemma 2.3 that a.s. $M_\infty = \mathbb{E}[Z|\mathcal{F}_\infty]$. □

We can extend the optional stopping theorem for closed martingale to any stopping times.

Proposition 4.26. *Let $M = (M_n, n \in \mathbb{N})$ be a closed martingale and write M_∞ for its a.s. limit. Let τ and ν be stopping times such that $\nu \leq \tau$. Then, we have a.s.:*

$$\mathbb{E}\left[M_\tau \middle| \mathcal{F}_\nu\right] = M_\nu. \tag{4.14}$$

Proof. According to Lemma 4.7, we have for any stopping time τ' that a.s.:

$$\mathbb{E}\left[M_\infty | \mathcal{F}_{\tau'}\right] = \sum_{n \in \overline{\mathbb{N}}} \mathbf{1}_{\{\tau'=n\}} \mathbb{E}\left[M_\infty | \mathcal{F}_n\right] = \sum_{n \in \overline{\mathbb{N}}} \mathbf{1}_{\{\tau'=n\}} M_n = M_{\tau'},$$

where we used (4.13) for the second equality. Using this result twice first with $\tau' = \tau$ and then with $\tau' = \nu$, we get, as $\mathcal{F}_\nu \subset \mathcal{F}_\tau$ according to property (iii) of Lemma 4.10, that a.s.:

$$\mathbb{E}\left[M_\tau | \mathcal{F}_\nu\right] = \mathbb{E}\left[\mathbb{E}\left[M_\infty | \mathcal{F}_\tau\right] | \mathcal{F}_\nu\right] = \mathbb{E}\left[M_\infty | \mathcal{F}_\nu\right] = M_\nu.$$

This gives the result. □

We have the following result when the martingale is bounded in L^p for some $p > 1$.

Proposition 4.27. *Let $M = (M_n,\, n \in \mathbb{N})$ be a martingale such that $\sup_{n \in \mathbb{N}} \mathbb{E}[|M_n|^p] < +\infty$ for some $p > 1$. Then, the martingale converges a.s. and in L^p towards a limit, say M_∞ and $M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$ a.s. for all $n \in \mathbb{N}$. We also have that $M^*_\infty = \sup_{n \in \mathbb{N}} |M_n|$ belongs to L^p and, with $C_p = (p/(p-1))^p$:*

$$\mathbb{E}\left[(M^*_\infty)^p\right] \leq C_p \mathbb{E}\left[|M_\infty|^p\right] \quad \text{as well as} \quad \mathbb{E}\left[|M_\infty|^p\right] = \sup_{n \in \mathbb{N}} \mathbb{E}[|M_n|^p].$$

Proof. Since M is bounded in L^1, we deduce from Theorem 4.21 that M converges a.s. towards a limit, say $M_\infty \in L^1$. We recall, see (4.7), that $M^*_n = \max_{k \in [\![0,n]\!]} |M_k|$. By monotone convergence, since $M^*_\infty = \lim_{n \to \infty} M^*_n$, we have that:

$$\mathbb{E}\left[(M^*_\infty)^p\right] = \lim_{n \to \infty} \mathbb{E}\left[(M^*_n)^p\right]. \tag{4.15}$$

According to Proposition 4.20 and since $\sup_{n \in \mathbb{N}} \mathbb{E}[|M_n|^p] < +\infty$, we deduce that:

$$\mathbb{E}\left[(M^*_\infty)^p\right] \leq C_p \sup_{n \in \mathbb{N}} \mathbb{E}[|M_n|^p] < +\infty.$$

This gives that M^*_∞ belongs to L^p. We deduce from (4.15) and the dominated convergence Theorem 1.46 (with $f_n = |M_n - M_\infty|^p$, $g_n = 2^{p-1}\left((M^*_n)^p + (M^*_\infty)^p\right)$ and $f_n \leq g_n$ as $(a+b)^p \leq 2^{p-1}(a^p + b^p)$ for $a, b \in \mathbb{R}_+$) that M converges in L^p towards M_∞. This implies in particular that $\mathbb{E}\left[|M_\infty|^p\right] = \lim_{n \to \infty} \mathbb{E}\left[|M_n|^p\right]$. Then, use that $\left(\mathbb{E}\left[|M_n|^p\right],\, n \in \mathbb{N}\right)$ is non-decreasing, see (4.8), to deduce that $\mathbb{E}\left[|M_\infty|^p\right] = \sup_{n \in \mathbb{N}} \mathbb{E}[|M_n|^p]$.

Since the martingale M converges in L^p towards M_∞, it also converges in L^1. We deduce then from Theorem 4.21 that $M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$ a.s. for all $n \in \mathbb{N}$. □

Bibliography

[1] J. Neveu. *Discrete-parameter martingales*. North-Holland Publishing Co., New York, revised edition, 1975.

[2] D. Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

Chapter 5

Optimal stopping

The goal of this chapter is to determine the best time, if any, at which one has to stop a game, seen as a stochastic process, in order to maximize a given criterion seen as a gain or a reward. The following two examples are typical of the problems which will be solved. Their solution are given respectively in Sections 5.1.3 and 5.3.2.

Example 5.1 (Marriage of a princess: the setting). In a faraway old age, a princess had to choose a prince for a marriage among $\zeta \in \mathbb{N}^*$ candidates. At step $1 \leq n < \zeta$, she interviews the n-th candidate and at the end of the interview she either accepts to marry this candidate or refuses. In the former case the process stops and she get married with the n-th candidate; in the latter case the rebuked candidate leaves forever and the princess moves on to step $n+1$. If $n = \zeta$, she has no more choice but to marry the last candidate. What is the best strategy or stopping rule for the princess if she wants to maximize the probability to marry the best prince?

This "Marriage problem", also known as the "Secretary problem", appeared in the late 1950's and early 1960's. See Ferguson [4] for an historical review as well as the corresponding Wikipedia page[1]. △

Example 5.2 (Castle to sell). A princess want to sell her castle, let X_n be the n-th price offer. However, preparing the castle for the visit of a potential buyer has a cost, say $c > 0$ per visit. So the gain of the selling at step $n \geq 1$ will be $G_n = X_n - nc$ or $G_n = \max_{1 \leq k \leq n} X_k - nc$ if the princess can recall a previous interested buyer. In this infinite time horizon setting, what is the best strategy for the princess in order to maximize her gain?

This "House-selling problem", see Chapter 4 in Ferguson [3] is also known as the "Job search problem" in economy, see Lippman and McCall [5]. △

For $n < \zeta \in \overline{\mathbb{N}} = \mathbb{N} \bigcup \{\infty\}$, we set $[\![n, \zeta]\!] = [n, \zeta] \bigcap \overline{\mathbb{N}}$ and $[\![n, \zeta[\![= [n, \zeta) \bigcap \overline{\mathbb{N}}$. We consider a game over the discrete time interval $[\![0, \zeta]\!]$ with horizon $\zeta \in \overline{\mathbb{N}}$, where at step $n \leq \zeta$ we can either stop and receive the gain G_n or continue to step $n+1$ if $n+1 \leq \zeta$. Eventually in the infinite horizon case, $\zeta = \infty$, if we never stop, we receive the gain G_∞. We assume the gains $G = (G_n, n \in [\![0, \zeta]\!])$ form a sequence of random variables on a probability space $(\Omega, \mathbb{P}, \mathcal{F})$ taking values in $[-\infty, +\infty)$.

We assume the information available is given by a filtration $\mathbb{F} = (\mathcal{F}_n, n \in [\![0, \zeta]\!])$ with $\mathcal{F}_n \subset \mathcal{F}$, and a strategy or stopping rule corresponds to a stopping time. Let $\mathbb{T}^\zeta$ be the set

[1] `https://en.wikipedia.org/wiki/Secretary_problem`

of all stopping times with respect to the filtration $\mathbb{F}$ taking values in $[\![0,\zeta]\!]$. We shall assume that $\mathbb{E}[G_\tau^+] < +\infty$ for all $\tau \in \mathbb{T}^\zeta$, where $x^+ = \max(0,x)$. In particular, the expectation $\mathbb{E}[G_\tau]$ is well defined and belongs to $[-\infty,+\infty)$. Thus, the maximal gain of the game G is:

$$V_* = \sup_{\tau\in\mathbb{T}^\zeta} \mathbb{E}[G_\tau]. \tag{5.1}$$

A stopping time $\tau' \in \mathbb{T}^\zeta$ is said optimal for G if $\mathbb{E}[G_{\tau'}] = V_*$ and thus $V_* = \max_{\tau\in\mathbb{T}^\zeta} \mathbb{E}[G_\tau]$.

The next theorem, which is a direct consequences of Corollaries 5.8 and 5.18, is the main result of this Chapter. For a real sequence $(a_n, n\in\overline{\mathbb{N}})$, we set $\limsup a_n = \lim_{n\nearrow\infty} \sup_{\infty>k\geq n} a_k$.

Theorem 5.3. *Let $\zeta \in \overline{\mathbb{N}}$, $G = (G_n, n \in [\![0,\zeta]\!])$ be a sequence of random variables taking values in $[-\infty,+\infty)$ and $\mathbb{F} = (\mathcal{F}_n, n\in[\![0,\zeta]\!])$ be a filtration. Assume the integrability condition:*

$$\mathbb{E}\Big[\sup_{n\in[\![0,\zeta]\!]} G_n^+\Big] < +\infty. \tag{5.2}$$

If $\zeta\in\mathbb{N}$ or if $\zeta=\infty$ and

$$\limsup G_n \leq G_\infty \quad a.s., \tag{5.3}$$

then, there exists an optimal stopping time.

We complete Theorem 5.3 by giving a description of the optimal stopping times when the sequence G is adapted to the filtration $\mathbb{F}$, (5.2) holds and (5.3) holds if $\zeta=\infty$. In this case, we consider the Snell envelope $S = (S_n, n\in[\![0,\zeta]\!]\bigcap\mathbb{N})$ which is a particular solution to the so-called optimal equations or Bellman equations:

$$S_n = \max\left(G_n, \mathbb{E}[S_{n+1}|\mathcal{F}_n]\right) \quad \text{for } n\in[\![0,\zeta[\![. \tag{5.4}$$

More precisely, in the finite horizon case S is defined by $S_\zeta = G_\zeta$ and the backward recursion (5.4); in the infinite horizon case S is defined by (5.17) which satisfies (5.4) according to Proposition 5.16. In this setting, we will consider the stopping times $\tau_* \leq \tau_{**}$ in $\mathbb{T}^\zeta$:

$$\tau_* = \inf\{n\in[\![0,\zeta[\![;\ S_n = G_n\}, \tag{5.5}$$
$$\tau_{**} = \inf\{n\in[\![0,\zeta[\![;\ S_n > \mathbb{E}[S_{n+1}|\mathcal{F}_n]\}, \tag{5.6}$$

with the convention $\inf\emptyset = \zeta$. We shall prove that they are optimal, see Propositions 5.6 and 5.17, and Exercises 5.1 and 5.5. Furthermore, if $V_* > -\infty$, then a stopping time τ is optimal if and only if $\tau_* \leq \tau \leq \tau_{**}$ a.s. and on $\{\tau<\infty\}$ we have a.s. $S_\tau = G_\tau$. See Exercises 5.1, 5.4 and 5.5. Thus, τ_* is the minimal optimal stopping time and τ_{**} the maximal one.

In the following two Remarks, we comment on the integrability condition (5.2) and we consider the case when the sequence G is not adapted to the filtration $\mathbb{F}$.

Remark 5.4. Notice that (5.2) implies that $\mathbb{E}[G_\tau^+] < +\infty$ for all $\tau\in\mathbb{T}^\zeta$. When $\zeta<\infty$, then (5.2) is equivalent to

$$\mathbb{E}[G_n^+] < +\infty \quad \text{for all } n\in[\![0,\zeta]\!]. \tag{5.7}$$

When $\zeta=\infty$, Condition (5.2) can be slightly weaken, see Proposition 5.17, when G is $\mathbb{F}$-adapted to Condition (H) page 94 which corresponds to the gain being bounded from above by a non-negative uniformly integrable martingale. $\Diamond$

Remark 5.5. When the sequence G is not adapted to the filtration $\mathbb{F}$, the idea is to check that an optimal stopping time for the adapted gain $G' = (G'_n, n \in [\![0, \zeta]\!])$ with $G'_n = \mathbb{E}[G_n | \mathcal{F}_n]$ is also an optimal stopping time for G, see Sections 5.1.2 and 5.2.4. $\Diamond$

The finite horizon case, $\zeta < \infty$, is presented in Section 5.1, and the infinite horizon case, $\zeta = \infty$, which is much more delicate in particular for the definition of S, is presented in Section 5.2. We consider the approximation of the infinite horizon case by finite horizon cases in Section 5.3, which includes the Markov chain setting developed in Section 5.3.3.

The presentation of this Chapter follows closely Ferguson [3] also inspired by Snell [7], see also Chow, Robbins and Siegmund [1, 6] and the references therein or for the Markovian setting Dynkin [2]. Concerning the infinite horizon case, we consider stopping time taking values in $\overline{\mathbb{N}}$ instead of $\mathbb{N}$ in most text books. Since in some standard applications, the gain of not stopping in finite time is $G_\infty = -\infty$ (which *de facto* implies the optimal stopping time is finite unless $V_* = -\infty$), we shall consider rewards G_n taking values in $[-\infty, +\infty)$, whereas in most text books it is assumed that $\mathbb{E}[|G_n|] < +\infty$ holds for all finite $n \leq \zeta$. The advantage of this setting is the simplicity of the hypothesis and the generality of the result given in Theorem 5.3. Its drawback is that we can not use the elegant martingale theory which is the corner stone of the Snell envelope approach, see Remark 5.7 and Exercise 5.1 and the presentation in Neveu [6]. Thus, we shall deal with integral technicalities in the infinite horizon case.

5.1 Finite horizon case

We assume in this section that $\zeta \in \mathbb{N}$ and that the gain process $G = (G_n, n \in [\![0, \zeta]\!])$ satisfies the integrability condition (5.7), or equivalently (5.2). We consider the filtration $\mathbb{F} = (\mathcal{F}_n, n \in [\![0, \zeta]\!])$. Recall $\mathbb{T}^\zeta$ is the set of stopping times with respect to the filtration $\mathbb{F}$ taking values in $[\![0, \zeta]\!]$. Notice that (5.7) implies that $\mathbb{E}[G_\tau^+] < +\infty$ for all $\tau \in \mathbb{T}^\zeta$.

We shall first consider in Section 5.1.1 that the gain process G is adapted to the filtration $\mathbb{F}$. This is not always the case. Indeed, in Example 5.1 on the marriage of a princess, the gain at step $n \in [\![1, \zeta]\!]$ is given by $G_n = \mathbf{1}_{\{\Sigma_n = 1\}}$, with Σ_n the random rank of the n-th candidate among the ζ candidates. In particular the rank Σ_n and thus the gain G_n are not observed unless $n = \zeta$, and thus the gain process is not adapted to the filtration generated by the observations. We extend the results of Section 5.1.1 to the case where G is not adapted to $\mathbb{F}$ in Section 5.1.2. Then, we solve the marriage problem in Section 5.1.3.

5.1.1 The adapted case

We assume that the sequence G is adapted to the filtration $\mathbb{F}$. We define the sequence $S = (S_n, n \in [\![0, \zeta]\!])$ recursively by $S_\zeta = G_\zeta$ and the optimal equations (5.4). The following Proposition gives a solution to the optimal stopping problem in the setting of this section.

Proposition 5.6. *Let $\zeta \in \mathbb{N}$ and $G = (G_n, n \in [\![0, \zeta]\!])$ be an adapted sequence such that $\mathbb{E}[G_n^+] < +\infty$ for all $n \in [\![0, \zeta]\!]$. The stopping time τ_* given by (5.5), with $(S_n, n \in [\![0, \zeta]\!])$ defined by $S_\zeta = G_\zeta$ and (5.4), is optimal and $V_* = \mathbb{E}[G_{\tau_*}] = \mathbb{E}[S_0]$.*

Proof. For $n \in [\![0, \zeta]\!]$, we define $\mathbb{T}_n$ as the set of all stopping times with respect to the filtration $\mathbb{F}$ taking values in $[\![n, \zeta]\!]$, as well as the stopping time $\tau_n = \inf\{k \in [\![n, \zeta]\!];\ S_k = G_k\}$. Notice that $n \leq \tau_n \leq \zeta$. We first prove by backward induction that:

$$S_n \geq \mathbb{E}[G_\tau|\mathcal{F}_n] \quad \text{a.s. for all } \tau \in \mathbb{T}_n, \tag{5.8}$$
$$S_n = \mathbb{E}[G_{\tau_n}|\mathcal{F}_n] \quad \text{a.s..} \tag{5.9}$$

Notice that (5.8) and (5.9) are clear for $n = \zeta$ as $S_\zeta = G_\zeta$.

Let $n \in [\![0, \zeta - 1]\!]$. We assume (5.8) and (5.9) hold for $n + 1$ and prove them for n. Let $\tau \in \mathbb{T}_n$ and consider $\tau' = \max(\tau, n + 1) \in \mathbb{T}_{n+1}$. As $\tau = \tau'$ on $\{\tau > n\}$, we have:

$$\mathbb{E}[G_\tau|\mathcal{F}_n] = G_n \mathbf{1}_{\{\tau=n\}} + \mathbb{E}[G_{\tau'}|\mathcal{F}_n]\mathbf{1}_{\{\tau>n\}}. \tag{5.10}$$

Using Inequality (5.8) with $n + 1$ and τ', we get that a.s.:

$$\mathbb{E}[G_{\tau'}|\mathcal{F}_n] = \mathbb{E}\left[\mathbb{E}[G_{\tau'}|\mathcal{F}_{n+1}]\mathcal{F}_n\right] \leq \mathbb{E}[S_{n+1}|\mathcal{F}_n]. \tag{5.11}$$

Using the optimal equations (5.4), we get a.s.:

$$\mathbb{E}[S_{n+1}|\mathcal{F}_n] \leq S_n. \tag{5.12}$$

Since (5.4) gives also $G_n \leq S_n$, we get using (5.10) that a.s.

$$\mathbb{E}[G_\tau|\mathcal{F}_n] \leq S_n. \tag{5.13}$$

This gives (5.8).

Consider τ_n instead of τ in (5.10). Then notice that on $\{\tau_n > n\}$, we have $\max(\tau_n, n+1) = \tau_{n+1}$. Then the inequality in (5.11) (with $\tau' = \tau_{n+1}$) is in fact an equality thanks to (5.9) (with $n + 1$). The inequality in (5.12) is also an equality on $\{\tau_n > n\}$ by definition of τ_n. Then use that $G_n = S_n$ on $\{\tau_n = n\}$, so that (5.13), with τ_n instead of τ, is also an equality. This gives (5.9). We then deduce that (5.8) and (5.9) hold for all $n \in [\![0, \zeta]\!]$.

Notice that $\tau_* = \tau_0$ by definition. We deduce from (5.8), with $n = 0$, that $\mathbb{E}[S_0] \geq \mathbb{E}[G_\tau]$ for all $\tau \in \mathbb{T}^\zeta$, and from (5.9), that $\mathbb{E}[S_0] = \mathbb{E}[G_{\tau_*}]$. This gives $V_* = \mathbb{E}[S_0]$ and τ_* is optimal. □

Remark 5.7 (Snell envelope). Let $\zeta \in \mathbb{N}$. Assume that $\mathbb{E}[|G_n|] < \infty$ for all $n \in [\![0, \zeta]\!]$. Notice from (5.4) that S is a super-martingale and that S dominates G. It is left to the reader to check that S is in fact the smallest super-martingale which dominates G. It is called the Snell enveloppe of G. For $n \in [\![0, \zeta[\![$, using that $S_n = \mathbb{E}[S_{n+1}|\mathcal{F}_n]$ on $\{\tau_* > n\}$, we have:

$$S_{n\wedge\tau_*} = S_{\tau_*}\mathbf{1}_{\{\tau_*\leq n\}} + S_n\mathbf{1}_{\{\tau_*>n\}} = S_{\tau_*}\mathbf{1}_{\{\tau_*\leq n\}} + \mathbb{E}[S_{n+1}\mathbf{1}_{\{\tau_*>n\}}|\mathcal{F}_n] = \mathbb{E}\left[S_{(n+1)\wedge\tau_*}|\mathcal{F}_n\right]. \tag{5.14}$$

This gives that $(S_{n\wedge\tau_*}, n \in [\![0, \zeta]\!])$ is a martingale. ◊

Exercise 5.1. Let $\zeta \in \mathbb{N}$. Assume that $\mathbb{E}[|G_n|] < \infty$ for all $n \in [\![0, \zeta]\!]$.

1. Prove that τ is an optimal stopping time if and only if $S_\tau = G_\tau$ a.s. and $(S_{n\wedge\tau}, n \in [\![0, \zeta]\!])$ is a martingale.

2. Deduce that τ_* is the minimal optimal stopping time (that is: if τ is optimal, then a.s. $\tau \geq \tau_*$).

3. Prove that τ_{**} defined by (5.6) is an optimal stopping time.
4. Using the Doob decomposition, see Remark 4.15, of the super-martingale S, prove that if $\tau \geq \tau_{**}$ is an optimal stopping time then $\tau = \tau_{**}$.
5. Arguing as in the proof of property (ii) from Lemma 5.13, prove that if τ and τ' are optimal stopping times so is $\max(\tau, \tau')$.
6. Deduce that τ is an optimal stopping time if and only if a.s. $\tau_* \leq \tau \leq \tau_{**}$ and $S_\tau = G_\tau$.

$\triangle$

5.1.2 The general case

If the sequence $G = (G_n, n \in [\![0, \zeta]\!])$ is not adapted to the filtration $\mathbb{F}$, then we shall consider the corresponding adapted sequence $G' = (G'_n, n \in [\![0, \zeta]\!])$ defined by:

$$G'_n = \mathbb{E}[G_n | \mathcal{F}_n].$$

Thanks to Jensen inequality, we have $\mathbb{E}[(G'_n)^+] \leq \mathbb{E}[G_n^+] < +\infty$ for all $n \in [\![0, \zeta]\!]$. Thus the sequence G' is is adapted to $\mathbb{F}$ and satisfies the integrability condition (5.7) or equivalently (5.2). Recall $\mathbb{T}^\zeta$ is the set of all stopping time with respect to the filtration $\mathbb{F}$ taking values in $[\![0, \zeta]\!]$. Thanks to Fubini, we get that for $\tau \in \mathbb{T}^\zeta$:

$$\mathbb{E}[G_\tau] = \sum_{n=0}^{\zeta} \mathbb{E}\left[G_n \mathbf{1}_{\{\tau = n\}}\right] = \sum_{n=0}^{\zeta} \mathbb{E}\left[G'_n \mathbf{1}_{\{\tau = n\}}\right] = \mathbb{E}[G'_\tau].$$

We thus deduce the maximal gain for the game G is also the maximal gain for the game G'. The following Corollary is then an immediate consequence of Proposition 5.6.

Corollary 5.8. *Let $\zeta \in \mathbb{N}$ and $G = (G_n, n \in [\![0, \zeta]\!])$ be such that $\mathbb{E}[G_n^+] < +\infty$ for all $n \in [\![0, \zeta]\!]$. Set $S_\zeta = \mathbb{E}[G_\zeta | \mathcal{F}_\zeta]$ and $S_n = \max\left(\mathbb{E}[G_n | \mathcal{F}_n], \mathbb{E}[S_{n+1} | \mathcal{F}_n]\right)$ for $0 \leq n < \zeta$. Then the stopping time $\tau_* = \inf\{n \in [\![0, \zeta]\!];\, S_n = \mathbb{E}[G_n | \mathcal{F}_n]\}$ is optimal and $V_* = \mathbb{E}[G_{\tau_*}] = \mathbb{E}[S_0]$.*

5.1.3 Marriage of a princess

We continue Example 5.1. The princess wants to maximize the probability to marry the best prince among $\zeta \in \mathbb{N}^*$ candidates. The corresponding gain at step n is $G_n = \mathbf{1}_{\{\Sigma_n = 1\}}$, with Σ_n the random rank of the n-th candidate among the ζ candidates. The random variable $\Sigma = (\Sigma_n, n \in [\![1, \zeta]\!])$ takes values in the set $\mathcal{S}_\zeta$ of permutation on $[\![1, \zeta]\!]$.

For a permutation $\sigma = (\sigma_n, n \in [\![1, \zeta]\!]) \in \mathcal{S}_\zeta$, we define the sequence of partial ranks $r(\sigma) = (r_1, \ldots, r_\zeta)$ such that r_n is the partial rank of σ_n in $(\sigma_1, \ldots, \sigma_n)$. In particular, we have $r_1 = 1$ and $r_\zeta = \sigma_\zeta$. Set $E = \prod_{n=1}^{\zeta} [\![1, n]\!]$ the state space of $r(\sigma)$. It is easy to check that r is a bijection from $\mathcal{S}_\zeta$ to E. Set $(R_1, \ldots, R_n) = r(\Sigma)$, so that R_n is the observed partial rank of the n-th candidate. In particular R_n corresponds to the observation of the princess at step n, and the information of the princess at step n is given by the σ-field $\mathcal{F}_n = \sigma(R_1, \ldots, R_n)$. In order to stick to the formalism of this chapter, we set $G_0 = -\infty$ and $\mathcal{F}_0$ the trivial σ-field.

We assume the princes are interviewed at random, that is the random permutation $\Sigma = (\Sigma_n, n \in [\![1, \zeta]\!])$ is uniformly distributed on $\mathcal{S}_\zeta$. Notice then that, for $n \in [\![1, \zeta[\![$, Σ_n is not a function of $(R_1, \ldots, R_n)$ and so it is not $\mathcal{F}_n$-measurable and thus the gain sequence $G = (G_n, n \in [\![0, \zeta]\!])$ is not adapted to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in [\![0, \zeta]\!])$.

Since r is a bijection, we deduce that $r(\Sigma)$ is uniform on E. Since E has a product form, we get that the random variables $R_1, \ldots, R_\zeta$ are independent and R_n is uniform on $[\![1, n]\!]$ for all $n \in [\![1, \zeta]\!]$. The event $\{\Sigma_n = 1\}$ is equal to $\{R_n = 1\} \bigcap_{k=n+1}^{\zeta} \{R_k > 1\}$. Using the independence of $(R_{n+1}, \ldots, R_\zeta)$ with $\mathcal{F}_n$, we deduce that for $n \in [\![1, \zeta]\!]$:

$$\mathbb{E}[G_n | \mathcal{F}_n] = \mathbb{E}[\mathbf{1}_{\{\Sigma_n = 1\}} | \mathcal{F}_n] = \mathbf{1}_{\{R_n = 1\}} \prod_{k=n+1}^{\zeta} \mathbb{P}(R_k > 1) = \frac{n}{\zeta} \mathbf{1}_{\{R_n = 1\}}.$$

By an elementary backward induction, we get from the definition of S_n given in Corollary 5.8 that, for $n \in [\![1, \zeta]\!]$, S_n is a function of R_n and more precisely $S_n = \max\left(\frac{n}{\zeta} \mathbf{1}_{\{R_n = 1\}}, s_{n+1}\right)$, with $s_{n+1} = \mathbb{E}[S_{n+1} | \mathcal{F}_n] = \mathbb{E}[S_{n+1}]$ as S_{n+1}, which is a function of R_{n+1}, is independent of $\mathcal{F}_n$. The sequence $(s_n, n \in [\![1, \zeta]\!])$ is non-increasing as $(S_n, n \in [\![1, \zeta]\!])$ is a super-martingale. We deduce that the optimal stopping time can be written as $\tau_* = \gamma_{n_*}$ for some n_*, where for $n \in [\![1, \zeta]\!]$, the stopping rule γ_n corresponds to first observe $n - 1$ candidate and then choose the next one who is better than those who have been observed (or the last if there is none): $\gamma_n = \inf\{k \in [\![n, \zeta]\!];\, R_k = 1 \text{ or } k = \zeta\}$. We set $\Gamma_n = \mathbb{E}[G_{\gamma_n}]$ the gain corresponding to the strategy γ_n. We have $\Gamma_1 = 1/\zeta$ and for $n \in [\![2, \zeta]\!]$:

$$\Gamma_n = \sum_{k=n}^{\zeta} \mathbb{P}(\gamma_n = k, \Sigma_k = 1) = \sum_{k=n}^{\zeta} \mathbb{P}(R_n > 1, \ldots, R_k = 1, \ldots, R_\zeta > 1) = \frac{n-1}{\zeta} \sum_{k=n}^{\zeta} \frac{1}{k-1},$$

where we used the independence for the last equality. Notice that $\zeta \Gamma_1 = \zeta \Gamma_\zeta = 1$. For $n \in [\![1, \zeta - 1]\!]$, we have $\zeta(\Gamma_n - \Gamma_{n+1}) = 1 - \sum_{j=n}^{\zeta-1} 1/j$. We deduce that Γ_n is maximal for $n_* = \inf\{n \geq 1;\, \Gamma_n \geq \Gamma_{n+1}\} = \inf\{n \geq 1;\, \sum_{j=n}^{\zeta-1} 1/j \leq 1\}$. We also have $V_* = \Gamma_{n_*}$.

For ζ large, we get $n_* \sim \zeta/\mathrm{e}$, so the optimal strategy is to observe a fraction of order $1/\mathrm{e} \simeq 37\%$ of the candidates, and then choose the next best one (or the last if there is none); the probability to get the best prince is then $V_* = \Gamma_{n_*} \simeq n_*/\zeta \simeq 1/\mathrm{e} \simeq 37\%$.

Exercise 5.2 (Choosing the second best instead of the best[2]). Assume the princess knows the best prince is very likely to get a better proposal somewhere else, so that she wants to select the second best prince among ζ candidates instead of the best one. For $x > 0$, we set $\lfloor x \rfloor$ the only integer $n \in \mathbb{N}$ such that $x - 1 < n \leq x$. Prove that the optimal stopping rule is to reject the first $n_0 = \lfloor (\zeta - 1)/2 \rfloor$ candidates and then chose the first second best so far prince or the last if none that is $\tau_* = \inf\{k > n_0;\, R_k = 2 \text{ or } k = \zeta\}$ and that the optimal gain is:

$$V_* = \frac{n_0(\zeta - n_0)}{\zeta(\zeta - 1)}.$$

So for ζ large, we get $V_* \simeq 1/4$. Selecting the third best leads to a more complex optimal strategy. △

[2] J. S. Rose. A problem of optimal choice assignment. *Operarions Research*, 30(1):172-181, 1982.

5.2 Infinite horizon case

We assume in this section that $\zeta = \infty$. Let $(\mathcal{F}_n, n \in \mathbb{N})$ be a filtration. For simplicity, we write $\mathbb{T} = \mathbb{T}^\infty$ for the set of stopping times taking values in $\overline{\mathbb{N}}$. Notice the definition of stopping time, and thus of the set $\mathbb{T}$, does not depend on the choice of $\mathcal{F}_\infty$ as long as this σ-field contains $\mathcal{F}_n$ for all $n \in \mathbb{N}$. For this reason, we shall take for $\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$ the smallest possible σ-field which contains $\bigcup_{n\in\mathbb{N}} \mathcal{F}_n$, see Proposition 1.2. We also use the following convention for the limit operators: $\lim_{n\to\infty}$ will be understood as $\lim_{n\to\infty;\, n<\infty}$, and for a real sequence $(a_n, n \in \overline{\mathbb{N}})$, we set $\limsup a_n = \lim_{n\to\infty} \sup_{\infty>k\geq n} a_k$ as well as $\liminf a_n = \lim_{n\to\infty} \inf_{\infty>k\geq n} a_k$.

The next two examples prove one can not remove the hypothesis (5.2) and (5.3) on the gain process to ensure the existence of an optimal stopping time.

Example 5.9. We consider the gain process $G = (G_n, n \in \overline{\mathbb{N}})$ given by $G_n = n/(n+1)$ for $n \in \mathbb{N}$ and $G_\infty = 0$. Clearly we have $V_* = 1$ and there is no optimal stopping time. Notice that (5.3) does not hold in this case. △

Example 5.10. Let $(X_n, n \in \mathbb{N}^*)$ be independent Bernoulli random variables such that $\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = 0) = 1/2$. We consider the gain process $G = (G_n, n \in \overline{\mathbb{N}})$ given by $G_0 = 0$, $G_n = (2^n - 1)\prod_{k=1}^n X_k$ for $n \in \mathbb{N}^*$ and a.s. $G_\infty = \lim_{n\to\infty} G_n = 0$. Let $\mathbb{F}$ be the natural filtration of the process G. We have $\mathbb{E}[G_n] = 1 - 2^{-n}$ so that $V_* \geq 1$. Notice G is a non-negative sub-martingale as:

$$\mathbb{E}[G_{n+1}|\mathcal{F}_n] = \frac{2^{n+1}-1}{2^{n+1}-2} G_n \geq G_n.$$

Thus, for any $\tau \in \mathbb{T}$, we have $\mathbb{E}[G_{\tau\wedge n}] \leq \mathbb{E}[G_n] \leq 1$. And by Fatou Lemma, we get $\mathbb{E}[G_\tau] \leq 1$. Thus, we deduce that $V_* = 1$.

Since $\mathbb{E}[G_{n+1}|\mathcal{F}_n] > G_n$ on $\{G_n \neq 0\}$ and $G_{n+1} = G_n$ on $\{G_n = 0\}$, we get at step n that the expected future gain at step $n+1$ is better than the gain G_n. Therefore it is more interesting to continue than to stop at step n. However this strategy will provide the gain $G_\infty = 0$, and is thus not optimal. We deduce there is no optimal stopping time.

Consider the stopping time $\tau = \inf\{n \geq 1;\, G_n = 0\}$. We have that τ is a geometric random variable with parameter $1/2$. Furthermore, we have $\sup_{n\in[\![0,\zeta]\!]} G_n^+ = 2^{\tau-1} - 1$ and thus $\mathbb{E}\Big[\sup_{n\in[\![0,\zeta]\!]} G_n^+\Big] = +\infty$. In particular, condition (5.2) does not hold in this case. △

The main result of this section is that if (5.2) and (5.3) hold, then there exists an optimal stopping time $\tau_* \in \mathbb{T}$, see Corollary 5.18. The main idea of the infinite horizon case, inspired by the finite horizon case, is to consider a process $S = (S_n, n \in [\![0,\zeta]\!])$ satisfying the optimal equations (5.4). But since the initialization of S given in the finite horizon case is now useless, we shall rely on a definition inspired by (5.8) and (5.9). However, we need to consider a measurable version of the supremum of $\mathbb{E}[G_\tau|\mathcal{F}_n]$, where τ is any stopping time such that $\tau \geq n$. This is developed in Section 5.2.1. In the technical Section 5.2.2, due to the fact we don't assume the gain to be integrable, following Ferguson [3], we use the notion of regular stopping time to prove the existence of an optimal stopping time in the adapted case. We connect this result with the optimal equations (5.4) in Section 5.2.3. Then, we consider the general case in Section 5.2.4.

5.2.1 Essential supremum

The following proposition asserts the existence of a minimal random variable dominating a family (which might be uncountable) of random variables in the sense of a.s. inequality. We set $\overline{\mathbb{R}} = [-\infty, +\infty]$.

Proposition 5.11. *Let $(X_t, t \in T)$ be a family of real-valued random variables indexed by a general set T. There exists a unique (up to the a.s. equivalence) real-valued random variable X_* such that:*

(i) For all $t \in T$, $\mathbb{P}(X_* \geq X_t) = 1$.

(ii) If there exists a random variable Y such that for all $t \in T$, $\mathbb{P}(Y \geq X_t) = 1$, then a.s. $Y \geq X_*$.

The random variable X_* of the previous proposition is called the essential supremum of $(X_t, t \in T)$ and is denoted by:

$$X_* = \operatorname*{ess\,sup}_{t \in T} X_t.$$

Example 5.12. If U is a uniform random variable on $[0, 1]$, and $X_t = \mathbf{1}_{\{U=t\}}$ for $t \in T = [0, 1]$, then we have that a.s. $\sup_{t \in T} X_t = 1$ and it is easy to check that a.s. $\operatorname{ess\,sup}_{t \in T} X_t = 0$. △

Proof of Proposition 5.11. Since we are only considering inequalities between real random variables, by mapping $\overline{\mathbb{R}}$ onto $[0, 1]$ with an increasing bijection, we can assume that X_t takes values in $[0, 1]$ for all $t \in T$.

Let $\mathcal{I}$ be the family of all countable sub-families of T. For each $I \in \mathcal{I}$, consider the (well defined) random variable $X_I = \sup_{t \in I} X_t$ and define $\alpha = \sup_{I \in \mathcal{I}} \mathbb{E}[X_I]$. There exists a sequence $(I_n, n \in \mathbb{N})$ such that $\lim_{n \to +\infty} \mathbb{E}[X_{I_n}] = \alpha$. The set $I_* = \bigcup_{n \in \mathbb{N}} I_n$ is countable and thus $I_* \in \mathcal{I}$. Set $X_* = X_{I_*}$. Since $\mathbb{E}[X_{I_n}] \leq \mathbb{E}[X_*] \leq \alpha$ for all $n \in \mathbb{N}$, we get $\mathbb{E}[X_*] = \alpha$.

For any $t \in T$, consider $J = I_* \bigcup \{t\}$, which belongs to $\mathcal{I}$, and notice that $X_J = \max(X_t, X_*)$. Since $\alpha = \mathbb{E}[X_*] \leq \mathbb{E}[X_J] \leq \alpha$, we deduce that $\mathbb{E}[X_*] = \mathbb{E}[X_J]$ and thus a.s. $X_J = X_*$, that is $\mathbb{P}(X_* \geq X_t) = 1$. This gives (i).

Let Y be as in (ii). Since I_* is countable, we get that a.s. $Y \geq X_*$. This gives (ii). □

5.2.2 The adapted case: regular stopping times

We assume in this section that the sequence $G = (G_n, n \in \overline{\mathbb{N}})$ is adapted to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in \overline{\mathbb{N}})$, with $\mathcal{F}_\infty = \bigvee_{n \in \mathbb{N}} \mathcal{F}_n$. We shall consider the following hypothesis which is slightly weaker than (5.2):

(H) There exists a non-negative integrable random variable M such that for all $n \in \overline{\mathbb{N}}$, we have a.s. $G_n^+ \leq \mathbb{E}[M|\mathcal{F}_n]$.

Condition (H) and (4.1) imply that for all $\tau \in \mathbb{T}$, we have a.s. $G_\tau^+ \leq \mathbb{E}[M|\mathcal{F}_\tau]$. Notice that if (5.2) holds then (H) holds with $M = \sup_{k \in \overline{\mathbb{N}}} G_k^+$.

For $n \in \mathbb{N}$, let $\mathbb{T}_n = \{\tau \in \mathbb{T};\, \tau \geq n\}$ be the set of stopping times larger than or equal to n. We say a stopping times $\tau \in \mathbb{T}_n$ is regular, which will be understood with respect to G, if for all finite $k \geq n$:

$$\mathbb{E}[G_\tau|\mathcal{F}_k] > G_k \quad \text{a.s. on } \{\tau > k\}.$$

We denote by $\mathbb{T}'_n$ the subset of $\mathbb{T}_n$ of regular stopping times. Notice that $\mathbb{T}'_n$ is non-empty as it contains n.

Lemma 5.13. *Assume that G is adapted and a.s. $\mathbb{E}[G_\tau^+] < +\infty$ for all $\tau \in \mathbb{T}$. Let $n \in \mathbb{N}$.*

(i) If $\tau \in \mathbb{T}_n$, then there exists a regular stopping time $\tau' \in \mathbb{T}'_n$ such that $\tau' \leq \tau$ and a.s. $\mathbb{E}[G_{\tau'}|\mathcal{F}_n] \geq \mathbb{E}[G_\tau|\mathcal{F}_n]$.

(ii) If $\tau', \tau'' \in \mathbb{T}'_n$ are regular, then the stopping time $\tau = \max(\tau', \tau'') \in \mathbb{T}'_n$ is regular and a.s. $\mathbb{E}[G_\tau|\mathcal{F}_n] \geq \max\left(\mathbb{E}[G_{\tau'}|\mathcal{F}_n], \mathbb{E}[G_{\tau''}|\mathcal{F}_n]\right)$.

Proof. We prove property (i). Let $\tau \in \mathbb{T}_n$ and set $\tau' = \inf\{k \geq n; \mathbb{E}[G_\tau|\mathcal{F}_k] \leq G_k\}$ with the convention that $\inf \emptyset = \infty$. Notice that τ' is a stopping time and that a.s. $n \leq \tau' \leq \tau$. On $\{\tau' = \infty\}$, we have $\tau = \infty$ and a.s. $G_{\tau'} = G_\infty = G_\tau$. For $\infty > m \geq n$, we have, on $\{\tau' = m\}$, that a.s. $\mathbb{E}[G_{\tau'}|\mathcal{F}_m] = G_m \geq \mathbb{E}[G_\tau|\mathcal{F}_m]$. We deduce that for all finite $k \geq n$ a.s. on $\{\tau' \geq k\}$:

$$\mathbb{E}[G_{\tau'}|\mathcal{F}_k] = \sum_{m \in [\![k,\infty]\!]} \mathbb{E}\left[\mathbb{E}[G_{\tau'}|\mathcal{F}_m]\mathbf{1}_{\{\tau'=m\}}|\mathcal{F}_k\right] \geq \sum_{m \in [\![k,\infty]\!]} \mathbb{E}\left[\mathbb{E}[G_\tau|\mathcal{F}_m]\mathbf{1}_{\{\tau'=m\}}|\mathcal{F}_k\right].$$

And thus, for all finite $k \geq n$:

$$\mathbb{E}[G_{\tau'}|\mathcal{F}_k]\mathbf{1}_{\{\tau' \geq k\}} \geq \mathbb{E}[G_\tau|\mathcal{F}_k]\mathbf{1}_{\{\tau' \geq k\}}. \tag{5.15}$$

We have on $\{\tau' > k\}$, $\mathbb{E}[G_\tau|\mathcal{F}_k] > G_k$. Then use (5.15) to get that τ' is regular. Take $k = n$ in (5.15) and use that $\tau' \geq n$ a.s. to get the last part of (i).

We prove property (ii). Let $\tau', \tau'' \in \mathbb{T}'_n$ and $\tau = \max(\tau', \tau'')$. By construction τ is a stopping time, see Proposition 4.4. We have for all $m \geq k \geq n$ and k finite:

$$\mathbb{E}\left[G_\tau \mathbf{1}_{\{\tau'=m\}}|\mathcal{F}_k\right] = \mathbb{E}\left[G_{\tau'}\mathbf{1}_{\{m=\tau' \geq \tau''\}}|\mathcal{F}_k\right] + \mathbb{E}\left[G_{\tau''}\mathbf{1}_{\{\tau'' > \tau'=m\}}|\mathcal{F}_k\right].$$

Using that $\tau'' \in \mathbb{T}'_n$, we get for all finite $m \geq k \geq n$:

$$\mathbb{E}[G_{\tau''}\mathbf{1}_{\{\tau''>\tau'=m\}}|\mathcal{F}_k] = \mathbb{E}\left[\mathbb{E}[G_{\tau''}|\mathcal{F}_m]\mathbf{1}_{\{\tau''>m\}}\mathbf{1}_{\{\tau'=m\}}|\mathcal{F}_k\right] \geq \mathbb{E}[G_m \mathbf{1}_{\{\tau''>\tau'=m\}}|\mathcal{F}_k].$$

We deduce that for all $m \geq k \geq n$ and k finite:

$$\mathbb{E}[G_\tau \mathbf{1}_{\{\tau'=m\}}|\mathcal{F}_k] \geq \mathbb{E}[G_{\tau'}\mathbf{1}_{\{\tau'=m\}}|\mathcal{F}_k]. \tag{5.16}$$

By summing (5.16) over m with $m > k$ and using that $\tau' \in \mathbb{T}'_n$, we get:

$$\mathbb{E}[G_\tau|\mathcal{F}_k]\mathbf{1}_{\{\tau'>k\}} \geq \mathbb{E}[G_{\tau'}|\mathcal{F}_k]\mathbf{1}_{\{\tau'>k\}} > G_k \mathbf{1}_{\{\tau'>k\}}.$$

By symmetry, we also get $\mathbb{E}[G_\tau|\mathcal{F}_k]\mathbf{1}_{\{\tau''>k\}} > G_k \mathbf{1}_{\{\tau''>k\}}$. Since $\{\tau > k\} = \{\tau' > k\} \bigcup \{\tau'' > k\}$, this implies that $\mathbb{E}[G_\tau|\mathcal{F}_k] > G_k$ a.s. on $\{\tau > k\}$. Thus, τ is regular.

By summing (5.16) over m with $m \geq k = n$, and using that $\tau' \geq n$ a.s., we get $\mathbb{E}[G_\tau|\mathcal{F}_n] \geq \mathbb{E}[G_{\tau'}|\mathcal{F}_n]$. By symmetry, we also have $\mathbb{E}[G_\tau|\mathcal{F}_n] \geq \mathbb{E}[G_{\tau''}|\mathcal{F}_n]$. We deduce the last part of (ii). □

The next lemma is the main result of this section.

Lemma 5.14. *We assume that G is adapted and hypothesis (H) and (5.3) hold. Then, for all $n \in \mathbb{N}$, there exists $\tau_n^\circ \in \mathbb{T}_n$ such that a.s.* $\operatorname{ess\,sup}_{\tau \in \mathbb{T}_n} \mathbb{E}[G_\tau | \mathcal{F}_n] = \mathbb{E}[G_{\tau_n^\circ} | \mathcal{F}_n]$.

Proof. We set $X_* = \operatorname{ess\,sup}_{\tau \in \mathbb{T}_n} \mathbb{E}[G_\tau | \mathcal{F}_n]$. According to the proof of Proposition 5.11, there exists a sequence $(\tau_k, k \in \mathbb{N})$ of elements of $\mathbb{T}_n$ such that $X_* = \sup_{k \in \mathbb{N}} \mathbb{E}[G_{\tau_k} | \mathcal{F}_n]$. Thanks to (i) of Lemma 5.13, there exists a sequence $(\tau'_k, k \in \mathbb{N})$ of regular stopping times, elements of $\mathbb{T}'_n$, such that $\mathbb{E}[G_{\tau'_k} | \mathcal{F}_n] \geq \mathbb{E}[G_{\tau_k} | \mathcal{F}_n]$. According to (ii) of Lemma 5.13, for all $k \in \mathbb{N}$, the stopping time $\tau''_k = \max_{0 \leq j \leq k} \tau'_j$ belongs to $\mathbb{T}'_n$, the sequence $(\mathbb{E}[G_{\tau''_k} | \mathcal{F}_n],\, k \in \mathbb{N})$ is non-decreasing and $\mathbb{E}[G_{\tau''_k} | \mathcal{F}_n] \geq \mathbb{E}[G_{\tau'_k} | \mathcal{F}_n] \geq \mathbb{E}[G_{\tau_k} | \mathcal{F}_n]$. In particular, we get $X_* = \sup_{k \in \mathbb{N}} \mathbb{E}[G_{\tau_k} | \mathcal{F}_n] \leq \sup_{k \in \mathbb{N}} \mathbb{E}[G_{\tau''_k} | \mathcal{F}_n] \leq X_*$, so that a.s. $X_* = \lim_{k \to \infty} \mathbb{E}[G_{\tau''_k} | \mathcal{F}_n]$.

Let $\tau_n^\circ \in \mathbb{T}_n$ be the limit of the non-decreasing sequence $(\tau''_k,\, k \in \mathbb{N})$. Set $Y_k = \mathbb{E}[M | \mathcal{F}_{\tau''_k}]$. We deduce from the optional stopping theorem for closed martingale, see Proposition 4.26, that $(Y_k, k \in \mathbb{N})$ is a martingale with respect to the filtration $(\mathcal{F}_{\tau_k},\, k \in \mathbb{N})$, which is closed thanks to property (ii) from Theorem 4.24. In particular, the sequence $(Y_k, k \in \mathbb{N})$ converges a.s. and in L^1 towards $Y_\infty = \mathbb{E}[M | \mathcal{F}_{\tau_n^\circ}]$ according to Corollary 4.25. Notice also that a.s. $\mathbb{E}\left[Y_k \middle| \mathcal{F}_n\right] = \mathbb{E}\left[Y_\infty \middle| \mathcal{F}_n\right]$. Then, we use Lemma 5.30 with $X_k = G_{\tau''_k}$ to get that a.s. $X_* \leq \mathbb{E}[\limsup_{k \to \infty} G_{\tau''_k} | \mathcal{F}_n]$. Thanks to (5.3), we have a.s. $\limsup_{k \to \infty} G_{\tau''_k} \leq G_{\tau_n^\circ}$. So we get that a.s. $X_* \leq \mathbb{E}[G_{\tau_n^\circ} | \mathcal{F}_n]$. To conclude use that by definition of X_*, we have $\mathbb{E}[G_{\tau_n^\circ} | \mathcal{F}_n] \leq X_*$ and thus $X_* = \mathbb{E}[G_{\tau_n^\circ} | \mathcal{F}_n]$. □

We have the following Corollary.

Corollary 5.15. *We assume that G is adapted and hypothesis (H) and (5.3) hold. Then, we have that τ_0° is optimal that is $V_* = \mathbb{E}[G_{\tau_0^\circ}]$.*

Proof. Lemma 5.14 gives that $\mathbb{E}[G_\tau] \leq \mathbb{E}[G_{\tau_0^\circ}]$ for all $\tau \in \mathbb{T}$. Thus τ_0° is optimal. □

Exercise 5.3. Assume that hypothesis (H) and (5.3) hold. Let $n \in \mathbb{N}$. Prove that the limit of a non-decreasing sequence of regular stopping times, elements of $\mathbb{T}'_n$, is regular. Deduce that τ_n° in Lemma 5.14 is regular, that is τ_n° belongs to $\mathbb{T}'_n$. △

5.2.3 The adapted case: optimal equations

We assume in this section that the sequence $G = (G_n, n \in \overline{\mathbb{N}})$ is adapted to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in \overline{\mathbb{N}})$, with $\mathcal{F}_\infty = \bigvee_{n \in \mathbb{N}} \mathcal{F}_n$. Recall that $\mathbb{T}_n = \{\tau \in \mathbb{T};\, \tau \geq n\}$ for $n \in \mathbb{N}$. We assume (H) holds. We set for $n \in \mathbb{N}$:

$$S_n = \operatorname*{ess\,sup}_{\tau \in \mathbb{T}_n} \mathbb{E}[G_\tau | \, \mathcal{F}_n]. \tag{5.17}$$

The next proposition is the main result of this section.

Proposition 5.16. *We assume that G is adapted and hypothesis (H) and (5.3) hold. Then, for all $n \in \mathbb{N}$, we have $\mathbb{E}[S_n^+] < +\infty$. The sequence $(S_n, n \in \mathbb{N})$ satisfies the optimal equations (5.4). We also have $V_* = \mathbb{E}[S_0]$.*

Proof. Recall that (H) implies $\mathbb{E}[G_\tau^+] < +\infty$ for all $\tau \in \mathbb{T}_n$. Then use Lemma 5.14 to deduce that $\mathbb{E}[S_n^+] = \mathbb{E}[G_{\tau_n^\circ}^+] < +\infty$. For $\tau \in \mathbb{T}_n$, we have (5.10) and (5.11) by definition of the essential supremum for S_{n+1}. We deduce that a.s. $\mathbb{E}[G_\tau | \, \mathcal{F}_n] \leq \max(G_n, \mathbb{E}[S_{n+1} | \, \mathcal{F}_n])$. This implies, see (ii) of Proposition 5.11, that a.s. $S_n \leq \max(G_n, \mathbb{E}[S_{n+1} | \, \mathcal{F}_n])$.

Thanks to Lemma 5.14, there exists $\tau_{n+1}^\circ \in \mathbb{T}_{n+1}$ such that a.s. $S_{n+1} = \mathbb{E}[G_{\tau_{n+1}^\circ}|\mathcal{F}_{n+1}]$. Since τ_{n+1}° (resp. n) belongs also to $\mathbb{T}_n$, we have $S_n \geq \mathbb{E}[G_{\tau_{n+1}^\circ}|\mathcal{F}_n] = \mathbb{E}[S_{n+1}|\mathcal{F}_n]$ (resp. $S_n \geq G_n$). This implies that $S_n \geq \max(G_n, \mathbb{E}[S_{n+1}|\,\mathcal{F}_n])$. And thus $(S_n, n \in \mathbb{N})$ satisfies the optimal equations.

Use Corollary 5.15 and Lemma 5.14 to get $V_* = \mathbb{E}[S_0]$. □

We conclude this section by giving an explicit optimal stopping time.

Proposition 5.17. *We assume that G is adapted and hypothesis (H) and (5.3) hold. Then τ_* defined by (5.5), with $(S_n, n \in \mathbb{N})$ given by (5.17), is optimal: $V_* = \mathbb{E}[G_{\tau_*}]$.*

Proof. If $V_* = -\infty$ then nothing has to be proven. So, we assume $V_* > -\infty$. According to Corollary 5.15, there exists an optimal stopping time τ.

In a first step, we check that $\tau' = \min(\tau, \tau_*)$ is also optimal. Since $\mathbb{E}[G_\tau^+] < +\infty$, by Fubini and the definition of S_n, we have:

$$\mathbb{E}\left[G_\tau \mathbf{1}_{\{\tau > \tau_*\}}\right] = \sum_{n\in\mathbb{N}} \mathbb{E}\left[G_\tau \mathbf{1}_{\{\tau>\tau_*=n\}}\right] = \sum_{n\in\mathbb{N}} \mathbb{E}\left[\mathbb{E}[G_\tau|\mathcal{F}_n]\mathbf{1}_{\{\tau>\tau_*=n\}}\right] \leq \sum_{n\in\mathbb{N}} \mathbb{E}\left[S_n \mathbf{1}_{\{\tau>\tau_*=n\}}\right].$$

Since $S_n = G_n$ on $\{\tau_* = n\}$ for $n \in \mathbb{N}$, we deduce that:

$$\mathbb{E}\left[G_\tau \mathbf{1}_{\{\tau>\tau_*\}}\right] \leq \sum_{n\in\mathbb{N}} \mathbb{E}\left[G_n \mathbf{1}_{\{\tau>\tau_*=n\}}\right] = \mathbb{E}\left[G_{\tau_*} \mathbf{1}_{\{\tau>\tau_*\}}\right].$$

This implies that:

$$\mathbb{E}\left[G_\tau\right] = \mathbb{E}\left[G_\tau \mathbf{1}_{\{\tau>\tau_*\}}\right] + \mathbb{E}\left[G_\tau \mathbf{1}_{\{\tau\leq\tau_*\}}\right] \leq \mathbb{E}\left[G_{\tau_*} \mathbf{1}_{\{\tau>\tau_*\}}\right] + \mathbb{E}\left[G_\tau \mathbf{1}_{\{\tau\leq\tau_*\}}\right] = \mathbb{E}\left[G_{\tau'}\right].$$

And thus τ' is optimal.

In a second step we check that a.s. $\tau' = \tau_*$. Let us assume that $\mathbb{P}(\tau' < \tau_*) > 0$. Recall τ_n° defined in Lemma 5.14. We define the stopping time τ'' by $\tau'' = \tau_*$ on $\{\tau' = \tau_*\}$ and $\tau'' = \tau_n^\circ$ on $\{n = \tau' < \tau_*\}$ for $n \in \mathbb{N}$. Since $\mathbb{E}[G_{\tau''}^+] < +\infty$, by Fubini and the definition of S_n, we have:

$$\mathbb{E}\left[G_{\tau''} \mathbf{1}_{\{\tau'<\tau_*\}}\right] = \sum_{n\in\mathbb{N}} \mathbb{E}\left[G_{\tau_n^\circ} \mathbf{1}_{\{n=\tau'<\tau_*\}}\right] = \sum_{n\in\mathbb{N}} \mathbb{E}\left[\mathbb{E}[G_{\tau_n^\circ}|\mathcal{F}_n]\mathbf{1}_{\{n=\tau'<\tau_*\}}\right] = \sum_{n\in\mathbb{N}} \mathbb{E}\left[S_n \mathbf{1}_{\{n=\tau'<\tau_*\}}\right].$$

Since $\mathbb{P}(\tau' < \tau_*) > 0$ and $S_n > G_n$ on $\{\tau_* > n\}$ for $n \in \mathbb{N}$, we deduce that:

$$\mathbb{E}\left[G_{\tau''} \mathbf{1}_{\{\tau'<\tau_*\}}\right] > \sum_{n\in\mathbb{N}} \mathbb{E}\left[G_n \mathbf{1}_{\{n=\tau'<\tau_*\}}\right] = \mathbb{E}\left[G_{\tau'} \mathbf{1}_{\{\tau'<\tau_*\}}\right]$$

unless $\mathbb{E}\left[G_{\tau''} \mathbf{1}_{\{\tau'<\tau_*\}}\right] = \mathbb{E}\left[G_{\tau'} \mathbf{1}_{\{\tau'<\tau_*\}}\right] = -\infty$. The latter case is not possible since $\mathbb{E}\left[G_{\tau'}\right] = V_* > -\infty$. Thus, we deduce that $\mathbb{E}\left[G_{\tau''} \mathbf{1}_{\{\tau'<\tau_*\}}\right] > \mathbb{E}\left[G_{\tau'} \mathbf{1}_{\{\tau'<\tau_*\}}\right]$. This implies (using again that $\mathbb{E}[G_{\tau'}] > -\infty$) that:

$$\mathbb{E}\left[G_{\tau''}\right] = \mathbb{E}\left[G_{\tau'} \mathbf{1}_{\{\tau'=\tau_*\}}\right] + \mathbb{E}\left[G_{\tau''} \mathbf{1}_{\{\tau'<\tau_*\}}\right] > \mathbb{E}\left[G_{\tau'} \mathbf{1}_{\{\tau'=\tau_*\}}\right] + \mathbb{E}\left[G_{\tau'} \mathbf{1}_{\{\tau'<\tau_*\}}\right] = \mathbb{E}\left[G_{\tau'}\right].$$

This is impossible as τ' is optimal. Thus, we have a.s. $\tau' = \tau_*$ and τ_* is optimal. □

Exercise 5.4. Assume that G is adapted and hypothesis (H) and (5.3) hold and $V_* > -\infty$.

1. Deduce from the proof of Proposition 5.17, that τ_* defined by (5.5) is the minimal optimal stopping time: if τ is an optimal stopping time then a.s. $\tau \geq \tau_*$.

2. Deduce that if $G_\infty = -\infty$ a.s., then a.s. τ_* is finite.

△

Exercise 5.5. Assume that G is adapted and hypothesis (H) and (5.3) hold. We set for $n \in \mathbb{N}$:

$$W_n = \operatorname*{ess\,sup}_{\tau \in \mathbb{T}_{n+1}} \mathbb{E}[G_\tau | \, \mathcal{F}_n] \tag{5.18}$$

with the convention that $\inf \emptyset = \infty$. Recall τ_* and τ_{**} defined by (5.5) and (5.6).

1. Prove that $W_n = \mathbb{E}[S_{n+1} | \mathcal{F}_n]$.

2. Prove that $(S_{n \wedge \tau_{**}}, n \in \mathbb{N})$ is such that $\mathbb{E}[S_0] = \mathbb{E}[S_{n \wedge \tau_{**}}]$ for all $n \in \mathbb{N}$.

3. Prove that $\mathbb{E}[S_0] \leq \mathbb{E}[\limsup S_{n \wedge \tau_{**}}] \leq \mathbb{E}[G_{\tau_{**}}]$. Deduce that τ_{**} is optimal.

4. Assume that $V_* > -\infty$. Prove that if τ is an optimal stopping time, then $\tau \wedge \tau_{**}$ is also optimal. Prove that a.s. $\tau \leq \tau_{**}$.

5. Assume that $V_* > -\infty$. Prove that τ is an optimal stopping time if and only if a.s. $S_\tau = G_\tau$ on $\{\tau < \infty\}$ and $\tau_* \leq \tau \leq \tau_{**}$.

△

Exercise 5.6. Assume that G is adapted and hypothesis (H) and (5.3) hold, as well as $V_* > -\infty$. Prove that τ_* defined by (5.5) is regular. △

5.2.4 The general case

We state the main result of this section. Let $\mathbb{T}$ denote the set of stopping times (taking values in $\overline{\mathbb{N}}$) with respect to the filtration $(\mathcal{F}_n, n \in \mathbb{N})$.

Corollary 5.18. *Let $G = (G_n, n \in \overline{\mathbb{N}})$ be a sequence of random variables such that (5.2) and (5.3) hold. Then there exists an optimal stopping time.*

Proof. According to the first paragraph of Section 5.2, without loss of generality, we can assume that $\mathcal{F}_\infty = \bigvee_{n \in \mathbb{N}} \mathcal{F}_n$. If G is adapted to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in \overline{\mathbb{N}})$ then use $M = \sup_{n \in \overline{\mathbb{N}}} G_n^+$, so that (H) holds, and Corollary 5.15 to conclude.

If the sequence G is not adapted to the filtration $\mathbb{F}$, then we shall consider the corresponding adapted sequence $G' = (G'_n, n \in \overline{\mathbb{N}})$ given by $G'_n = \mathbb{E}[G_n | \mathcal{F}_n]$ for $n \in \overline{\mathbb{N}}$. Notice G' is well defined thanks to (5.2). Thanks to (5.2), we can use Fubini lemma to get for $\tau \in \mathbb{T}$:

$$\mathbb{E}[G_\tau] = \sum_{n \in \overline{\mathbb{N}}} \mathbb{E}[G_n \mathbf{1}_{\{\tau = n\}}] = \sum_{n \in \overline{\mathbb{N}}} \mathbb{E}[G'_n \mathbf{1}_{\{\tau = n\}}] = \mathbb{E}[G'_\tau].$$

We thus deduce the maximal gain for the game G is also the maximal gain for the game G'.

Let $M = \mathbb{E}\big[\sup_{n\in\overline{\mathbb{N}}} G_n^+|\mathcal{F}_\infty\big]$. Notice then that (H) holds with G replaced by G'. To conclude using Corollary 5.15, it is enough to check that (5.3) holds with G replaced by G'.

For $n \geq k$ finite, we have $G'_n \leq \mathbb{E}\big[\sup_{\ell\in[\![k,\infty]\!]} G_\ell\big|\,\mathcal{F}_n\big]$. Since $\mathbb{E}\big[\sup_{\ell\in[\![k,\infty]\!]} G_\ell^+\big]$ is finite thanks to (5.2), we deduce from Lemma 5.31 that:

$$\limsup_n G'_n \leq \limsup_n \mathbb{E}\big[\sup_{\ell\in[\![k,\infty]\!]} G_\ell\big|\,\mathcal{F}_n\big] \leq \mathbb{E}\big[\sup_{\ell\in[\![k,\infty]\!]} G_\ell\big|\,\mathcal{F}_\infty\big].$$

Since k is arbitrary, we get:

$$\limsup_n G'_n \leq \limsup_k \mathbb{E}\big[\sup_{\ell\in[\![k,\infty]\!]} G_\ell\big|\,\mathcal{F}_\infty\big] \leq \mathbb{E}\big[\limsup_k \sup_{\ell\in[\![k,\infty]\!]} G_\ell\big|\,\mathcal{F}_\infty\big] \leq \mathbb{E}[G_\infty|\,\mathcal{F}_\infty] = G'_\infty,$$

where we used Lemma 5.30 (with $X_k = \sup_{\ell\in[\![k,\infty]\!]} G_\ell$ and $Y_k = Y = M$) for the second inequality and (5.3) for the last. Thus (5.3) holds with G replaced by G'. This finishes the proof. □

Exercise 5.7. Let $G = (G_n, n \in \overline{\mathbb{N}})$ be a sequence of random variables such that (5.2) and (5.3) hold. Let $\tau_* = \inf\{n \in \mathbb{N};\, \operatorname{ess\,sup}_{\tau\in\mathbb{T}_n} \mathbb{E}[G_\tau|\mathcal{F}_n] = \mathbb{E}[G_n|\mathcal{F}_n]\}$ with $\inf\emptyset = \infty$. Prove that τ_* is optimal. △

5.3 From finite horizon to infinite horizon

In the finite horizon case, the solution to the optimal equations (5.4) are defined recursively in a constructive way. There is no such constructive way in the infinite horizon case. Thus, it is natural to ask if the infinite horizon case can be seen as the limit of finite horizon cases, when the horizon ζ goes to infinity. We shall give sufficient condition for this to hold in Section 5.3.1 for the adapted case then derive a solution to the castle selling problem of Example 5.2 in Section 5.3.2 and a solution in a Markov chain setting in Section 5.3.3.

5.3.1 From finite horizon to infinite horizon

We assume in this section that the gain sequence $G = (G_n, n \in \overline{\mathbb{N}})$ is adapted to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in \overline{\mathbb{N}})$, with $\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$. We also assume that (5.2), or the weaker Condition (H) page 94, holds. We consider the following assumptions which are stronger than (5.3):

$$\limsup G_n = G_\infty \quad \text{a.s.} \tag{5.19}$$

$$\lim_{n\to\infty} G_n = G_\infty \quad \text{a.s..} \tag{5.20}$$

Remark 5.19. We comment on the conditions (5.19) and (5.20). In particular, (5.20) holds if (5.3) holds and a.s. $G_\infty = -\infty$. We now prove that if (5.3) holds and

$$\lim_{n\to\infty} \mathbb{E}[G_\infty|\mathcal{F}_n] = G_\infty \quad \text{a.s.,} \tag{5.21}$$

then we can modify the gain so that the maximal gain is the same and (5.19) holds for the modified gain. Notice the convergence (5.21) holds in particular if G_∞ is integrable, thanks to Corollary 4.25.

Assume that Condition (H) page 94 holds for G. We consider the gain $G' = (G'_n, n \in \overline{\mathbb{N}})$ with $G'_n = \max(G_n, \mathbb{E}[G_\infty|\mathcal{F}_n])$ which satisfies Condition (H) with $M' = M + G_\infty^+$ as well as (5.19), since (5.21) holds. According to Proposition 5.17, there exists an optimal stopping time, say τ', for the gain G'. The maximal gain is $V'_* = \mathbb{E}[G'_{\tau'}]$. Set $\tau = \tau'$ on $\bigcup_{n\in\mathbb{N}}\{\tau' = n,\, G'_n = G_n\}$ and $\tau = +\infty$ otherwise. Roughly speaking, the stopping rule τ can be described as follows: on $\{\tau' = n\}$, then either $G_n = G'_n$, and then one stops the game at time n to get the gain G_n, or $G_n < G'_n$ and then one never stops the game to get the gain G_∞. Notice τ is a stopping time. We have:

$$\begin{aligned}
\mathbb{E}[G_\tau] &= \sum_{n\in\overline{\mathbb{N}}} \mathbb{E}[G_n \mathbf{1}_{\{\tau=n\}}] \\
&= \sum_{n\in\mathbb{N}} \mathbb{E}[G_n \mathbf{1}_{\{\tau'=n, G_n=G'_n\}}] + \sum_{n\in\mathbb{N}} \mathbb{E}[G_\infty \mathbf{1}_{\{\tau'=n, G_n<G'_n\}}] + \mathbb{E}[G_\infty \mathbf{1}_{\{\tau'=\infty\}}] \\
&= \sum_{n\in\mathbb{N}} \mathbb{E}[G_n \mathbf{1}_{\{\tau'=n, G_n=G'_n\}}] + \sum_{n\in\mathbb{N}} \mathbb{E}[\mathbb{E}[G_\infty|\mathcal{F}_n] \mathbf{1}_{\{\tau'=n, G_n<G'_n\}}] + \mathbb{E}[G_\infty \mathbf{1}_{\{\tau'=\infty\}}] \\
&= \sum_{n\in\mathbb{N}} \mathbb{E}[G'_n \mathbf{1}_{\{\tau'=n\}}] + \mathbb{E}[G_\infty \mathbf{1}_{\{\tau'=\infty\}}] \\
&= \mathbb{E}[G'_{\tau'}].
\end{aligned}$$

As $\mathbb{E}[G_\tau] = \mathbb{E}[G'_{\tau'}]$, we get that $\mathbb{E}[G'_{\tau'}] \leq V_*$. Since $G'_n \geq G_n$ and τ' is optimal, we also get that $\mathbb{E}[G'_{\tau'}] \geq V_*$. We deduce that $V'_* = \mathbb{E}[G'_{\tau'}] = V_* = \mathbb{E}[G_\tau]$, which implies that τ is optimal.

Thus, if (5.21) holds, then (5.19) holds with G' instead of G, and if (H) holds for G, then we can recover an optimal stopping times for G from an optimal stopping times for G', the maximal gain being the same. $\diamondsuit$

Recall $\mathbb{T}_n = \{\tau \in \mathbb{T};\, \tau \geq n\}$ is the set of stopping time equal to or larger than $n \in \mathbb{N}$ and $\mathbb{T}^\zeta = \{\tau \in \mathbb{T};\, \tau \leq \zeta\}$ is the set of stopping times bounded by $\zeta \in \mathbb{N}$. For $\zeta \in \mathbb{N}$ and $n \in [\![0, \zeta]\!]$ we define $\mathbb{T}_n^\zeta = \mathbb{T}_n \bigcap \mathbb{T}^\zeta$ as well as:

$$S_n^\zeta = \operatorname*{ess\,sup}_{\tau\in\mathbb{T}_n^\zeta} \mathbb{E}[G_\tau|\, \mathcal{F}_n]. \tag{5.22}$$

From Sections 5.1.1 and 5.2.3, we get that $S_\zeta^\zeta = G_\zeta$ and $S^\zeta = (S_n^\zeta, n \in [\![0, \zeta]\!])$ satisfies the optimal equations (5.4). For $n \in \mathbb{N}$, the sequence $(S_n^\zeta, \zeta \in [\![n, \infty[\![)$ is non-decreasing and denote by S_n^* its limit. For $n \in \mathbb{N}$, we have a.s. $S_n^* = \operatorname{ess\,sup}_{\tau\in\mathbb{T}_n^{(\mathrm{b})}} \mathbb{E}[G_\tau|\, \mathcal{F}_n]$, where $\mathbb{T}_n^{(\mathrm{b})} = \mathbb{T}_n \bigcap \mathbb{T}^{(\mathrm{b})}$ and $\mathbb{T}^{(\mathrm{b})} \subset \mathbb{T}$ is the subset of bounded stopping times. By construction of S_n, we have for all $n \in \mathbb{N}$:

$$S_n^* \leq S_n, \tag{5.23}$$

The sequence $(\tau_*^\zeta, \zeta \in \mathbb{N})$, with $\tau_*^\zeta = \inf\{n \in [\![0, \zeta]\!];\, S_n^\zeta = G_n\}$, is non-decreasing and thus converge to a limit, say $\tau_*^* \in \overline{\mathbb{N}}$ and

$$\tau_*^* = \lim_{\zeta\to\infty} \tau_*^\zeta = \inf\{n \in \mathbb{N};\, S_n^* = G_n\}. \tag{5.24}$$

Thanks to (5.23) we deduce that a.s. $\tau_*^* \leq \tau_*$. We set $V_*^\zeta = \mathbb{E}[S_0^\zeta] = \sup_{\tau\in\mathbb{T}^\zeta} \mathbb{E}[G_\tau]$ and $V_* = \mathbb{E}[S_0] = \sup_{\tau\in\mathbb{T}} \mathbb{E}[G_\tau]$. Let V_*^* be the non-decreasing limit of the sequence $(V_*^\zeta, \zeta \in \mathbb{N})$, so that $V_*^* \leq V_*$.

Remark 5.20. Assume that (5.2) holds and G_n is integrable for all $n \in \mathbb{N}$. Since $G_n \leq S_n^\zeta \leq \mathbb{E}\big[\sup_{k\in\overline{\mathbb{N}}} G_k^+ | \mathcal{F}_n\big] = M_n$ for all $\zeta \geq n$, using dominated convergence, we deduce from (5.2) that $(S_n^*, n \in \mathbb{N})$ satisfies the optimal equations (5.4) with $\zeta = \infty$. In fact, it is easy to check that $S^* = (S_n^*, n \in \mathbb{N})$ is the smallest sequence satisfying the optimal equations (5.4) with $\zeta = \infty$. Following Remark 5.7, we deduce that S^* is the smallest super-martingale which dominates $(G_n, n \in \mathbb{N})$. And the process $(S^*_{n\wedge\tau_*^*}, n \in \mathbb{N})$ is a martingale, which is a.s. converging thanks to (5.2). $\Diamond$

Definition 5.21. *The infinite horizon case is the limit of the finite horizon cases if $V_*^* = V_*$.*

It is not true in general that $V_*^* = V_*$, see Example 5.22 below taken from Neveu [6].

Example 5.22. Let $(X_n, n \in \mathbb{N}^*)$ be independent random variables such that $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = 1/2$ for all $n \in \mathbb{N}$. Let $c = (c_n, n \in \mathbb{N}^*)$ be a strictly increasing sequence such that $0 < c_n < 1$ for all $n \in \mathbb{N}^*$ and $\lim_{n\to\infty} c_n = 1$. We define $G_0 = 0$, $G_\infty = 0$, and for $n \in \mathbb{N}^*$:

$$G_n = \min\big(1, W_n\big) - c_n,$$

with $W_n = \sum_{k=1}^n X_k$. Notice that $G_n \leq 1$ and a.s. $\limsup G_n = G_\infty$ so that (5.2) and (5.19) hold. (Notice also that $\mathbb{E}[|G_n|]$ for all $n \in \overline{\mathbb{N}}$.) Since $\mathbb{E}[W_{n+1}|\mathcal{F}_n] = W_n$, we deduce from Jensen inequality that a.s. $\mathbb{E}[\min(1, W_{n+1})|\mathcal{F}_n] \geq \min(1, W_n)$. Then use that the sequence c is strictly increasing to get that for all $n \in \mathbb{N}$ a.s. $G_n > \mathbb{E}[G_{n+1}|\mathcal{F}_n]$. Using a backward induction argument and the optimal equations, we get that $S_n^\zeta = G_n$ for all $n \in [\![0, \zeta]\!]$ and $\zeta \in \mathbb{N}$ and thus $\tau_*^\zeta = 0$. We deduce that $S_n^* = G_n$ for all $n \in \mathbb{N}$, $\tau_*^* = 0$ and $V_*^* = 0$.

Since (5.2) and (5.3) hold, we deduce there exists an optimal stopping time for the infinite horizon case. The stopping time $\tau = \inf\{n \in \mathbb{N}^*; W_n = 1\}$ is a.s. strictly positive and finite. On $\{\tau = n\}$, we have that $G_n = 1 - c_n$ as well as $G_m \leq 0 < G_n$ for all $m \in [\![0, n[\![$ and $G_m \leq 1 - c_m < G_n$ for all $m \in]\!]n, \infty]\!]$. We deduce that $G_\tau = \sup_{\tau'\in\mathbb{T}} G_{\tau'}$, that is $\tau = \tau_*$ is optimal. Notice that $V_* > V_*^* = 0$ and a.s. $\tau_* > \tau_*^* = 0$. Thus, the infinite horizon case is not the limit of the finite horizon cases. $\triangle$

We give sufficient conditions so that $V_*^* = V_*$. Recall that (5.2) implies Condition (H).

Proposition 5.23. *Let $(G_n, n \in \mathbb{N})$ be an adapted sequence of $\mathbb{R}$-valued random variables and define G_∞ by (5.19). Assume that (H) holds and that the sequence $(T_n, n \in \mathbb{N})$, with $T_n = \sup_{k\geq n} G_k - G_n$, is uniformly integrable. If there exists an a.s. finite optimal stopping time or if (5.20) holds, then the infinite horizon case is the limit of the finite horizon cases.*

Proof. If $V_* = -\infty$, nothing has to be proven. Let us assume that $V_* > -\infty$. According to Proposition 5.17, there exists an optimal stopping time, say τ. Since $\mathbb{E}[G_{\min(\tau,n)}] \leq V_*^n$, we get:

$$\begin{aligned} 0 \leq V_* - V_*^n \leq \mathbb{E}[G_\tau - G_{\min(\tau,n)}] &= \mathbb{E}\left[\mathbf{1}_{\{n<\tau<\infty\}}(G_\tau - G_n)\right] + \mathbb{E}\left[\mathbf{1}_{\{\tau=\infty\}}(G_\infty - G_n)\right] \\ &\leq \mathbb{E}\left[\mathbf{1}_{\{n<\tau<\infty\}}T_n\right] + \mathbb{E}\left[\mathbf{1}_{\{\tau=\infty\}}(G_\infty - G_n)^+\right]. \end{aligned}$$

As $(T_n, n \in \mathbb{N})$ is uniformly integrable, we deduce from property (iii) of Proposition 7.18 that $(\mathbf{1}_{\{n<\tau<\infty\}}T_n, n \in \mathbb{N})$ is also uniformly integrable. Since a.s. $\lim_{n\to+\infty} \mathbf{1}_{\{n<\tau<\infty\}} = 0$ and thus $\lim_{n\to+\infty} \mathbf{1}_{\{n<\tau<\infty\}}T_n = 0$, we deduce from Proposition 7.21 that this latter convergence holds also in L^1 that is $\lim_{n\to+\infty} \mathbb{E}\left[\mathbf{1}_{\{n<\tau<\infty\}}T_n\right] = 0$.

If τ is a.s. finite, then we have $\mathbb{E}\left[\mathbf{1}_{\{\tau=\infty\}}(G_\infty - G_n)^+\right] = 0$. Otherwise, if (5.20) holds, then the sequence $(\mathbf{1}_{\{\tau=\infty\}}(G_\infty - G_n)^+, n \in \mathbb{N})$ converges a.s. to 0. Since $\mathbf{1}_{\{\tau=\infty\}}(G_\infty - G_n)^+ \leq |T_n|$ and $(T_n, n \in \mathbb{N})$ is uniformly integrable, we deduce from property (iii) of Proposition 7.18 that the sequence $(\mathbf{1}_{\{\tau=\infty\}}(G_\infty - G_n)^+, n \in \mathbb{N})$ is uniformly integrable. Use Proposition 7.21 to get it converges towards 0 in L^1: $\lim_{n\to+\infty} \mathbb{E}\left[\mathbf{1}_{\{\tau=\infty\}}(G_\infty - G_n)^+\right] = 0$. In both cases, we deduce that $\lim_{n\to\infty} V_* - V_*^n = 0$. This gives the result. □

The following exercise complete Proposition 5.23 by giving the convergence of the minimal optimal stopping time in the finite horizon case to τ_* the minimal optimal stopping time in the infinite horizon case defined in (5.5).

Exercise 5.8. Let $(G_n, n \in \mathbb{N})$ be an adapted sequence of random variables taking values in $\mathbb{R}$ and define G_∞ by (5.19). Assume that (H) holds and that the sequence $(T_n, n \in \mathbb{N})$, with $T_n = \sup_{k\geq n} G_k - G_n$, is uniformly integrable. Recall τ_*^* defined in (5.24).

1. If τ_* is a.s. finite, prove that a.s. $S^*_{n\wedge\tau_*} = S_{n\wedge\tau_*}$ for all $n \in \mathbb{N}$ and thus a.s. $\tau_*^* = \tau_*$.

2. If (5.20) holds, prove that $S_n^* = S_n$ for all $n \in \mathbb{N}$ and thus a.s. $\tau_*^* = \tau_*$.

△

We give an immediate Corollary of Proposition 5.23.

Corollary 5.24. *Let $(G_n, n \in \mathbb{N})$ be an adapted sequence of $\mathbb{R}$-valued random variables and define G_∞ by (5.19). Assume that for $n \in \mathbb{N}$ we have $G_n = Z_n - W_n$, with $(Z_n, n \in \mathbb{N})$ adapted, $\mathbb{E}[\sup_{n\in\mathbb{N}} |Z_n|] < +\infty$ and $(W_n, n \in \mathbb{N})$ an adapted non-decreasing sequence of non-negative random variables. If there exists an a.s. finite optimal stopping time or if (5.20) holds, then the infinite horizon case is the limit of the finite horizon cases.*

Proof. For $k \geq n$, we have $G_k - G_n \leq Z_k - Z_n \leq 2\sup_{\ell\in\mathbb{N}} |Z_\ell|$. This gives that the sequence $(T_n = \sup_{k\geq n} G_k - G_n, n \in \mathbb{N})$ is non-negative and bounded by $2\sup_{\ell\in\mathbb{N}} |Z_\ell|$, hence it is uniformly integrable. We conclude using Proposition 5.23. □

Using super-martingale theory, we can prove directly the following result (which is not a direct consequence of the previous Corollary with $W_n = 0$).

Proposition 5.25. *Let $(G_n, n \in \mathbb{N})$ be an adapted sequence of random variables taking values in $\mathbb{R}$ and define G_∞ by (5.19). Assume that $\mathbb{E}[\sup_{n\in\mathbb{N}} |G_n|] < +\infty$. Then the infinite horizon case is the limit of the finite horizon cases. Furthermore, we have that $(S_n, n \in \mathbb{N})$ given by (5.17) is a.s. equal to $(S_n^*, n \in \mathbb{N})$ given by (5.22), and thus the optimal stopping time τ_* defined by (5.5) is a.s. equal to τ_*^* defined by (5.24).*

Proof. According to Remark 5.20, the process $S^* = (S_n^*, n \in \mathbb{N})$ satisfies the optimal equations (5.4) with $\zeta = \infty$. Since it is bounded by $\sup_{n\in\mathbb{N}} |G_n|$ which is integrable, it is a super-martingale and it converges a.s. to a limit say S_∞^*. We have $S_n^* \geq G_n$ for all $n \in \mathbb{N}$, which implies thanks to (5.19) that $S_\infty^* \geq G_\infty$.

Let $n \in \mathbb{N}$. We have for all stopping times $\tau \geq n$ that a.s. $S_n^* \geq \lim_{m\to\infty} \mathbb{E}[S^*_{m\wedge\tau} | \mathcal{F}_n] = \mathbb{E}[S_\tau^* | \mathcal{F}_n]$, where we used the optional stopping theorem for the inequality, and the dominated convergence from property (vi) in Proposition 2.7 (with $Y = \sup_{n\in\mathbb{N}} |G_n|$ and $X_n = S_n^*$) for the equality. This implies that, for all stopping times $\tau \geq n$, a.s. $S_n^* \geq \mathbb{E}[G_\tau|\mathcal{F}_n]$, which thanks to Proposition 5.11 implies that a.s. $S_n^* \geq S_n$. Thanks to (5.23), we get that a.s. $S_n^* \leq S_n$ and thus a.s. $S_n^* = S_n$ for all $n \in \mathbb{N}$. By dominated convergence, we have $V_*^* = \lim_{\zeta\to\infty} \mathbb{E}[S_0^\zeta] = \mathbb{E}[S_0^*] = V_*$. Thus, the infinite horizon case is the limit of the finite horizon cases. Using (5.24), we get that a.s. $\tau_* = \tau_*^*$. □

Exercise 5.9. Extend Proposition 5.25 to the non adapted case. △

5.3.2 Castle to sell

Continuation of Example 5.2. We model the proposal of the n-th buyer of the castle by a random variable X_n. We assume $(X_n, n \in \mathbb{N}^*)$ is a sequence of independent random variables distributed as a random variable X which takes values in $[-\infty, +\infty)$ with $\mathbb{E}[(X^+)^2] < +\infty$ and $\mathbb{P}(X > -\infty) > 0$. We assume each visit of the castle has a fixed cost $c > 0$. We first consider the case, where a previous buyer can be called back, so that the gain at step $n \in \mathbb{N}^*$ is given by $G_n = M_n - nc$, with $M_n = \max_{1\leq k\leq n} X_k$. We set $G_\infty = -\infty$. We consider the σ-field $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ for $n \in \mathbb{N}^*$ and $\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}^*} \mathcal{F}_n$. (Notice that to stick to the presentation of this section, we could set $G_0 = -\infty$ and $\mathcal{F}_0$ the trivial σ-field.)

Notice that $\max(x, y) = (x - y)^+ + y$ for $x \in [-\infty, +\infty)$ and $y \in \mathbb{R}$. In particular, if Y is a $\mathbb{R}$-valued random variable independent of X, we get $\mathbb{E}[\max(X, Y)|Y] = f(Y) + Y$ with $f(x) = \mathbb{E}[(X - x)^+]$. We deduce for $n \in \mathbb{N}^*$ that on $\{M_n > -\infty\}$:

$$\mathbb{E}[G_{n+1}|\mathcal{F}_n] = \mathbb{E}[\max(X_{n+1}, M_n)|M_n] - (n+1)c = f(M_n) - c + G_n. \tag{5.25}$$

We set $x_0 = \sup\{x \in \mathbb{R};\, \mathbb{P}(X \geq x) > 0\}$ and $x_0 \in (-\infty, +\infty]$ as $\mathbb{P}(X > -\infty) > 0$. Since $\mathbb{E}[X^+]$ is finite, we get that the function $f(x) = \mathbb{E}[(X - x)^+]$ is continuous strictly decreasing on $(-\infty, x_0)$ and such that $\lim_{x\to-\infty} f(x) = +\infty$ and $\lim_{x\to x_0} f(x) = 0$. By convention, we set $f(-\infty) = +\infty$. Since a.s. $\lim_{n\to\infty} M_n = x_0$, we get that a.s. $\lim_{n\to\infty} f(M_n) = 0$. Thus the stopping time $\tau = \inf\{n \in \mathbb{N}^*,\, f(M_n) \leq c\}$ is a.s. finite. From the properties of f, we deduce there exists a unique $c_* \in \mathbb{R}$ such that $f(c_*) = c$. Using that $(f(M_n), n \in \mathbb{N}^*)$ is non-increasing and that it jumps at record times of the sequence $(X_n, n \in \mathbb{N}^*)$, we get the representation:

$$\tau = \inf\{n \in \mathbb{N}^*,\, X_n \geq c_*\}.$$

We shall prove that τ is optimal and:

$$\tau_* = \tau \text{ a.s.} \quad \text{and} \quad V_* = \mathbb{E}[G_\tau] = c_*.$$

Since τ is geometric with parameter $\mathbb{P}(X \geq c_*)$, we have $\mathbb{E}[\tau] = 1/\mathbb{P}(X \geq c_*) < +\infty$ and:

$$\mathbb{E}[G_\tau] = \mathbb{E}[X_\tau] - c\mathbb{E}[\tau] = \frac{\mathbb{E}[X\mathbf{1}_{\{X\geq c_*\}}] - c}{\mathbb{P}(X \geq c_*)} = \frac{\mathbb{E}[(X - c_*)^+] - c}{\mathbb{P}(X \geq c_*)} + c_* = c_*,$$

where we used that $\mathbb{E}[(X - c_*)^+] = f(c_*) = c$ for the last equality. Furthermore, for $n \in \mathbb{N}^*$, we deduce from (5.25) that a.s. :

$$\mathbb{E}[G_{n+1}|\mathcal{F}_n] > G_n \quad \text{on} \quad \{n < \tau\} \bigcap \{G_n > -\infty\}, \tag{5.26}$$

$$\mathbb{E}[G_{n+1}|\mathcal{F}_n] \leq G_n \quad \text{on} \quad \{n \geq \tau\}. \tag{5.27}$$

We now state a technical Lemma whose proof is postponed to the end of this section.

Lemma 5.26. *Let X be a random variable taking values in $[-\infty, +\infty)$. Let $(X_n, n \in \mathbb{N}^*)$ be a sequence of random variables distributed as X. Let $c \in]0, +\infty[$. Set $G_n = \max_{1 \le k \le n} X_k - nc$ for $n \in \mathbb{N}^*$. If $\mathbb{E}[(X^+)^2] < +\infty$, then $\mathbb{E}[\sup_{n \in \mathbb{N}^*} G_n^+] < +\infty$ and $\limsup G_n = -\infty$.*

According to Lemma 5.26, we have that (5.2) and (5.3) hold. According to Proposition 5.17, τ_* given by (5.5) is optimal. This implies that $V_* = \mathbb{E}[G_{\tau_*}] \ge \mathbb{E}[G_\tau] > -\infty$ and since a.s. $G_\infty = -\infty$, we get that τ_* is finite. We deduce also from (5.26) that a.s. $\tau_* \ge \tau$.

We have with $c' = c/2$:

$$-\infty < \mathbb{E}[G_{\tau_*}] = \mathbb{E}[\max_{k \in [\![1, \tau_*]\!]} X_k - \tau_* c'] - \mathbb{E}[\tau_* c'] \le \mathbb{E}[\sup_{n \in \mathbb{N}^*} (\max_{1 \le k \le n} X_k - nc')^+] - \mathbb{E}[\tau_*]c'.$$

Using Lemma 5.26 with c replaced by c', we get that $\mathbb{E}[\sup_{n \in \mathbb{N}^*} (\max_{1 \le k \le n} X_k - nc')^+]$ is finite and thus $\mathbb{E}[\tau_*]$ is finite. Let $n \in \mathbb{N}^*$. On $\{\tau = n\}$, we have for finite $k \ge n$ that $G_\tau - \tau_* c \le G_{\tau_* \wedge k} \le \sup_{n \in \mathbb{N}} G_n^+$ and thus a.s.:

$$\mathbf{1}_{\{\tau = n\}} \mathbb{E}[\sup_{k \ge n} |G_{\tau_* \wedge k}| \,|\, \mathcal{F}_n] < +\infty. \tag{5.28}$$

Mimicking (5.14) with G instead of S and using that $\tau_* \ge \tau$, we deduce from (5.27) that a.s., on $\{\tau = n\}$, $\mathbb{E}[G_{\tau_* \wedge k} | \mathcal{F}_n] \le G_n$ for all finite $k \ge n$. Letting k goes to infinity, since τ_* is a.s. finite, we deduce by dominated convergence, using (5.28), that $\mathbb{E}[G_{\tau_*} | \mathcal{F}_n] \le G_n$ a.s. on $\{\tau = n\}$. Since τ is finite, this gives $\mathbb{E}[G_{\tau_*}] \le \mathbb{E}[G_\tau]$. Since τ_* is optimal, we deduce that τ is also optimal. This gives $V_* = \mathbb{E}[G_\tau] = c_*$. Notice also that a.s. $\tau = \tau_*$ as τ_* is the minimal optimal stopping time according to Exercise 5.4.

If one can not call back a previous buyer, then the gain is $G''_n = X_n - nc$. Let V''_* be the corresponding maximal gain. On the one hand, since $G''_n \le G_n$ for all $n \in \mathbb{N}$, we deduce that $V''_* \le V_*$. On the other hand, we have $G''_\tau = G_\tau = G_{\tau_*}$. This implies that $V''_* \ge \mathbb{E}[G''_\tau] = \mathbb{E}[G_\tau] = V_*$. We deduce that $V''_* = c_*$ and τ is also optimal in this case.

In this last part, we assume furthermore that $\mathbb{E}[|X|] < +\infty$. We shall prove directly, as Corollary 5.24 can not be used here, that the infinite horizon case is the limit of the finite horizon cases. We first consider the case where previous buyers can be called back, so the gain is $G_n = \max_{1 \le k \le n} X_k - nc$ for $n \in \mathbb{N}^*$. For $n \in \mathbb{N}^*$, we have a.s. that $X_1 - \tau c \le G_{\tau \wedge n} \le \sup_{n \in \overline{\mathbb{N}}} G_n^+$. By dominated convergence, we get that $\lim_{n \to \infty} \mathbb{E}[G_{\tau \wedge n}] = \mathbb{E}[G_\tau] = V_*$. We deduce that $V_*^* \ge \lim_{n \to \infty} \mathbb{E}[G_{\tau \wedge n}] = V_*$ and $V_*^* = V_*$ as $V_* \ge V_*^*$. Therefore the infinite horizon case is the limit of the finite horizon cases. (Notice that if $1 > \mathbb{P}(X = -\infty) > 0$, then the infinite horizon case is no more the limit of the finite horizon cases as $V_*^n = -\infty$ for all $n \in \mathbb{N}^*$.)

We now consider the case where previous buyers can not be called back, so the gain is $G''_n = X_n - nc$ for $n \in \mathbb{N}^*$. Let $V''_* = V_*$ (resp. V''^n_*) denote the maximal gain when the horizon is infinite (resp. equal to n). We have:

$$0 \le V''_* - V''^n_* \le \mathbb{E}[G''_{\tau_*} - G''_{\tau_* \wedge n}] \le \mathbb{E}\left[\mathbf{1}_{\{n < \tau_* < \infty\}}(X_{\tau_*} - X_n)\right] = \mathbb{E}\left[\mathbf{1}_{\{n < \tau_* < \infty\}}(X_{\tau_*} - X_1)\right],$$

where we used that $G''_{\tau_*} - G''_n \le X_{\tau_*} - X_n$ on $\{n < \tau_*\}$ for the second inequality and that conditionally on $\{n < \tau_* < \infty\}$, (X_{τ_*}, X_n) and (X_{τ_*}, X_1) have the same distribution for the last equality. Since X_* and X_1 are integrable, we get that $\lim_{n \to +\infty} \mathbb{E}\left[\mathbf{1}_{\{n < \tau_* < \infty\}}(X_{\tau_*} - X_1)\right] = 0$ by dominated convergence. We deduce that the infinite horizon case is the limit of the finite horizon cases.

Proof of Lemma 5.26. Assume that $\mathbb{E}[(X^+)^2] < +\infty$. Since $X_n - nc \leq G_n \leq \max_{1\leq k \leq n}(X_k - kc)$ for all $n \in \mathbb{N}^*$, we deduce that $\sup_{n\in\mathbb{N}^*} G_n = \sup_{n\in\mathbb{N}^*}(X_n - nc)$. This gives:

$$\mathbb{E}\Big[\sup_{n\in\mathbb{N}^*} G_n^+\Big] = \mathbb{E}\Big[\sup_{n\in\mathbb{N}^*}(X_n - nc)^+\Big] \leq \mathbb{E}\Big[\sum_{n\in\mathbb{N}^*}(X_n - nc)^+\Big] = \mathbb{E}\Big[\sum_{n\in\mathbb{N}^*}(X - nc)^+\Big],$$

where we used Fubini (twice) and that X_n is distributed as X in the last equality. Then use that for $x \in \mathbb{R}$:

$$\sum_{n\in\mathbb{N}^*}(x-n)^+ \leq \sum_{n\in\mathbb{N}^*} x^+ \mathbf{1}_{\{n<x^+\}} \leq (x^+)^2,$$

to get $\mathbb{E}\Big[\sum_{n\in\mathbb{N}^*}(X - nc)^+\Big] \leq \mathbb{E}[(X^+)^2]/c < +\infty$. So we obtain $\mathbb{E}\Big[\sup_{n\in\mathbb{N}^*} G_n^+\Big] < +\infty$.

Set $G'_n = \max_{1\leq k\leq n} X_k - nc/2$. Using the previous result (with c replaced by $c/2$), we deduce that $\sup_{n\in\mathbb{N}^*}(G'_n)^+$ is integrable and thus a.s. $\limsup G'_n < +\infty$. Since $G_n = G'_n - nc/2$, we get that a.s. $\limsup G_n \leq \limsup G'_n - \lim nc/2 = -\infty$. □

With the notation of Lemma 5.26, one can prove that if the random variables $(X_n, n \in \mathbb{N}^*)$ are independent then $\mathbb{E}[\sup_{n\in\mathbb{N}^*} G_n^+] < +\infty$ implies that $\mathbb{E}[(X^+)^2] < +\infty$.

5.3.3 The Markovian case

Let $(\mathcal{F}_n, n \in \mathbb{N})$ be a filtration. Recall $\mathbb{T}$ is the set of stopping times and $\mathbb{T}^\zeta$ is the set of stopping times bounded by $\zeta \in \overline{\mathbb{N}}$. Let $(X_n, n \in \mathbb{N})$ be a Markov chain with state space E (at most countable) and transition kernel P. Let φ be a non-negative function defined on E. We shall consider the optimal stopping problem for the game with gain $G_n = \varphi(X_n)$ for $n \in \mathbb{N}$ and $G_\infty = \limsup G_n$ with horizon $\zeta \in \overline{\mathbb{N}}$. We set:

$$\varphi_0 = \varphi \quad \text{and} \quad \varphi_{n+1} = \max(\varphi, P\varphi_n) \quad \text{for } n \in \mathbb{N}.$$

We have the following result for the finite horizon case.

Lemma 5.27. *Let $\zeta \in \mathbb{N}$, $x \in E$ and φ a non-negative function defined on E. Assume that $\mathbb{E}_x[\varphi(X_n)] < +\infty$ for all $n \in [\![0, \zeta]\!]$. Then, we have:*

$$\varphi_\zeta(x) = \sup_{\tau\in\mathbb{T}^\zeta} \mathbb{E}_x[\varphi(X_\tau)] = \mathbb{E}_x[\varphi(X_{\tau_*^\zeta})],$$

$$\textit{with} \quad \tau_*^\zeta = \inf\{n \in [\![0,\zeta]\!];\ X_n \in \{\varphi = \varphi_{\zeta-n}\}\}.$$

Proof. We keep notations from Section 5.3. Recall definition (5.22) of S_n^ζ for the finite horizon $\zeta \in \mathbb{N}$. We deduce from Proposition 5.6 and $S_\zeta^\zeta = G_\zeta$ that $S_n^\zeta = \varphi_{\zeta-n}(X_n)$ for all $0 \leq n \leq \zeta$ and that the optimal stopping time is $\tau_*^\zeta = \inf\{n \in [\![0,\zeta]\!];\ \varphi_{\zeta-n}(X_n) = \varphi(X_n)\}$. □

We give a technical lemma.

Lemma 5.28. *The sequence of functions $(\varphi_n, n \in \mathbb{N})$ is non-decreasing and converges to a limit say φ_* such that $\varphi_* = \max(\varphi, P\varphi_*)$. For any non-negative function g such that $g \geq \max(\varphi, Pg)$, we have that $g \geq \varphi_*$.*

Proof. By an elementary induction argument, we get that the sequence $(\varphi_n, n \in \mathbb{N})$ is non-decreasing. Let φ_* be its limit. By monotone convergence, we get that $\varphi_* = \max(\varphi, P\varphi_*)$. Let g be a non-negative function g such that $g \geq \max(\varphi, Pg)$, we have by induction that $g \geq \varphi_n$ and thus $g \geq \varphi_*$. $\square$

We now give the main result of this section on the infinite horizon case. Recall $\mathbb{T}^{(b)}$ is the set of bounded stopping times.

Proposition 5.29. *Let $x \in E$ and φ a non-negative function defined on E. Assume that $\mathbb{E}_x[\sup_{n\in\mathbb{N}} \varphi(X_n)] < +\infty$. Then, the maximal gain under $\mathbb{P}_x$ is given by:*

$$\varphi_*(x) = \sup_{\tau\in\mathbb{T}^{(b)}} \mathbb{E}_x[\varphi(X_\tau)] = \sup_{\tau\in\mathbb{T}} \mathbb{E}_x[\varphi(X_\tau)] = \mathbb{E}[\varphi(X_{\tau_*})],$$

with the optimal stopping time:

$$\tau_* = \inf\{n \in \mathbb{N};\, X_n \in \{\varphi = \varphi_*\}\}, \tag{5.29}$$

and the conventions $\inf\emptyset = \infty$ *and* $\varphi(X_\infty) = \limsup \varphi(X_n)$. *Furthermore, the infinite horizon case is the limit of the finite horizon case and a.s. $\tau_* = \tau_*^*$.*

Proof. We keep notations from the proof of Lemma 5.27. Lemma 5.28 implies that $S_n^* = \lim_{\zeta\to\infty} S_n^\zeta = \varphi_*(X_n)$. Recall that by definition $\tau_*^* = \lim_{\zeta\to\infty} \tau_*^\zeta$. According to Proposition 5.25, the infinite horizon case is the limit of the finite horizon cases and the optimal stopping time τ_* is given by (5.5) that is by (5.29) with the conventions $\inf\emptyset = \infty$ and $\varphi(X_\infty) = \limsup \varphi(X_n)$. We also get it is a.s. equal to τ_*^* and that $V_* = \mathbb{E}[G_{\tau_*}] = \mathbb{E}_x[S_0^*] = \varphi_*(x)$. $\square$

Exercise 5.10. Let $x \in E$ and φ a non-negative function defined on E. Assume that $\mathbb{E}_x[\sup_{n\in\mathbb{N}} \varphi(X_n)] < +\infty$ and consider the gain sequence $(G_n, n \in \overline{\mathbb{N}})$ with $G_n = \varphi(X_n)$ and $G_\infty = \limsup G_n$. Recall the minimal stopping time τ_* defined by (5.5) or equivalently (5.29) and the maximal stopping time τ_{**} defined by (5.6). Prove that:

$$\tau_* = \inf\{n \in \mathbb{N};\, X_n \in \{\varphi \geq P\varphi_*\}\} \quad \text{and} \quad \tau_{**} = \inf\{n \in \mathbb{N};\, X_n \in \{\varphi > P\varphi_*\}\},$$

with the convention $\inf\emptyset = \infty$. $\triangle$

5.4 Appendix

We give in this section some technical Lemmas related to integration. Let $(\Omega, \mathbb{P}, \mathcal{F})$ be a probability space.

Lemma 5.30. *Let X and $(X_n, n \in \mathbb{N})$ be real-valued random variables. Let Y and $(Y_n, n \in \mathbb{N})$ be non-negative integrable random variables. Assume that a.s. $X_n^+ \leq Y_n$ for all $n \in \mathbb{N}$, $\lim_{n\to\infty} Y_n = Y$ and $\lim_{n\to\infty} \mathbb{E}[Y_n|\mathcal{H}] = \mathbb{E}[Y|\mathcal{H}]$, where $\mathcal{H} \subset \mathcal{F}$ is a σ-field. Then we have that a.s.:*

$$\limsup_{n\to\infty} \mathbb{E}[X_n|\mathcal{H}] \leq \mathbb{E}\left[\limsup_{n\to\infty} X_n \middle| \mathcal{H}\right].$$

Proof. By Fatou Lemma, we get $\liminf_{n\to\infty} \mathbb{E}\left[X_n^-|\mathcal{H}\right] \geq \mathbb{E}\left[\liminf_{n\to\infty} X_n^-|\mathcal{H}\right]$. We also have:

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{E}\left[X_n^+|\,\mathcal{H}\right] &= -\liminf_{n\to\infty} \mathbb{E}\left[Y_n - X_n^+|\,\mathcal{H}\right] - \lim_{n\to\infty} \mathbb{E}\left[Y_n|\,\mathcal{H}\right] \\
&\leq -\mathbb{E}\left[\liminf_{n\to\infty}(Y_n - X_n^+)|\,\mathcal{H}\right] - \mathbb{E}\left[Y|\,\mathcal{H}\right] \\
&= \mathbb{E}\left[\limsup_{n\to\infty} X_n^+|\,\mathcal{H}\right] - \mathbb{E}\left[\lim_{n\to\infty} Y_n|\,\mathcal{H}\right] - \mathbb{E}\left[Y|\,\mathcal{H}\right] = \mathbb{E}\left[\limsup_{n\to\infty} X_n^+|\,\mathcal{H}\right],
\end{aligned}$$

where we used Fatou lemma for the inequality. To conclude, use that a.s.: :

$$\limsup_{n\to\infty} \mathbb{E}[X_n|\mathcal{H}] \leq \limsup_{n\to\infty} \mathbb{E}[X_n^+|\mathcal{H}] - \liminf_{n\to\infty} \mathbb{E}[X_n^-|\mathcal{H}]$$

and $\limsup_{n\to\infty} X_n^+ - \liminf_{n\to\infty} X_n^- = \limsup_{n\to\infty} X_n$. □

Let $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$, with $\mathcal{F}_n \subset \mathcal{F}$, be a filtration. We set $\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$ the smallest possible σ-field which contains $\bigcup_{n\in\mathbb{N}} \mathcal{F}_n$.

Lemma 5.31. *Let M be random variable taking values in $[-\infty, +\infty)$ such that $\mathbb{E}[M^+] < +\infty$. Let $M_n = \mathbb{E}[M|\,\mathcal{F}_n]$ for $n \in \overline{\mathbb{N}}$. Then, we have that a.s. $\limsup M_n \leq M_\infty$.*

Proof. Let $a \in \mathbb{R}$. By Jensen inequality, we have that $M_n \vee a \leq \mathbb{E}[M \vee a|\,\mathcal{F}_n]$. According to Corollary 4.25, we have a.s. $\limsup M_n \leq \limsup M_n \vee a \leq \mathbb{E}[M \vee a|\,\mathcal{F}_\infty]$. By monotone convergence, we deduce by letting a goes to $-\infty$, that $\limsup M_n \leq \mathbb{E}[M|\,\mathcal{F}_\infty] = M_\infty$. □

Bibliography

[1] Y. S. Chow, H. Robbins, and D. Siegmund. *Great expectations: the theory of optimal stopping.* Houghton Mifflin Co., Boston, Mass., 1971.

[2] E. B. Dynkin. Optimal choice of the stopping moment of a Markov process. *Dokl. Akad. Nauk SSSR*, 150:238–240, 1963.

[3] T. Ferguson. Optimal stopping and applications. `http://www.math.ucla.edu/~tom/Stopping/Contents.html`.

[4] T. S. Ferguson. Who solved the secretary problem? *Statist. Sci.*, 4(3):282–289, 08 1989.

[5] S. A. Lippman and J. J. McCall. Job search in a dynamic economy. *J. Econom. Theory*, 12(3):365–390, 1976.

[6] J. Neveu. *Discrete-parameter martingales.* North-Holland Publishing Co., New York, revised edition, 1975.

[7] J. L. Snell. Applications of martingale system theorems. *Trans. Amer. Math. Soc.*, 73:293–312, 1952.

Chapter 6

Brownian motion

6.1 Gaussian process

6.1.1 Gaussian vector

We recall that X is a Gaussian (or normal) random variable if it is a real-valued random variable whose distribution has density f_{m,σ^2} with respect to the Lebesgue measure on $\mathbb{R}$ given by:

$$f_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, \mathrm{e}^{-(x-m)^2/(2\sigma^2)} \quad \text{for } x \in \mathbb{R},$$

with parameters $m \in \mathbb{R}$ and $\sigma > 0$. The random variable X is square integrable and the parameter m is the mean of X and σ^2 its variance. The law of X is often denoted by $\mathcal{N}(m, \sigma^2)$. By convention, the constant $m \in \mathbb{R}$ will also be considered as a (degenerate) Gaussian random variable with $\sigma^2 = 0$ and we shall denote its distribution by $\mathcal{N}(m, 0)$. The characteristic function ψ_{m,σ^2} of X with distribution $\mathcal{N}(m, \sigma^2)$ is given by:

$$\psi_{m,\sigma^2}(u) = \mathbb{E}\left[\mathrm{e}^{iuX}\right] = \exp\left(ium - \frac{1}{2}\,\sigma^2 u^2\right) \quad \text{for } u \in \mathbb{R}. \tag{6.1}$$

In the next definition we recall the extension of the Gaussian distribution in higher dimension. We recall that a matrix $\Sigma \in \mathbb{R}^{d\times d}$, with $d \geq 1$, is positive semi-definite if it is symmetric and $\langle u, \Sigma u\rangle \geq 0$ for all $u \in \mathbb{R}^d$, where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product on $\mathbb{R}^d$.

Definition 6.1. *Let $d \geq 1$. Let $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ be a positive semi-definite matrix. A $\mathbb{R}^d$-valued random variable X has Gaussian distribution $\mathcal{N}(\mu, \Sigma)$ if its characteristic function $\psi_{\mu,\Sigma}$ is given by:*

$$\psi_{\mu,\Sigma}(u) = \mathbb{E}\left[\mathrm{e}^{i\langle u, X\rangle}\right] = \exp\left(i\langle u, \mu\rangle - \frac{1}{2}\langle u, \Sigma u\rangle\right) \quad \textit{for } u \in \mathbb{R}^d. \tag{6.2}$$

If X is a Gaussian random variable with distribution $\mathcal{N}(\mu, \Sigma)$, then X is square integrable with mean $\mathbb{E}[X] = \mu$ and covariance matrix (see Definition 1.61) $\mathrm{Cov}(X, X) = \Sigma$. Furthermore using the development of the exponential function in series, we get that for all $\lambda \in \mathbb{C}^d$, the random variable $\mathrm{e}^{\langle \lambda, X\rangle}$ is integrable, and we have:

$$\mathbb{E}\left[\mathrm{e}^{\langle \lambda, X\rangle}\right] = \exp\left(\langle \lambda, \mu\rangle + \frac{1}{2}\langle \lambda, \Sigma\lambda\rangle\right). \tag{6.3}$$

Using (6.2) with u replaced by $M^t u$, we deduce the following lemma, which asserts that every affine transformation of a Gaussian random variable is still a Gaussian random variable.

Lemma 6.2. *Let $p, d \in \mathbb{N}^*$. Let X be a $\mathbb{R}^d$-valued Gaussian random variable with distribution $\mathcal{N}(\mu, \Sigma)$. Let $M \in \mathbb{R}^{p\times d}$ and $c \in \mathbb{R}^p$. Then $Y = c + MX$ is a $\mathbb{R}^p$-valued Gaussian random variable with parameter $\mathbb{E}[Y] = c + M\mu$ and $\mathrm{Cov}(Y, Y) = M\Sigma M^\top$.*

The next remark ensures that for all $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ a positive semi-definite matrix, the distribution $\mathcal{N}(\mu, \Sigma)$ is meaningful.

Remark 6.3. Let $d \geq 1$. Let $(G_1, \ldots, G_d)$ be independent real-valued Gaussian random variables with the same distribution $\mathcal{N}(0, 1)$. Using (6.2), we get that the random vector $G = (G_1, \ldots, G_d)$ is Gaussian with distribution $\mathcal{N}(0, I_d)$ and $I_d \in \mathbb{R}^{d\times d}$ the identity matrix.

Let $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ a positive semi-definite matrix. There exists an orthogonal matrix $O \in \mathbb{R}^{d\times d}$ (that is $O^\top O = OO^\top = I_d$) and a diagonal matrix $\Delta \in \mathbb{R}^{d\times d}$ with non-negative entries such that $\Sigma = O\Delta^2 O^\top$. According to Lemma 6.2, we get that $\mu + O\Delta G$ has distribution $\mathcal{N}(\mu, \Sigma)$. $\diamondsuit$

We have the following result on the convergence in distribution of Gaussian vectors.

Lemma 6.4. *Let $d \geq 1$. The family of Gaussian probability distributions $\{\mathcal{N}(\mu, \Sigma);\ \mu \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d\times d}$ positive semi-definite$\}$ is closed for the convergence in distribution. Furthermore, if $(X_n, n \in \mathbb{N})$ are Gaussian random variables on $\mathbb{R}^d$, then the sequence $(X_n, n \in \mathbb{N})$ converges in distribution towards a limit, say X, if and only if X is a Gaussian random variable, $(\mathbb{E}[X_n], n \in \mathbb{N})$ and $(\mathrm{Cov}(X_n, X_n), n \in \mathbb{N})$ converge respectively towards $\mathbb{E}[X]$ and $\mathrm{Cov}(X, X)$.*

Proof. We consider the one-dimensional case $d = 1$. (The general case $d \geq 1$ which is proved similarly is left to the reader.) Let $(X_n, n \in \mathbb{N})$ be a sequence of real-valued Gaussian random variables which converges in distribution towards a limit, say X. Let $m_n = \mathbb{E}[X_n]$ and $\sigma_n^2 = \mathrm{Var}(X_n)$, and ψ_n be the characteristic function of X_n. As the sequence of functions $(\psi_n, n \in \mathbb{N})$ converges pointwise towards ψ, the characteristic function of X, we get that $\lim_{n\to+\infty} |\psi_n(u)| = |\psi(u)|$. This and (6.1) readily implies that the sequence $(\sigma_n, n \in \mathbb{N})$ converges to a limit $\sigma \in [0, +\infty]$. Use that $\psi(0) = 1$ to deduce that σ is finite.

We shall now prove that the sequence $(m_n, n \in \mathbb{N})$ converges. We deduce from the first part of the proof that $(\mathrm{e}^{ium_n} = \mathrm{e}^{u^2\sigma_n^2/2}\, \psi_n(u),\ n \in \mathbb{N})$ converges pointwise towards $\varphi(u) = \mathrm{e}^{u^2\sigma^2/2}\, \psi(u)$. Notice that φ is continuous on $\mathbb{R}$ and that $|\varphi(u)| = 1$ for $u \in \mathbb{R}$. Let G be Gaussian random variable with distribution $\mathcal{N}(0, 1)$. By dominated convergence, we get that for all $x \in \mathbb{R}$ and $a > 0$:

$$\mathrm{e}^{-(m_n-x)^2a^2/2} = \mathbb{E}\left[\mathrm{e}^{i(m_n-x)aG}\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[\mathrm{e}^{-ixaG}\, \varphi(aG)\right]. \tag{6.4}$$

This implies that the sequence $(m_n, n \in \mathbb{N})$ either converges in $\mathbb{R}$ or $\lim_{n\to\infty} |m_n| = +\infty$. In the latter case, we deduce from (6.4) that $\mathbb{E}\left[\mathrm{e}^{-ixaG}\, \varphi(aG)\right] = 0$ for all $a > 0$. Letting a goes to 0, we deduce by dominated convergence, as $|\varphi| = 1$, from the continuity of φ at 0 that $\varphi(0) = 0$ which is a contradiction. Thus the sequence $(m_n, n \in \mathbb{N})$ converges to a limit $m \in \mathbb{R}$. We deduce that, for all $u \in \mathbb{R}$, $\psi_n(u)$ converges towards $\mathrm{e}^{ium-\sigma^2u^2/2}$ which is thus equal to $\psi(u)$. We deduce from (6.1) that X has distribution $\mathcal{N}(m, \sigma^2)$.

We have proved that if the sequence $(X_n, n \in \mathbb{N})$ of real-valued Gaussian random variables converges in distribution towards X, then X is a Gaussian random variable and $(\mathbb{E}[X_n], n \in \mathbb{N})$ as well as $(\mathrm{Cov}(X_n, X_n), n \in \mathbb{N})$ converge respectively towards $\mathbb{E}[X]$ and $\mathrm{Cov}(X, X)$. The converse is a direct consequence of (6.1). $\square$

We give in the next remark an alternative characterization for Gaussian vectors.

Remark 6.5. Let $d \geq 1$ and X a $\mathbb{R}^d$-valued random variable. If $\langle u, X\rangle$ is Gaussian for all $u \in \mathbb{R}^d$, then X has a Gaussian distribution.

Indeed, since $\langle u, X\rangle$ is square integrable for all $u \in \mathbb{R}^d$, we deduce that X is square integrable. Let $\mu = \mathbb{E}[X]$ and $\Sigma = \mathrm{Cov}(X, X)$. Notice that Σ is positive semi-definite as $\langle u, \Sigma u\rangle = \mathrm{Var}(\langle u, X\rangle) \geq 0$ for $u \in \mathbb{R}^d$. Since $\langle u, X\rangle$ is Gaussian with mean $\langle \mu, \mathbb{E}[X]\rangle$ and variance $\mathrm{Var}(\langle u, X\rangle)$, its distribution is $\mathcal{N}(\langle u, \mu\rangle, \langle u, \Sigma u\rangle)$. We deduce from (6.1) (with u and X replaced respectively by 1 and $\langle u, X\rangle$) that (6.2) holds. Thus, by Definition 6.1, X is a Gaussian random vector with distribution $\mathcal{N}(\mu, \Sigma)$. $\lozenge$

It is easy to characterize the independence for Gaussian vectors.

Lemma 6.6. *Let $d \geq 1$, $p \geq 1$, X be a $\mathbb{R}^d$-valued random variable and Y be a $\mathbb{R}^p$-valued random variable. Assume that (X, Y) has a Gaussian distribution. Then X and Y are independent if and only if* $\mathrm{Cov}(X, Y) = 0$.

Proof. Since $\mathrm{Cov}(X, Y) = 0$, we get, with $W = (X, Y)$, that:

$$\mathrm{Cov}(W, W) = \begin{pmatrix} \mathrm{Cov}(X, X) & 0 \\ 0 & \mathrm{Cov}(Y, Y) \end{pmatrix},$$

and thus for all $w = (u, v) \in \mathbb{R}^{d+p}$:

$$\langle w, \mathrm{Cov}(W, W)w\rangle = \langle u, \mathrm{Cov}(X, X)u\rangle + \langle v, \mathrm{Cov}(Y, Y)v\rangle.$$

Using (6.2) (three times), we get that $= \mathbb{E}[\mathrm{e}^{i\langle w, W\rangle}] = \mathbb{E}[\mathrm{e}^{i\langle u, X\rangle}]\mathbb{E}[\mathrm{e}^{i\langle v, Y\rangle}]$ for all $w = (u, v) \in \mathbb{R}^{d+p}$. Since the characteristic function characterizes the distribution of $\mathbb{R}^q$-valued random variables for $q \in \mathbb{N}^*$, we deduce that (X, Y) has the same distribution as (X', Y') where X' and Y' are independent and respectively distributed as X and Y. This implies that X and Y are independent.

The converse is immediate. $\square$

6.1.2 Gaussian process and Brownian motion

We refer to [6, 4, 5] for a general theory of Gaussian processes. The next definition gives an extension of Gaussian vectors to processes.

Definition 6.7. *Let T be a set. Consider the Borel σ-field $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$, and the product space $E = \mathbb{R}^T$ with the corresponding product σ-field $\mathcal{E} = \mathcal{B}(\mathbb{R})^{\otimes T}$. We say a measurable E-valued random variable $X = (X_t, t \in T)$ is a Gaussian process indexed by T if for all finite set $J \subset T$, the vector $(X_t, t \in J)$ is Gaussian. In this case the mean process m is given by $m = (m(t) = \mathbb{E}[X_t];\ t \in T)$ and the covariance kernel K is given by $K = (K(s, t) = \mathrm{Cov}(X_s, X_t);\ s, t \in T)$.*

Lemma 6.8. *The distribution of a Gaussian process is characterized by its mean process and covariance kernel.*

Proof. Let $X = (X_t, t \in T)$ be a Gaussian process indexed by T with mean process m and covariance kernel K. For all $J \subset T$ finite, the vector $(X_t, t \in J)$ is Gaussian and its distribution is characterized by its mean $(m(t), t \in J)$ and its covariance $(K(s,t);\, s,t \in J)$, hence by m and K. Then use Lemma 1.29 to get that the distribution of X is characterized by m and K. □

Let $K = (K(s,t);\, s,t \in T)$ be the covariance kernel of a Gaussian process indexed by T. From the proof of Lemma 6.8, we get that $(K(s,t);\, s,t \in J)$ is a positive semi-definite matrix for all finite $J \subset T$. We deduce that the covariance kernel K is a positive semi-definite function, that is $K(s,t) = K(t,s)$ for all $s,t \in T$ and for all finite set $J \subset T$ and all $\mathbb{R}$-valued vector $(a_t, t \in J)$, we have:

$$\sum_{s,t \in J} a_s a_t\, K(s,t) \geq 0.$$

We admit the converse, see Corollary 3.5 in [6].

Theorem 6.9. *Let T be a set, m a real-valued function defined on T and K a positive semi-definite function defined on T. Then there exist a probability space and a Gaussian process defined on this probability space with mean process m and covariance kernel K.*

One very interesting Gaussian process is the so called Brownian motion. We first give its covariance kernel.

Lemma 6.10. *The function $K = (K(s,t);\, s,t \in \mathbb{R}_+)$ defined by $K(s,t) = s \wedge t$ is a covariance kernel on $\mathbb{R}_+$.*

Proof. Let λ be the Lebesgue measure on $\mathbb{R}_+$. We recall that $\langle f, g\rangle = \int fg\, \mathrm{d}\lambda$ defines a scalar product on $L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \lambda)$. Set $f_t = \mathbf{1}_{[0,t]}$ for $t \in \mathbb{R}_+$, and notice that $K(s,t) = \langle f_s, f_t \rangle$ for all $s,t \in \mathbb{R}_+$. The function K is clearly symmetric and for all $n \in \mathbb{N}^*$, $t_1, \ldots, t_n \in \mathbb{R}_+$, $a_1, \ldots, a_n \in \mathbb{R}$, we have:

$$\sum_{1 \leq i,j \leq n} a_i a_j\, K(t_i, t_j) = \int \Big(\sum_{1 \leq i \leq n} a_i f_{t_i} \Big)^2 \mathrm{d}\lambda \geq 0.$$

Thus the function K is positive semi-definite. □

The existence of the Brownian motion, see below, is justified by Theorem 6.9 and Lemma 6.10. We say a Gaussian process is centered if its mean function is constant equal to 0.

Definition 6.11. *A standard Brownian motion $B = (B_t, t \in \mathbb{R}_+)$ defined a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a centered Gaussian process with covariance kernel K given in Lemma 6.10. A Brownian motion with drift $b \in \mathbb{R}$ and diffusion coefficient $\sigma \in \mathbb{R}_+^*$ is distributed as $(bt + \sigma B_t,\, t \in \mathbb{R}_+)$.*

We derive some elementary computations for a standard Brownian motion $B = (B_t,\, t \in \mathbb{R}_+)$. For $t \geq s \geq u \geq 0$, we have:

$$\operatorname{Var}(B_t - B_s) = \operatorname{Var}(B_t) + \operatorname{Var}(B_s) - 2\operatorname{Cov}(B_t, B_s) = t - s = \operatorname{Var}(B_{t-s}), \tag{6.5}$$

$$\operatorname{Cov}(B_t - B_s, B_u) = \operatorname{Cov}(B_t, B_u) - \operatorname{Cov}(B_s, B_u) = 0. \tag{6.6}$$

6.2 Properties of Brownian motion

We refer to [3, 7] for a general presentation of Brownian motion.

6.2.1 Continuity

There is a technical difficulty when one says the Brownian motion is a.s. continuous, because one sees the Brownian motion as a $\mathbb{R}^{\mathbb{R}_+}$-valued random variable and one can prove that the set of continuous functions is not measurable with respect to σ-field $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}$ on $\mathbb{R}^{\mathbb{R}_+}$. For this reason, we shall consider directly the set of continuous functions.

For an interval $I \subset \mathbb{R}_+$, let $\mathcal{C}^0(I) = \mathcal{C}^0(I, \mathbb{R})$ be the set of $\mathbb{R}$-valued continuous functions defined on I. We define the uniform norm $\|\cdot\|_\infty$ on $\mathcal{C}^0(I)$ as $\|f\|_\infty = \sup_{x \in I} |f(x)|$ for $f \in \mathcal{C}^0(I)$. It is easy to check that $(\mathcal{C}^0(I), \|\cdot\|_\infty)$ is a Banach space. And we denote by $\mathcal{B}(\mathcal{C}^0(I))$ the corresponding Borel σ-field. Notice that $\mathcal{C}^0(I)$ is a subset of $\mathbb{R}^I$ (but it does not belong to $\mathcal{B}(\mathbb{R})^{\otimes I}$). We consider $\mathcal{C}^0(I) \cap \mathcal{B}(\mathbb{R})^{\otimes I} = \{\mathcal{C}^0(I) \bigcap A;\, A \in \mathcal{B}(\mathbb{R})^{\otimes I}\}$ which is a σ-field on $\mathcal{C}^0(I)$; it is called the restriction of $\mathcal{B}(\mathbb{R})^{\otimes I}$ to $\mathcal{C}^0(I)$. We admit the following lemma which states that the Borel σ-field of the Banach space $\mathcal{C}^0(I)$ is $\mathcal{C}^0(I) \cap \mathcal{B}(\mathbb{R})^{\otimes I}$, see Example 1.3 in [1]. (The proof given in [1] has to be adapted when I is not compact).

Lemma 6.12. *Let I be an interval of $\mathbb{R}_+$. We have $\mathcal{B}(\mathcal{C}^0(I)) = \mathcal{C}^0(I) \bigcap \mathcal{B}(\mathbb{R})^{\otimes I}$.*

Since a probability measure on $\mathbb{R}^I$ is characterized by the distribution of the corresponding finite marginals, one can then prove that a probability measure on $\mathcal{C}^0(I)$ is also characterized by the distribution of the corresponding finite marginals. We also admit the following theorem, which says that the Brownian motion is a.s. continuous, see Theorem I.2.2 and Corollary I.2.6 in [7].

Theorem 6.13. *There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a $\mathcal{C}^0(\mathbb{R}_+)$-valued process $B = (B_t, t \in \mathbb{R}_+)$ defined on it which is a Brownian motion (when one sees B as a $\mathbb{R}^{\mathbb{R}_+}$-valued process). Furthermore, for any $\varepsilon > 0$, B is a.s. Hölder with index $1/2 - \varepsilon$ and a.e. not Hölder with index $1/2 + \varepsilon$.*

In particular, the Brownian motion has no derivative.

6.2.2 Limit of simple random walks

Let Y be a $\mathbb{R}$-valued square integrable random variable such that $\mathbb{E}[Y] = 0$ and $\operatorname{Var}(Y) = 1$. Let $(Y_n, n \in \mathbb{N})$ be independent random variables distributed as Y. We consider the random walk $S = (S_n, n \in \mathbb{N})$ defined by:

$$S_0 = 0 \quad \text{and} \quad S_n = \sum_{k=1}^{n} Y_k \quad \text{for } n \in \mathbb{N}^*.$$

We consider a time-space scaling $X^{(n)} = (X_t^{(n)}, t \in \mathbb{R}_+)$ of the process S given by, for $n \in \mathbb{N}^*$:

$$X_t^{(n)} = \frac{1}{\sqrt{n}} S_{\lfloor nt \rfloor}.$$

We have the following important result.

Proposition 6.14. *Let $B = (B_t, t \in \mathbb{R}_+)$ be a standard Brownian motion. The sequence of processes $(X^{(n)}, n \in \mathbb{N}^*)$ converges in distribution for the finite dimensional marginals towards B: for all $k \in \mathbb{N}^*$, $t_1, \dots, t_k \in \mathbb{R}_+$, the sequence of vectors $((X_{t_1}^{(n)}, \dots, X_{t_k}^{(n)}), n \in \mathbb{N}^*)$ converges in distribution towards $(B_{t_1}, \dots, B_{t_k})$.*

Proof. We deduce from the central limit theorem that $(\lfloor nt \rfloor^{-1/2} S_{\lfloor nt \rfloor}, n \in \mathbb{N}^*)$ converges in distribution towards a Gaussian random variable with distribution $\mathcal{N}(0,1)$. This implies that $(X_t^{(n)}, n \in \mathbb{N}^*)$ converges in distribution towards B_t. This gives the convergence in distribution of the 1-dimensional marginals of $X^{(n)}$ towards those of B.

Let $t \geq s \geq 0$. By construction, we have that $X_t^{(n)} - X_s^{(n)}$ is independent of $\sigma(X_u^{(n)}, u \in [0,s])$ and distributed as $a_n(t,s) X_{t-s}^{(n)}$ with $a_n(t,s) = \lfloor n(t-s) \rfloor / (\lfloor nt \rfloor - \lfloor ns \rfloor)$ if $\lfloor nt \rfloor - \lfloor ns \rfloor > 0$ and $a_n(t,s) = 1$ otherwise. Since $\lim_{n \to \infty} a_n(t,s) = 1$, we deduce that $\left((X_s^{(n)}, X_t^{(n)} - X_s^{(n)}), n \in \mathbb{N}^*\right)$ converges in distribution towards (G_1, G_2), where G_1 and G_2 are independent Gaussian random variable with $G_1 \sim \mathcal{N}(0,s)$ and $G_2 \sim \mathcal{N}(0, t-s)$. Notice that (G_1, G_2) is distributed as $(B_s, B_t - B_s)$. Indeed $(B_s, B_t - B_s)$ is Gaussian vector as the linear transformation of the Gaussian vector (B_s, B_t); it has mean $(0,0)$ and we have $\mathrm{Var}(B_s) = s$, $\mathrm{Var}(B_t - B_s) = t - s$, see (6.5), and $\mathrm{Cov}(B_s, B_t - B_s) = 0$, see (6.6), so the mean and covariance matrix of (G_1, G_2) and $(B_s, B_t - B_s)$ are the same. This gives they have the same distribution. We deduce that $\left((X_s^{(n)}, X_t^{(n)}), n \in \mathbb{N}^*\right)$ converges in distribution towards (B_s, B_t). This gives the convergence in distribution of the 2-dimensional marginals of $X^{(n)}$ towards those of B.

The convergence in distribution of the k-dimensional marginals of $X^{(n)}$ towards those of B is then an easy extension which is left to the reader. $\square$

In fact we can have a much stronger statement concerning this convergence by considering a continuous linear interpolation of the processes $X^{(n)}$. For $n \in \mathbb{N}^*$, we define the continuous process $\tilde{X}^{(n)} = (\tilde{X}_t^{(n)}, t \in \mathbb{R}_+)$ by $\tilde{X}_t^{(n)} = X_t^{(n)} + C_t^{(n)}$, where $C_t^{(n)} = \frac{1}{\sqrt{n}}(nt - \lfloor nt \rfloor) Y_{\lfloor nt \rfloor + 1}$. Notice that $\mathbb{E}[|C_t^{(n)}|^2] \leq n^{-1}$ so that $(C_t^{(n)}, n \in \mathbb{N}^*)$ converges in probability towards 0 for all $t \in \mathbb{R}_+$. We deduce that the sequence $(\tilde{X}^{(n)}, n \in \mathbb{N}^*)$ converges in distribution for the finite dimensional marginals towards B. The Donsker's theorem state this convergence in distribution holds for the process seen as continuous functions. For a function $f = (f(t), t \in \mathbb{R}_+)$ defined on $\mathbb{R}_+$, we write $f_{[0,1]} = (f(t), t \in [0,1])$ for its restriction to $[0,1]$. We admit the following result, see Theorem 8.2 in [1].

Theorem 6.15 (Donsker (1951))**.** *The sequence of processes $\left(\tilde{X}_{[0,1]}^{(n)}, n \in \mathbb{N}^*\right)$ converges in distribution, on the space $\mathcal{C}^0([0,1])$, towards $B_{[0,1]}$, where B is a standard Brownian motion.*

In particular, we get that for all continuous functional F defined on $\mathcal{C}^0([0,1])$, we have that $(F(\tilde{X}^{(n)}_{[0,1]}), n \in \mathbb{N}^*)$ converges in distribution towards $F(B_{[0,1]})$. For example the following real-valued functionals, say F, on $\mathcal{C}^0([0,1])$ are continuous, for $f \in \mathcal{C}^0([0,1])$:

$$F(f) = \|f\|_\infty, \quad F(f) = \sup_{[0,1]}(f), \quad F(f) = \int_{[0,1]} f \, \mathrm{d}\lambda, \quad F(f) = f(t_0) \text{ for some } t_0 \in [0,1].$$

6.2.3 Markov property

Let $\mathbb{F} = (\mathcal{F}_t, t \in \mathbb{R}_+)$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$, that is a non-decreasing of family of σ-fields, subsets of $\mathcal{F}$. A process $(X_t, t \in \mathbb{R}_+)$ defined on Ω is said $\mathbb{F}$-adapted if X_t is $\mathcal{F}_t$-measurable for all $t \in \mathbb{R}_+$.

Definition 6.16. *Let $X = (X_t, t \in \mathbb{R}_+)$ be a $\mathbb{R}$-valued process adapted to the filtration $\mathbb{F} = (\mathcal{F}_t, t \in \mathbb{R}_+)$.*

(i) We say that X is a Markov process with respect to the filtration $\mathbb{F}$ if for all $t \in \mathbb{R}_+$, conditionally on the σ-field $\sigma(X_t)$ the σ-fields $\mathcal{F}_t$ and $\sigma(X_u, u \geq t)$ are independent.

(ii) We say that X has independent increments if for all $t \geq s \geq 0$, $X_t - X_s$ is independent of $\mathcal{F}_s$.

(iii) We say that X has stationary increments if for all $t \geq s$, $X_t - X_s$ is distributed as $X_{t-s} - X_0$.

In the previous definition, usually one takes $\mathbb{F}$ the natural filtration of X, that is $\mathcal{F}_t = \sigma(X_u, u \in [0,t])$. Clearly, if a process has independent increments, it has the Markov property (with respect to its natural filtration).

Lemma 6.17. *The Brownian motion is a Markov process (with respect to its natural filtration), with independent and stationary increments.*

Proof. Let $B = (B_t, t \in \mathbb{R}_+)$ be a standard Brownian motion and $\mathbb{F} = (\mathcal{F}_t, t \in \mathbb{R}_+)$ its natural filtration, that is $\mathcal{F}_t = \sigma(B_u,\, u \in [0,t])$. It is enough to check that it has independent and stationary increments. Let $t \geq s \geq 0$. Since B is a Brownian process, we deduce that $B_t - B_s$ is Gaussian, and we have $\mathrm{Var}(B_t - B_s) = t - s = \mathrm{Var}(B_{t-s})$, see (6.5). Since B is centered, we deduce that $B_t - B_s$ and B_{t-s} have the same distribution $\mathcal{N}(0, t-s)$. Thus B has stationary increments. Since B is a Gaussian process and, according to (6.5), $\mathrm{Cov}(B_u, B_t - B_s) = 0$ for all $u \in [0,s]$, we deduce that $B_t - B_s$ is independent of $\mathcal{F}_s = \sigma(B_u, u \in [0,s])$. Thus, B has independent increments. The extension to a general Brownian motion is immediate. $\square$

We mention that the Brownian motion is the only continuous random process with independent and stationary increments (the proof of this fact is beyond those notes), and that the study of general random process with independent and stationary increments is a very active domain of the probabilities.

6.2.4 Brownian bridge and simulation

Let $B = (B_t, t \in \mathbb{R}_+)$ be a standard Brownian motion. Let $T > 0$ be given. The Brownian bridge over $[0, T]$ is the distribution of the process $W^T = (W_t^T, t \in [0, T])$ defined by:

$$W_t^T = B_t - \frac{t}{T} B_T.$$

See Exercise 8.38 for the recursive simulation of the Brownian motion using Brownian bridge approach.

6.2.5 Martingale and stopping times

We shall admit all the results presented in this section and refer to [2, 3, 7, 8, 9]. We consider a standard Brownian motion $B = (B_t, t \in \mathbb{R}_+)$ seen as a random variable on $\mathcal{C}^0(\mathbb{R}_+)$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The Brownian filtration $\mathbb{F} = (\mathcal{F}_t, t \in \mathbb{R}_+)$ of the Brownian motion is given by $\mathcal{F}_t$ generated by $\sigma(B_u, u \in [0, t])$. We set $\mathcal{F}_\infty = \bigvee_{t \in \mathbb{R}_+} \mathcal{F}_t$.

A non-negative real-valued random variable τ is a stopping time with respect to the filtration $\mathbb{F}$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \in \mathbb{R}_+$. The σ-field $\mathcal{F}_\tau$ of the events which are prior to a stopping time τ is defined by:

$$\mathcal{F}_\tau = \{B \in \mathcal{F}_\infty;\ B \cap \{\tau \leq t\} \in \mathcal{F}_t \quad \text{for all} \quad t \in \mathbb{R}_+\}.$$

Remark 6.18. We recall the convention that $\inf \emptyset = +\infty$. Let A be an open set of $\mathbb{R}$. The entry time $\tau_A = \inf\{t \geq 0;\ B_t \in A\}$ is a stopping time [1]. Indeed, we have for $t \geq 0$ that:

$$\{\tau_A \leq t\} = \bigcup_{s \in \mathbb{Q}_+,\, s \leq t} \{B_s \in A\} \in \mathcal{F}_t.$$

$\Diamond$

A real-valued process $M = (M_t, t \in \mathbb{R}_+)$ is called a $\mathbb{F}$-martingale if it is $\mathbb{F}$-adapted (that is M_t is $\mathcal{F}_t$-measurable for all $t \in \mathbb{R}_+$) and for all $t \geq s \geq 0$, M_t is integrable and:

$$\mathbb{E}[M_t | \mathcal{F}_s] = M_s \quad a.s.. \tag{6.7}$$

If (6.7) is replaced by $\mathbb{E}[M_t | \mathcal{F}_s] \geq M_s$ a.s., then M is called an $\mathbb{F}$-sub-martingale.
If (6.7) is replaced by $\mathbb{E}[M_t | \mathcal{F}_s] \leq M_s$ a.s., then M is called an $\mathbb{F}$-super-martingale.

In this setting, we admit the following optional stopping theorem, see [7].

Theorem 6.19. *If M is a continuous $\mathbb{F}$-martingale and T, S are two bounded stopping times such that $S \leq T$, then we have:*

$$\mathbb{E}[M_T | \mathcal{F}_S] = M_S \quad a.s.. \tag{6.8}$$

In particular, we get that $\mathbb{E}[M_T] = \mathbb{E}[M_0]$.

[1] One can prove that if $X = (X_t, t \in \mathbb{R}_+)$ is an a.s. continuous process taking values in a metric space E and A a Borel subset of E then: the entry time $\tau_A = \inf\{t \geq 0;\ B_t \in A\}$ is a stopping time with respect to the natural filtration $\mathbb{F} = (\mathcal{F}_t, t \in \mathbb{R}_+)$ where $\mathcal{F}_t = \sigma(X_u,\, u \in [0, t])$; and the hitting time $T_A = \inf\{t > 0;\ B_t \in A\}$ is a stopping time with respect to the filtration $(\mathcal{F}_{t+}, t \in \mathbb{R}_+)$ where $\mathcal{F}_{t+} = \bigcap_{s > t} \mathcal{F}_s$.

We admit that the Brownian motion has the strong Markov property, see [7].

Theorem 6.20. *Let T be a finite stopping time. Then $\tilde{B} = (\tilde{B}_t = B_{T+t} - B_T, t \in \mathbb{R}_+)$ is a standard Brownian motion independent of $\mathcal{F}_T$.*

6.3 Wiener integrals

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which is defined a Brownian motion $B = (B_t, t \in \mathbb{R}_+)$. In Section 6.3.1, we shall give a precise meaning of the Wiener integral $\int_0^t f(s)\, \mathrm{d}B_s$ for some general function f, whereas the Brownian motion is not differentiable. This integral (and its generalization known as the Itô integral when the integrand f is also random) has been intensively used in physics, biology, finance, applied mathematics, ... The Wiener integral can also bee seen as an extension of the stochastic discrete integrals (see Lemma 4.14) to the continuous case. This approach, which is not developed here, known as stochastic calculus with respect to martingales is another very important generalization of the integrals with respect to the Brownian motion. We refer to [7, 8, 9] for a complete exposition. We present in Section 6.3.2 an application to the Langevin equation which describes the evolution of the speed of a particle in a fluid. In Section 6.3.3, we use the Wiener integral to define the Cameron-Martin change of probability measure and compute the Laplace transform of the hitting time of a line by the Brownian motion.

6.3.1 Gaussian space

Let λ be the Lebesgue measure on $\mathbb{R}$. The vector space $L^2(\lambda) = L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \lambda)$ endowed with the scalar product $\langle f, g\rangle_\lambda = \int_{\mathbb{R}_+} fg\, \mathrm{d}\lambda$ is an Hilbert space. We consider the vector space $\mathcal{I} = \mathrm{Vect}(\mathbf{1}_{[0,t)},\, t \in \mathbb{R}_+) \subset L^2$ of finite linear combinations of indicators of intervals $[0, t)$ for $t \in \mathbb{R}_+$ that is:

$$\mathcal{I} = \left\{ \sum_{k=1}^{n} a_{k-1} \mathbf{1}_{[t_{k-1}, t_k)} \text{ for some } n \in \mathbb{N}^*,\ 0 = t_0 < \cdots < t_n \text{ and } a_0, \ldots, a_{n-1} \in \mathbb{R} \right\}. \tag{6.9}$$

We have the following density result.

Lemma 6.21. *The vector space $\mathcal{I}$ is dense in the Hilbert space L^2.*

Proof. Assume the vector space $\mathcal{I}$ is not dense in the Hilbert space L^2, that is, there exists $f \in L^2$ orthogonal to $\mathcal{I}$ and non zero. We get that for all $t \geq 0$, $\int_{\mathbb{R}_+} \mathbf{1}_{[0,t]}\, f_+\, \mathrm{d}\lambda = \int_{\mathbb{R}_+} \mathbf{1}_{[0,t]}\, f_+\, \mathrm{d}\lambda$, and denote by c_t this common value. Since f is non-zero, there exits T such that $c_T > 0$. The two probability measures $\mathbf{1}_{[0,T]} f_+\, \lambda / c_T$ and $\mathbf{1}_{[0,T]} f_-\, \lambda / c_T$ coincide on the sets $\mathcal{C} = \{[0, t],\, t \in \mathbb{R}_+\}$. Since the σ-field generated by $\mathcal{C}$ is the Borel σ-field on $\mathbb{R}_+$, we deduce from the monotone class theorem, see Corollary 1.14, that $\mathbf{1}_{[0,T]} f_+\, \lambda = \mathbf{1}_{[0,T]} f_-\, \lambda$, and thus $\mathbf{1}_{[0,T]} f_+ = \mathbf{1}_{[0,T]} f_-$. By definition of f_+ and f_-, this implies that $f = 0$ a.e. on $[0, T]$, and thus $c_T = 0$. Since this is absurd, we deduce that the vector space $\mathcal{I}$ is dense in the Hilbert space L^2. $\square$

We deduce from Proposition 1.50 that $L^2(\mathbb{P}) = L^2(\Omega, \mathcal{F}, \mathbb{P})$ endowed with the scalar product $\langle X, Y\rangle_{\mathbb{P}} = \mathbb{E}[XY]$ for $X, Y \in L^2(\mathbb{P})$ is an Hilbert space. Let $\mathcal{I}_B = \mathrm{Vect}(B_t,\, t \in$

$\mathbb{R}_+) \in L^2(\mathbb{P})$ be the vector space of finite linear combinations of marginals of B, that is:

$$\mathcal{I}_B = \Big\{ \sum_{k=1}^n a_{k-1}(B_{t_k} - B_{t_{k-1}}) \text{ for some } n \in \mathbb{N}^*,\ 0 = t_0 < \cdots < t_n \text{ and } a_0, \ldots, a_{n-1} \in \mathbb{R} \Big\}.$$

Let H_B be the closure in $L^2(\mathbb{P})$ of $\mathcal{I}_B$. Notice that H_B is also an Hilbert space. The space H_B is a Gaussian space in the following sense.

Lemma 6.22. *Let $d \in \mathbb{N}^*$ and $X_1, \ldots, X_d \in H_B$. Then the random vector $(X_1, \ldots, X_d)$ is a Gaussian centered vector.*

Proof. We first consider the case $d = 1$. Since $X_1 \in H_B$, there exists a sequence $(Y_k, k \in \mathbb{N})$ of elements of $\mathrm{Vect}(B_t,\, t \in \mathbb{R}_+)$ which converges in $L^2(\mathbb{P})$ towards X_1. Thanks to Lemma 6.2, Y_k is a centered Gaussian random variable for all $k \in \mathbb{N}$. Thanks to Lemma 6.4, we get that X_1 is also a centered Gaussian random variable. The general case $d \in \mathbb{N}^*$ is proved using similar arguments. □

For $f = \sum_{k=1}^n a_{k-1}\mathbf{1}_{[t_{k-1},t_k)}$, element of $\mathcal{I}$, we define the integral of f with respect to the Brownian motion as $I(f) = \sum_{k=1}^n a_{k-1}(B_{t_k} - B_{t_{k-1}})$. Notice that I is a linear map from $\mathcal{I}$ to $\mathcal{I}_B$. We have $I(\mathbf{1}_{[0,t)}) = B_t$ for all $t \in \mathbb{R}_+$ and thus for $t, s \in \mathbb{R}_+$:

$$\langle I(\mathbf{1}_{[0,t)}), I(\mathbf{1}_{[0,s)})\rangle_{\mathbb{P}} = \mathbb{E}[B_t, B_s] = s \wedge t = \langle \mathbf{1}_{[0,t)}, \mathbf{1}_{[0,s)}\rangle_\lambda.$$

By linearity, we deduce that for $f, g \in \mathcal{I}$, $\langle I(f), I(g)\rangle_{\mathbb{P}} = \langle f, g\rangle_\lambda$. Therefore, I is a linear isometric map from $\mathcal{I}$ to $\mathcal{I}_B$. It admits a unique linear isometric extension from $\overline{\mathcal{I}} = L^2(\lambda)$ to $\overline{\mathcal{I}_B} = H_B$, which we still denote by I. The Wiener integral of a function $f \in L^2(\lambda)$ with respect to the Brownian motion B is any random variable a.s. equal to $I(f) \in H_B$. We shall use the notation $I(f) = \int_{\mathbb{R}_+} f(s)\,\mathrm{d}B_s = \int_{\mathbb{R}_+} f\,\mathrm{d}B$, even if the Brownian motion has no derivative. We shall also use the notation $\int_0^t f\,\mathrm{d}B = \int_0^t f(s)\,\mathrm{d}B_s = \int_{\mathbb{R}_+} \mathbf{1}_{[0,t)}(s) f(s)\,\mathrm{d}B_s$. With this convention, we have $\int_0^t \mathrm{d}B_s = B_t$.

We have the following properties of the Wiener integral.

Proposition 6.23. *Let $f, g \in L^2(\lambda)$.*

(i) The random variable $\int_{\mathbb{R}_+} f\,\mathrm{d}B$ is Gaussian with distribution $\mathcal{N}\left(0, \int_{\mathbb{R}_+} f^2\,\mathrm{d}\lambda\right)$.

(ii) The Gaussian random variables $\int_{\mathbb{R}_+} f\,\mathrm{d}B$ and $\int_{\mathbb{R}_+} g\,\mathrm{d}B$ are independent if and only if $\int_{\mathbb{R}_+} fg\,\mathrm{d}\lambda = 0$.

(iii) Let h be a measurable real-valued function defined on $\mathbb{R}_+$ locally square integrable (that is $\int_0^t h^2\,\mathrm{d}\lambda < +\infty$ for all $t \in \mathbb{R}_+$). The process $M = (M_t = \int_0^t h\,\mathrm{d}B, t \in \mathbb{R}_+)$ is a martingale.

(iv) The process M given in (iii) is a centered Gaussian process with covariance kernel $K = (K(s,t);\, s, t \in \mathbb{R}_+)$ given by $K(s,t) = \int_0^{s\wedge t} h^2\,\mathrm{d}\lambda$.

Proof. Proof of property (i). Let $f \in L^2(\lambda)$. Since $I(f)$ belongs to H_B, we deduce from Lemma 6.22, that $I(f)$ is a centered Gaussian random variable. Its variance is given by $\mathbb{E}[I(f)^2] = \langle I(f), I(f)\rangle_{\mathbb{P}} = \langle f, f\rangle_\lambda = \int_{\mathbb{R}_+} f^2\,\mathrm{d}\lambda$.

Proof of property (ii). Let $f, g \in L^2(\lambda)$. Since $I(f)$ and $I(g)$ belongs to the Gaussian space H_B and are centered, we deduce from Lemmas 6.6 and 6.22, that $I(f)$ and $I(g)$ are independent if and only if $\mathbb{E}[I(f)I(g)] = 0$. Then use that $\mathbb{E}[I(f)I(g)] = \langle I(f), I(g)\rangle_{\mathbb{P}} = \langle f, g\rangle_\lambda = \int_{\mathbb{R}_+} fg \, \mathrm{d}\lambda$ to conclude.

Proof of property (iii). Notice that $M_t \in L^2(\mathbb{P})$ for all $t \geq 0$. Let $t \geq s \geq 0$ be fixed. Since $h\mathbf{1}_{[s,t)}$ belongs to L^2, there exists a sequence $(f_n, n \in \mathbb{N})$ of elements of $\mathcal{I}$ which converges to $h\mathbf{1}_{[s,t)}$ in L^2. Clearly the sequence $(f_n\mathbf{1}_{[s,t)}, n \in \mathbb{N})$ converges also to $h\mathbf{1}_{[s,t)}$ in L^2. Since $f_n\mathbf{1}_{[s,t)}$ belongs to $\mathcal{I}$, we get that $I(f_n\mathbf{1}_{[s,t)})$ is $\sigma(B_u - B_s,\, u \in [s,t])$-measurable by construction. This implies that $I(h\mathbf{1}_{[s,t)})$ is also $\sigma(B_u - B_s,\, u \in [s,t])$-measurable. We deduce that $M_t - M_s$ is $\sigma(B_u - B_s,\, u \in [s,t])$-measurable and (taking $s = 0$ and $t = s$) that M_s is $\mathcal{F}_s$-measurable. In particular the process M is adapted to the filtration $\mathbb{F}$. Using that the Brownian motion has independent increments, we get that the σ-fields $\sigma(B_u - B_s,\, u \in [s,t])$ and $\mathcal{F}_s$ are independent. We deduce that $\mathbb{E}[M_t - M_s | \, \mathcal{F}_s] = \mathbb{E}[M_t - M_s] = \mathbb{E}[M_t] - \mathbb{E}[M_s] = 0$. This gives that M is a martingale.

Proof of (iv). Since $M_t \in H_B$ for all $t \geq 0$, we deduce from Lemma 6.22 that M is a centered Gaussian process. Its covariance kernel is given for $s, t \in \mathbb{R}_+$ by $K(s,t) = \mathbb{E}[M_s M_t] = \langle I(h\mathbf{1}_{[0,s)}), I(h\mathbf{1}_{[0,t)})\rangle_{\mathbb{P}} = \langle h\mathbf{1}_{[0,s)}, h\mathbf{1}_{[0,t)}\rangle_\lambda = \int_0^{s\wedge t} h^2 \, \mathrm{d}\lambda$. □

We give a natural representation of $\int_0^t f \, \mathrm{d}B$ when f is of class $\mathcal{C}^1$.

Proposition 6.24. *Assume that $f \in \mathcal{C}^1(\mathbb{R}_+)$. We have the following integration by part formula, for all $t \in \mathbb{R}_+$, a.s.:*

$$\int_0^t f(s) \, \mathrm{d}B_s = f(t)B_t - \int_0^t B_s f'(s) \, \mathrm{d}s.$$

Remark 6.25. If $f \in \mathcal{C}^1(\mathbb{R}_+)$, then the process $\tilde{M} = (\tilde{M}_t = f(t)B_t - \int_0^t B_s f'(s) \, \mathrm{d},\, t \in \mathbb{R}_+)$ is a.s. continuous. Consider the martingale $M = (M_t = \int_0^t f(s) \, \mathrm{d}B_s,\, t \in \mathbb{R}_+)$. From Proposition 6.24, we get that for all $t \in \mathbb{R}_+$, a.s. $\tilde{M}_t = M_t$. We say that $\tilde{M}$ is a continuous version[2] of M. ♢

Proof of Proposition 6.24. For $t \geq 0$, we set $Z_t = f(t)B_t - \int_0^t B_s f'(s) \, \mathrm{d}s - \int_0^t f(s) \, \mathrm{d}B_s$. We have $Z_t \in H_B$. We compute for $u \geq 0$:

$$\begin{aligned}\langle Z_t, B_u\rangle_{\mathbb{P}} = \mathbb{E}[Z_t B_u] &= f(t)\mathbb{E}[B_t B_u] - \int_0^t \mathbb{E}[B_s B_u] f'(s) \, \mathrm{d}s - \mathbb{E}[I(f\mathbf{1}_{[0,t)})I(\mathbf{1}_{[0,u)})] \\ &= f(t)(t \wedge u) - \int_0^t (s \wedge u) f'(s) \, \mathrm{d}s - \int_0^{t\wedge u} f \, \mathrm{d}\lambda \\ &= 0.\end{aligned}$$

Since $\mathrm{Vect}(B_t,\, t \in \mathbb{R}_+)$ is dense in H_B, we deduce that $\langle Z_t, X\rangle_{\mathbb{P}} = 0$ for all $X \in H_B$. As Z_t belongs to H_B, we can take $X = Z_t$ and deduce that a.s. $Z_t = 0$. This ends the proof. □

[2]It can be proven that if f is a measurable real-valued locally square integrable function defined on $\mathbb{R}_+$, then there exists a continuous version of the martingale $(\int_0^t f(s) \, \mathrm{d}B_s,\, t \in \mathbb{R}_+)$.

6.3.2 An application: the Langevin equation

We consider the Langevin equation in dimension 1 which describes the evolution of the speed V of a particle with mass m in a fluid with friction and multiple homogeneous random collisions from molecules of the fluid:

$$m\mathrm{d}V_t = -\lambda V_t \mathrm{d}t + F(t)\,\mathrm{d}t \quad \text{for } t \in \mathbb{R}_+,$$

where $\lambda > 0$ is a damping coefficient, which can be seen as a frictional or drag force, and $F(t)$ is a random force with Gaussian distribution. This force $F(t)\,\mathrm{d}t$ is modeled by a Brownian motion $\rho\,\mathrm{d}B_t$, with $\rho > 0$. Taking $a = \lambda/m > 0$ and $\sigma = \rho/m > 0$, we get the stochastic differential equation:

$$\mathrm{d}V_t = -aV_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t \quad \text{for } t \in \mathbb{R}_+. \tag{6.10}$$

We say that a random locally integrable process $V = (V_t, t \in \mathbb{R}_+)$ is solution to the Langevin equation (6.10) with initial condition V_0 if a.s.:

$$V_t = V_0 - a\int_0^t V_s\,\mathrm{d}s + \sigma B_t.$$

We have the following solution to the Langevin equation.

Proposition 6.26. *Let $a > 0$ and $\sigma > 0$. The equation (6.10) with initial condition V_0 has a unique locally integrable solution $V = (V_t, t \in \mathbb{R}_+)$ given a.s. by:*

$$V_t = V_0\,\mathrm{e}^{-at} + \sigma\int_0^t \mathrm{e}^{-a(t-s)}\,\mathrm{d}B_s \quad \text{a.s. for } t \in \mathbb{R}_+.$$

Proof. We define $Y = (Y_t, t \in \mathbb{R}_+)$ by $Y_t = V_0\,\mathrm{e}^{-at} + \sigma\int_0^t \mathrm{e}^{-a(t-s)}\,\mathrm{d}B_s$ for all $t \in \mathbb{R}_+$. Using Proposition 6.24, we have that a.s. for $t \geq 0$:

$$\begin{aligned} Y_t &= V_0\,\mathrm{e}^{-at} + \sigma\,\mathrm{e}^{-at}\left(\mathrm{e}^{at} B_t - a\int_0^t B_s\,\mathrm{e}^{as}\,\mathrm{d}s\right) \\ &= V_0\,\mathrm{e}^{-at} + \sigma B_t - a\sigma\int_0^t B_s\,\mathrm{e}^{-a(t-s)}\,\mathrm{d}s. \end{aligned}$$

We deduce that $a\int_0^t Y_s\,\mathrm{d}s = V_0(1 - \mathrm{e}^{-at}) + a\sigma\int_0^t B_s\,\mathrm{d}s - a\sigma X_t$, where, using Fubini theorem:

$$X_t = a\int_0^t \mathrm{d}s \int_0^s B_u\,\mathrm{e}^{-a(s-u)}\,\mathrm{d}u = \int_0^t \mathrm{d}u\, B_u\,\mathrm{e}^{au}\int_u^t a\,\mathrm{e}^{-as}\,\mathrm{d}s = \int_0^t B_u\left(1 - \mathrm{e}^{-a(t-u)}\right)\mathrm{d}u.$$

We get that a.s. for all $t \geq 0$: $a\int_0^t Y_s\,\mathrm{d}s = V_0 - Y_t + \sigma B_t$. This gives that Y is a solution to (6.10) with initial condition V_0. Let $Y' = (Y'_t, t \in \mathbb{R}_+)$ be another locally integrable solution. Taking $Z_t = Y_t - Y'_t$, we get that $Z = (Z_t, t \in \mathbb{R}_+)$ is locally integrable and that a.s. $Z_0 = 0$ and for all $t \geq 0$: $Z_t = -a\int_0^t Z_s\,\mathrm{d}s$. This gives that a.s. $Z_t = 0$ for all $t \geq 0$. Hence there exists at most one locally integrable solution to (6.10) with initial condition V_0. □

The process V given in Proposition 6.26 is called the Ornstein-Uhlenbeck process. See Exercise 8.39 for results on this process.

The Ornstein-Uhlenbeck process can be defined for all times in $\mathbb{R}$ by $(\frac{\sigma}{\sqrt{2a}}\,\mathrm{e}^{-at}\,B_{\mathrm{e}^{2at}},\, t \in \mathbb{R})$. This definition is coherent thanks to (8.3).

If we consider the position of the particle at time $t \geq 0$, say X_t, we get that $X_t = X_0 + \int_0^t V_s \mathrm{d}s$, which gives the following result whose proof is immediate.

Lemma 6.27. *Let $a > 0$ and $\sigma > 0$. The path of the particle $X = (X_t, t \in \mathbb{R}_+)$ governed by the Langevin equation (6.10) is given by a.s.:*

$$\begin{aligned} X_t &= X_0 + \int_0^t V_s \mathrm{d}s \\ &= X_0 + \frac{V_0}{a}\left(1 - \mathrm{e}^{-at}\right) + \frac{\sigma}{a} B_t - \frac{\sigma}{a}\int_0^t \mathrm{e}^{-a(t-r)}\,\mathrm{d}B_r \quad \textit{for } t \in \mathbb{R}_+. \end{aligned}$$

Remark 6.28. Recall that $a = \lambda/m$ and $\sigma = \rho/m$, with m the mass of the particle. Denote by $X^{(m)} = (X_t^{(m)}, t \in \mathbb{R}^+)$ the path of the particle with mass m. We have:

$$X_t^{(m)} = X_0 + \frac{V_0}{a}\left(1 - \mathrm{e}^{-at}\right) + \rho B_t - \rho \int_0^t \mathrm{e}^{-a(t-r)}\,\mathrm{d}B_r.$$

Letting m goes to 0 (that is considering an infinitesimal particle), or equivalently a goes to infinity, with X_0 and ρ fixed, we get that $X_t^{(m)}$ converges in $L^2(\mathbb{P})$ towards the Einstein model for the motion of a infinitesimal particle in a fluid $X^{(0)} = (X_t^{(0)}, t \in \mathbb{R}^+)$ given by:

$$X_t^{(0)} = X_0 + \rho B_t.$$

Indeed, we have, as a goes to infinity, that $\lim_{a \to +\infty} \frac{V_0}{a}\left(1 - \mathrm{e}^{-at}\right) = 0$ and

$$\lim_{a \to +\infty} \mathbb{E}\left[\left(\int_0^t \mathrm{e}^{-a(t-r)}\,\mathrm{d}B_r\right)^2\right] = \lim_{a \to +\infty} \int_0^t \mathrm{e}^{-2a(t-r)}\,\mathrm{d}r = 0,$$

which implies that $X_t^{(m)}$ converges to $X_0 + \rho B_t$ in $L^2(\mathbb{P})$ as m goes to zero. $\Diamond$

6.3.3 Cameron-Martin Theorem

The Cameron-Martin theorem is a particular case of the family of Girsanov theorem whose spirit is a change of probability measure using exponential martingales.

Let h be a measurable real-valued function defined on $\mathbb{R}_+$ locally square integrable (that is $\int_0^t h^2\,\mathrm{d}\lambda < +\infty$ for all $t \in \mathbb{R}_+$). We consider the non-negative process $M^h = (M_t^h, t \in \mathbb{R}_+)$ defined by:

$$M_t^h = \exp\left(\int_0^t h(s)\,\mathrm{d}B_s - \frac{1}{2}\int_0^t h(s)^2\,\mathrm{d}s\right). \tag{6.11}$$

Lemma 6.29. *The process M^h defined in (6.11) is a non-negative martingale.*

Proof. According to property (iii) of Proposition 6.23, we get that M is adapted to the Brownian filtration. Notice that for $t \geq s \geq 0$, we have:

$$M_t^h = M_s^h \, \mathrm{e}^{G-(\sigma^2/2)}$$

with $G = \int_s^t h(s)\, \mathrm{d}B_s$ and $\sigma^2 = \int_s^t h(s)^2 \, \mathrm{d}s$. Arguing as in the proof of property (iii) of Proposition 6.23, we get that G is independent of $\mathcal{F}_s$ and has distribution $\mathcal{N}(0, \sigma^2)$. We deduce that a.s.:

$$\mathbb{E}\left[M_t^h | \, \mathcal{F}_s\right] = M_s^h \, \mathbb{E}\left[\mathrm{e}^{G-(\sigma^2/2)}\right] = M_s^h,$$

where the last equality is a consequence of (6.3) with $\lambda = 1$ and $X = G$. This implies that M is a martingale. By construction, it is non-negative. □

The next theorem assert that a Brownian motion with a drift can be seen as a Brownian motion under a different probability measure.

Theorem 6.30 (Cameron-Martin). *Let $t \geq 0$ and F be a real-valued non-negative measurable function defined on $\left(\mathcal{C}^0([0,t]), \mathcal{B}(\mathcal{C}^0([0,t]))\right)$. We have:*

$$\mathbb{E}\left[F\left(\left(B_u + \int_0^u h(s)\, \mathrm{d}s, \; u \in [0,t]\right)\right)\right] = \mathbb{E}\left[M_t^h \, F\left((B_u, \; u \in [0,t])\right)\right]. \tag{6.12}$$

Remark 6.31. Let $\tilde{\mathbb{P}}$ be a probability measure defined on $(\Omega, \mathcal{F}_t)$ by $\tilde{\mathbb{P}}(A) = \mathbb{E}[M_t^h \mathbf{1}_A]$ for $A \in \mathcal{F}_t$. (Check this define indeed a probability measure.) Let $\tilde{\mathbb{E}}$ denote the corresponding expectation. Theorem 6.30 gives that:

$$\mathbb{E}\left[F\left(\left(B_u + \int_0^u h(s)\, \mathrm{d}s, \; u \in [0,t]\right)\right)\right] = \tilde{\mathbb{E}}\left[F\left((B_u, \; u \in [0,t])\right)\right].$$

In particular, the process $t \mapsto B_t - \int_0^t h(s)\, \mathrm{d}s$ is a Brownian motion under $\tilde{\mathbb{P}}$. ◊

Partial proof of Theorem 6.30. We assume[3] that it is enough to check (6.12) for functions F of the form $F((Y_u, u \in [0,t])) = \exp\left(\int_0^t f(u)\, \mathrm{d}Y_u\right)$ with $f \in \mathcal{I}$ and $\mathcal{I}$ defined by (6.9). So, we have $F\left((B_u, \; u \in [0,t])\right) = \exp\left(\int_0^t f(u)\, \mathrm{d}B_u\right)$ and $F\left(\left(B_u + \int_0^u h(s)\, \mathrm{d}s, \; u \in [0,t]\right)\right) = \exp\left(\int_0^t f(u)\, \mathrm{d}B_u + \int_0^t f(u)h(u)\, \lambda(\mathrm{d}u)\right)$. We get:

$$\begin{aligned}
\mathbb{E}\left[F\left(\left(B_u + \int_0^u h(s)\, \mathrm{d}s, \; u \in [0,t]\right)\right)\right] &= \mathbb{E}\left[\mathrm{e}^{\int_0^t f(u)\, \mathrm{d}B_u + \int_0^t fh \, \mathrm{d}\lambda}\right] \\
&= \exp\left(\frac{1}{2}\int_0^t f^2 \, \mathrm{d}\lambda + \int_0^t fh \, \mathrm{d}\lambda\right) \\
&= \exp\left(\frac{1}{2}\int_0^t (f+h)^2 \, \mathrm{d}\lambda - \frac{1}{2}\int_0^t h^2 \, \mathrm{d}\lambda\right) \\
&= \mathbb{E}\left[M_t^h \, \mathrm{e}^{\int_0^t f(u)\, \mathrm{d}B_u}\right] \\
&= \mathbb{E}\left[M_t^h \, F\left((B_u, \; u \in [0,t])\right)\right],
\end{aligned}$$

[3] This is the functional version of the monotone class theorem.

where we used that M^f (resp. M^{f+h}) is a martingale for the second (resp. fourth) equality. □

As an application, we can compute the Laplace transform (and hence the distribution) of the hitting time of a line for the Brownian motion. Let $a > 0$ and $\delta \in \mathbb{R}$. We consider:

$$\tau_a^\delta = \inf\{t \in \mathbb{R}_+;\, B_t = a + \delta t\}.$$

Notice that τ_a^δ is a stopping times as $\{\tau_a^\delta \leq t\} = \bigcap_{n\in\mathbb{N}^*} \bigcup_{s\in\mathbb{Q}_+,\, s\leq t} \{B_s - a - \delta s \geq -1/n\}$ which belongs to $\mathcal{F}_t$ for all $t \geq 0$. When $\delta = 0$, we write τ_a for τ_a^δ, and using the continuity of B, we get also that $\tau_a = \inf\{t \in \mathbb{R}_+;\, B_t \geq a\}$.

Proposition 6.32. *Let $a > 0$.*

(i) We have that for all $\lambda \geq 0$:
$$\mathbb{E}\left[\mathrm{e}^{-\lambda\tau_a}\right] = \mathrm{e}^{-a\sqrt{2\lambda}}. \tag{6.13}$$

(ii) Let $\delta \in \mathbb{R}$. We have $\mathbb{P}(\tau_a^\delta < +\infty) = \exp(-2a\delta^+)$ and for $\lambda \geq 0$:
$$\mathbb{E}\left[\mathrm{e}^{-\lambda\tau_a^\delta}\right] = \exp\left(-a\left(\delta + \sqrt{2\lambda + \delta^2}\right)\right).$$

Proof. We first prove (i). Let $\lambda \geq 0$. Consider the process $M = (M_t, t \in \mathbb{R}_+)$ with $M_t = \exp\left(\sqrt{2\lambda}B_t - \lambda t\right)$. Using (6.11), we have $M = M^h$ with h constant equal to $\sqrt{2\lambda}$. Thus M is a continuous martingale. This implies that the process $N = (N_t = M_{\tau_a\wedge t}, t \in \mathbb{R}_+)$ is a continuous martingale thanks to the optional stopping Theorem 6.19. It converges a.s. towards $N_\infty = \mathrm{e}^{a\sqrt{2\lambda} - \lambda\tau_a}\,\mathbf{1}_{\{\tau_a<+\infty\}}$ as $B_{\tau_a} = a$ on $\{\tau_a < +\infty\}$. Since the process N takes values in $[0, \mathrm{e}^{a\sqrt{2\lambda}}]$, we deduce it converges also in L^1 towards N_∞. By dominated convergence, we get that $\mathbb{E}[N_\infty] = \mathbb{E}[N_0] = 1$ and thus:

$$\mathbb{E}\left[\mathrm{e}^{a\sqrt{2\lambda} - \lambda\tau_a}\,\mathbf{1}_{\{\tau_a<+\infty\}}\right] = 1.$$

Taking $\lambda = 0$ in the previous equality implies that τ_a is a.s. finite. This gives (i).

We now prove (ii). Let f be a non-negative measurable function defined on $\mathbb{R}$. We have:

$$\begin{aligned}
\mathbb{E}[f(\tau_a^\delta \wedge t)] &= \mathbb{E}[f(\inf\{u \in \mathbb{R}_+;\, B_u - \delta u = a\} \wedge t)]\\
&= \mathbb{E}[M_t^{-\delta} f(\tau_a \wedge t)]\\
&= \mathbb{E}[M_{\tau_a\wedge t}^{-\delta} f(\tau_a \wedge t)],
\end{aligned}$$

where we used the Cameron-Martin theorem with $h = -\delta$ for the second equality, the optional stopping Theorem 6.19 (with $T = t$ and $S = \tau_a \wedge t$ and the martingale $M^{-\delta}$ which has a continuous version thanks to Remark 6.25) for the third. Taking $f(x) = \mathrm{e}^{-\lambda x}$ with $\lambda \geq 0$, we get:

$$\mathbb{E}\left[\mathrm{e}^{-\lambda(\tau_a^\delta\wedge t)}\right] = \mathbb{E}\left[\mathrm{e}^{-\lambda(\tau_a\wedge t) - \delta B_{\tau_a\wedge t} - \frac{\delta^2}{2}(\tau_a\wedge t)}\right].$$

Assume $\delta \leq 0$. Letting t goes to infinity in the previous equality and using that τ_a is a.s. finite and $B_{\tau_a} = a$, we get by dominated convergence (for the left hand-side and the right hand-side) that:

$$\mathbb{E}\left[\mathrm{e}^{-\lambda\tau_a^\delta}\right] = \mathbb{E}\left[\mathrm{e}^{-(\lambda+\frac{\delta^2}{2})\tau_a - \delta a}\right].$$

Then use (6.13) to get that $\mathbb{E}\left[\mathrm{e}^{-\lambda\tau_a^\delta}\right] = \exp\left(-\delta a - a\sqrt{2\lambda+\delta^2}\right)$. Letting λ goes down to 0, we deduce that $\mathbb{P}(\tau_a^\delta < \infty) = 1$.

The case $\delta > 0$ is more technical. The idea is to consider the stopping time $\tau_{a,b}^\delta = \inf\{t \in \mathbb{R}_+;\, B_t \notin (b + \delta t, a + \delta t)\}$ for $b < a$; compute the Laplace transform of $\tau_{a,b}^\delta$ and then use that the non-decreasing sequence $(\tau_{a,b}^\delta, b \in (-\infty, a))$ converges to τ_a^δ when b goes to $-\infty$. The details are left to the reader. $\square$

Bibliography

[1] P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[2] A. Klenke. *Probability theory.* Universitext. Springer, London, second edition, 2014. A comprehensive course.

[3] J.-F. Le Gall. *Brownian motion, martingales, and stochastic calculus.* Springer, 2016.

[4] M. A. Lifshits. *Gaussian random functions*, volume 322 of *Mathematics and its Applications.* Kluwer Academic Publishers, 1995.

[5] M. A. Lifshits. *Lecture on Gaussian processes.* Springer Briefs in Mathematics. Springer, 2012.

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

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

[8] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 1.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000.

[9] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes, and martingales. Vol. 2.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000.

Chapter 7

Appendix

7.1 More on measure theory

7.1.1 Construction of probability measures

We give in this section, without proofs, the main theorem which allows to build the usual measures such as Lebesgue measure and product measure.

Definition 7.1. *A collection, $\mathcal{A}$, of subsets of Ω is called a Boolean algebra if:*

(i) $\Omega \in \mathcal{A}$*;*

(ii) $A \in \mathcal{A}$ *implies* $A^c \in \mathcal{A}$*;*

(iii) $A, B \in \mathcal{A}$ *implies* $A \cup B \in \mathcal{A}$*.*

It is easy to check that a Boolean algebra is stable by finite intersection. A probability distribution can be defined on a Boolean algebra (to be compared with Definition 1.7).

Definition 7.2. *Let $\mathcal{A}$ be a Boolean algebra. A probability measure on* $(\Omega, \mathcal{A})$ *is a map* P *defined on $\mathcal{A}$ taking values in* $[0, +\infty]$ *such that:*

(i) $\mathrm{P}(\Omega) = 1$*;*

(ii) Additivity: for all $A, B \in \mathcal{A}$ disjoint, $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$*;*

(iii) Continuity at $\emptyset$: for all sequences $(A_n, n \in \mathbb{N})$ *such that $A_n \in \mathcal{A}$, $A_{n+1} \subset A_n$ for all $n \in \mathbb{N}$ and $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$, then the sequence* $(\mathrm{P}(A_n), n \in \mathbb{N})$ *converges to 0.*

The following extension theorem allows to extend a probability measure on a Boolean algebra to a probability measure on the σ-field generated by the Boolean algebra. Its proof can be found in Section I.5 of [2] or in Section 3 of [1].

Theorem 7.3 (Carathéodory extension theorem)**.** *Let* P *be a probability measure defined on a Boolean algebra $\mathcal{A}$ of Ω. There exists a unique probability measure $\mathbb{P}$ on* $(\Omega, \sigma(\mathcal{A}))$ *such that $\mathbb{P}$ and* P *coincide on $\mathcal{A}$.*

This extension theorem allows to prove the existence of the Lebesgue measure.

Proposition 7.4 (Lebesgue measure)**.** *There exists a unique probability measure $\mathbb{P}$ on the measurable space $([0,1), \mathcal{B}([0,1)))$, called Lebesgue measure, such that $\mathbb{P}([a,b)) = b-a$ for all $0 \leq a \leq b \leq 1$.*

Before giving the proof of Proposition 7.4, we provide a sufficient condition for a real-valued additive function defined on a Boolean algebra to be continuous at $\emptyset$.

Lemma 7.5. *Let $\mathcal{A}$ be a Boolean algebra on $\mathbb{R}^d$, $d \geq 1$. Let P be a $[0,+\infty]$-valued function defined on $\mathcal{A}$ such that $\mathrm{P}(\Omega) = 1$ and P is additive (that is (ii) of Definition 7.2 holds). If for all $A \in \mathcal{A}$ and $\varepsilon > 0$, there exists a compact set $K \subset \mathbb{R}^d$ and $B \in \mathcal{A}$ such that $B \subset K \subset A$ and $\mathrm{P}(A \cap B^c) \leq \varepsilon$, then P is a probability on $\mathcal{A}$ (that P is continuous at $\emptyset$).*

Proof. Let $(A_n, n \in \mathbb{N})$ be a non-increasing $\mathcal{A}$-valued sequence such that $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$. We shall prove that $\lim_{n\to+\infty} \mathrm{P}(A_n) = 0$.

Let $\varepsilon > 0$. For all $k \in \mathbb{N}$, there exists a compact set K_k and $B_k \in \mathcal{A}$ such that $B_k \subset K_k \subset A_k$ and $\mathrm{P}(A_k \cap B_k^c) \leq \varepsilon/2^k$. Since $\bigcap_{k\in\mathbb{N}} A_k = \emptyset$, we get $\bigcap_{k\in\mathbb{N}} K_k = \emptyset$. As a sequence of compact sets with empty intersection has a finite sub-sequence with finite intersection, we deduce there exists $n_0 \in \mathbb{N}$ such that $\bigcap_{k=0}^{n_0} K_k = \emptyset$. This implies that $\bigcap_{k=0}^{n_0} B_k = \emptyset$ and thus $\bigcup_{k=0}^{n} B_k^c = \mathbb{R}^d$ for all $n \geq n_0$. We get that, for $n \geq n_0$:

$$\mathrm{P}(A_n) = \mathrm{P}\left(A_n \cap (\bigcup_{k=0}^{n} B_k^c)\right) \leq \sum_{k=0}^{n} \mathrm{P}(A_n \cap B_k^c) \leq \sum_{k=0}^{n} \mathrm{P}(A_k \cap B_k^c) \leq \sum_{k=0}^{n} \varepsilon 2^{-k} \leq 2\varepsilon,$$

that is $\mathrm{P}(A_n) \leq 2\varepsilon$ for all $n \geq n_0$. Since $\varepsilon > 0$ is arbitrary, we deduce that $\lim_{n\to+\infty} \mathrm{P}(A_n) = 0$, which ends the proof of the lemma. □

Proof of Proposition 7.4. Let $\mathcal{A}$ be the set of finite union of intervals $[a,b)$ with $0 \leq a \leq b \leq 1$. Notice $\mathcal{A}$ is a Boolean algebra which generates the Borel σ-field $\mathcal{B}([0,1))$. Define $\mathrm{P}([a,b)) = b-a$ for $0 \leq a \leq b \leq 1$. It is elementary to check that P can be uniquely extended to $\mathcal{A}$ into an a additive $[0,+\infty]$-valued function, which we still denote by P. Notice that $\mathrm{P}([0,1)) = 1$. To conclude, it is enough to that P is continuous at $\emptyset$.

For $A \in \mathcal{A}$, non empty, there exists $n_0 \in \mathbb{N}^*$, $0 \leq a_i < b_i < a_{i+1}$ for $i \in [\![1, n_0]\!]$, with the convention $a_{n_0+1} = 1$, such that $A = \bigcup_{i\in[\![1,n_0]\!]} [a_i, b_i)$. Let $\varepsilon > 0$. Taking $K = \bigcup_{i\in[\![1,n_0]\!]} [a_i, a_i \vee (b_i - \varepsilon 2^{-i})]$ and $B = \bigcup_{i\in[\![1,n_0]\!]} [a_i, a_i \vee (b_i - \varepsilon 2^{-i}))$ we get that $B \in \mathcal{A}$, $B \subset K \subset A$ and $\mathrm{P}(A \bigcap B^c) \leq \varepsilon$. We deduce that the hypothesis of Lemma 7.5 are satisfied. Thus P is a probability on $\mathcal{A}$. Therefore Theorem 7.3 implies there exists a unique probability on $[0,1)$ which is an extension of P. □

Remark 7.6. Let λ_1 denote the Lebesgue measure on $[0,1)$. Then, the Lebesgue measure on $\mathbb{R}$, λ, is defined by: for all Borel set A of $\mathbb{R}$, $\lambda(A) = \sum_{x\in\mathbb{Z}} \lambda_1\left((A+x) \cap [0,1)\right)$, where $A+x = \{z+x, z \in A\}$. It is easy to check that λ is σ-additive (and thus a measure according to Definition 1.7). Notice that $\lambda([a,b)) = b-a$ for all $a \leq b$. Using Exercise 8.2, we get that the Lebesgue measure is the only measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that this latter property holds.

The construction of the Lebesgue measure, λ, on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ for $d \leq 1$, which is the unique σ-finite measure such that $\lambda(\prod_{i=1}^{d} [a_i, b_i)) = \prod_{i=1}^{d} (b_i - a_i)$ for all $a_i \leq b_i$, follows the same steps and is left to the reader. ◊

Using the extension theorem, we get the existence of the product probability measure of the product measurable space.

Proposition 7.7. *Let* $((\Omega_i, \mathcal{G}_i, \mathbb{P}_i), i \in I)$ *be a collection of probability spaces and set* $\Omega = \prod_{i\in I} \Omega_i$ *as well as* $\mathcal{G} = \bigotimes_{i\in I} \mathcal{G}_i$. *There exists a unique probability measure* $\mathbb{P}$ *on* $(\Omega, \mathcal{G})$ *such that* $\mathbb{P}\left(\prod_{i\in I} A_i\right) = \prod_{i\in I} \mathbb{P}_i\,(A_i)$, *where* $A_i \in \mathcal{G}_i$ *for all* $i \in I$ *and* $A_i = \Omega_i$ *but for a finite number of indices.*

The probability $\mathbb{P}$ is called the product probability measure and it is denoted by $\bigotimes_{i\in I} \mathbb{P}_i$. The probability space $(\Omega, \mathcal{G}, \mathbb{P})$ is called the product probability space. Proposition 7.7 can be extended to the finite product of σ-finite measures, see also Theorem 1.53 for an alternative construction in this particular case.

Proof. Let $\mathcal{A}$ be the set of finite unions of sets of the form $\prod_{i\in I} A_i$, where $A_i \in \mathcal{G}_i$ for all $i \in I$ and $A_i = \Omega_i$ but for a finite number of indices. Notice that $\mathcal{A}$ is a Boolean algebra which generates the product σ-field $\mathcal{G}$. Define $\mathrm{P}(\prod_{i\in I} A_i) = \prod_{i\in I} \mathbb{P}_i(A_i)$. It is elementary to check that P can be uniquely extended to $\mathcal{A}$ into an a additive $[0, +\infty]$-valued function, which we still denote by P. Notice that $\mathrm{P}(\Omega) = 1$. To conclude, it is enough to prove that P is continuous at $\emptyset$.

We first assume that $I = \mathbb{N}^*$. For $n \in \mathbb{N}$, we set $\Omega^n = \prod_{k>n} \Omega_k$ and $\mathcal{A}^n$ the Boolean algebra of the finite unions of set $\prod_{k>n} A'_k$ with $A'_k \in \mathcal{G}_k$ for all $k > n$ and $A'_k = \Omega_k$ but for finite number of indices. Define $\mathrm{P}^n\left(\prod_{k>n} A'_k\right) = \prod_{k>n} \mathbb{P}_k\,(A'_k)$. It is elementary to check that P^n can be uniquely extended to $\mathcal{A}$ into an a additive $[0, +\infty]$-valued function, which we still denote by P^n.

Let us prove that P is continuous at $\emptyset$ by contraposition. Let $(A_n, n \in \mathbb{N}^*)$ be a non-increasing $\mathcal{A}$-valued sequence and $\varepsilon > 0$ such that $\lim_{n\to+\infty} \mathrm{P}\left(\bigcap_{k=1}^n A_k\right) \geq \varepsilon$. We shall prove that $\bigcap_{n\in\mathbb{N}^*} A_n$ is non-empty.

For $\omega_1 \in \Omega_1$, consider $A_n^1(\omega_1) = \{\omega^1 \in \Omega^1;\, (\omega_1, \omega^1) \in A_n\}$ the section of A_n on Ω_1 at ω_1. It is elementary to deduce from $A_n \subset \mathcal{A}$ that for all $\omega_1 \in \Omega_1$ we have $A_n^1(\omega_1) \in \mathcal{A}^1$. It is also not difficult to prove that $B_{n,1} = \{\omega_1 \in \Omega_1; \mathbb{P}^1(A_n^1(\omega_1)) \geq \varepsilon/2\}$ belongs to $\mathcal{G}_1$. Since the sequence $(A_n, n \in \mathbb{N})$ is non-increasing, we get the sequence $(B_n^1, n \in \mathbb{N}^*)$ is also non-increasing. Since A_n is a subset of $B_{n,1} \times \Omega^1 \bigcup \{(\omega_1, \omega^1);\, \omega_1 \notin B_{n,1} \text{ and } \omega^1 \in A_n^1(\omega_1)\}$, we get that:

$$\varepsilon \leq \mathbb{P}(A_n) \leq \mathbb{P}_1(B_{n,1}) + (1 - \mathbb{P}_1(B_{n,1}))\frac{\varepsilon}{2},$$

and thus $\mathbb{P}_1(B_{n,1}) \geq \varepsilon/2$ for all $n \in \mathbb{N}^*$. We deduce from the continuity of $\mathbb{P}_1$ at $\emptyset$, that there exists $\bar{\omega}_1 \in \bigcap_{n\in\mathbb{N}^*} B_{n,1}$. Furthermore, the sequence $(A_n^1(\bar{\omega}_1), n \in \mathbb{N}^*)$ of elements of $\mathcal{A}^1$ is non-increasing and such that $\lim_{n\to+\infty} \mathbb{P}^1\left(\bigcap_{k=1}^n A_k^1(\bar{\omega}_1)\right) \geq \varepsilon/2$ and $\{\bar{\omega}_1\} \times \bigcap_{n\in\mathbb{N}^*} A_n^1(\bar{\omega}_1) \subset \bigcap_{n\in\mathbb{N}^*} A_n$. By iterating the previous argument, we get that for all $k \in \mathbb{N}^*$, there exists $\bar{\omega}_k \in \Omega_k$ such that:

$$\{(\bar{\omega}_1, \ldots, \bar{\omega}_k)\} \times \bigcap_{n\in\mathbb{N}^*} A_n^k(\bar{\omega}_1, \ldots, \bar{\omega}_k) \subset \bigcap_{n\in\mathbb{N}^*} A_n,$$

where $A_n^k(\bar{\omega}_1, \ldots, \bar{\omega}_k) = \{\omega^k \in \Omega^k;\, (\bar{\omega}_1, \ldots, \bar{\omega}_k, \omega^k) \in A_n\}$ is the section of A_n on $\prod_{i=1}^k \Omega_i$ at $(\bar{\omega}_1, \ldots, \bar{\omega}_k)$. This implies that $(\bar{\omega}_k, k \in \mathbb{N}^*) \in \bigcap_{n\in\mathbb{N}^*} A_n$, and thus $\bigcap_{n\in\mathbb{N}^*} A_n$ is non-empty. The proposition is thus true when I is countable.

According to the previous arguments, it is clear the proposition is also true when I is finite. Let us assume that I is uncountable. For all (countable) sequence $(A_n, n \in \mathbb{N}^*)$ of elements of $\mathcal{A}$, there exists a set $J \subset I$ at most countable such that the sets A_n are finite unions of sets of type $\prod_{j\in J} A'_j \prod_{i\in I\setminus J} \Omega_i$ with $A'_j \in \mathcal{G}_j$ for all $j \in J$. Thus we have $A_n = A_n^J \prod_{i\in I\setminus J} \Omega_i$, with $A_n^J = \{\omega_J \in \prod_{j\in J} \Omega_j; \{\omega_J\} \times \prod_{i\in I\setminus J} \Omega_i \subset A_n\}$. And the continuity of $\mathbb{P}$ at $\emptyset$ is the a consequence of the first part of the proof as J is at most countable. □

Based on Proposition 7.7, the next exercise provides an alternative proof of Proposition 7.4 on the existence of the Lebesgue measure on $[0, 1)$.

Exercise 7.1. Set $\Omega_i = \{0, 1\}$, $\mathcal{G}_i = \mathcal{P}(\Omega_i)$ and $\mathbb{P}_i(\{0\}) = \mathbb{P}_i(\{1\}) = 1/2$ for $i \in \mathbb{N}^*$. Consider the product probability space $(\Omega = \prod_{i\in\mathbb{N}^*} \Omega_i$, $\mathcal{G} = \bigotimes_{i\in\mathbb{N}^*} \mathcal{G}_i$, $\mathbb{P} = \bigotimes_{i\in\mathbb{N}^*} \mathbb{P}_i)$. Define the function φ from Ω to $[0, 1)$ by:

$$\varphi((\omega_i, i \in \mathbb{N}^*)) = \sum_{i\in\mathbb{N}^*} 2^{-i}\omega_i.$$

By considering intervals $[k2^{-i}, j2^{-i})$, check that φ is measurable and, using Corollary 1.14, that $\mathbb{P}_\varphi$, the image of $\mathbb{P}$ by φ, is the Lebesgue measure on $([0, 1), \mathcal{B}([0, 1))$. △

7.1.2 Proof of the Carathéodory extension Theorem 7.3

We first give some properties on Boolean algebra.

Proposition 7.8. *Let $\mathcal{A}$ be a Boolean algebra on Ω and* P *a probability measure on $(\Omega, \mathcal{A})$. We have the following properties.*

(i) $\mathrm{P}(A \cup B) + \mathrm{P}(A \cap B) = \mathrm{P}(A) + \mathrm{P}(B)$ *for all $A, B \in \mathcal{A}$.*

(ii) $\mathrm{P}(A) \leq \mathrm{P}(B)$ *for all $A, B \in \mathcal{A}$ such that $A \subset B$.*

(iii) Let $(A_i, i \in I)$ be a family at most countable of elements of $\mathcal{A}$ such that $\bigcup_{i\in I} A_i \in \mathcal{A}$. One has $\mathrm{P}\left(\bigcup_{i\in I} A_i\right) \leq \sum_{i\in I} \mathrm{P}(A_i)$ with an equality if the sets $(A_i, i \in I)$ are pairwise disjoint.

Proof. Properties (i) and (ii) are consequence of the additive property of P.

It is enough to prove property (iii) with I countable. Let $(B_n, n \in \mathbb{N})$ a sequence of elements of $\mathcal{A}$ pairwise disjoint and such that $\bigcup_{n\in\mathbb{N}} B_n \in \mathcal{A}$. The sequence $\left(\bigcup_{k>n} B_k, n \in \mathbb{N}\right)$ is non-decreasing and converges towards $\emptyset$. The continuity property at $\emptyset$ of P implies that $\lim_{n\to+\infty} \mathrm{P}\left(\bigcup_{k>n} B_k\right) = 0$. By additivity, we get:

$$\mathrm{P}\left(\bigcup_{k\in\mathbb{N}} B_k\right) = \mathrm{P}\left(\bigcup_{k=0}^{n} B_k\right) + \mathrm{P}\left(\bigcup_{k>n} B_k\right) = \sum_{k=0}^{n} \mathrm{P}(B_k) + \mathrm{P}\left(\bigcup_{k>n} B_k\right).$$

Letting n goes to infinity, we get $\mathrm{P}\left(\bigcup_{k\in\mathbb{N}} B_k\right) = \sum_{k\in\mathbb{N}} \mathrm{P}(B_k)$.

Let $(A_n, n \in \mathbb{N})$ be a sequence of elements of $\mathcal{A}$ such that $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$. Set $B_0 = A_0$ and for $n \geq 1$, $B_n = A_n \bigcap \left(\bigcup_{k=0}^{n-1} A_k\right)^c$. We have $B_n \subset A_n$ as well as $\bigcup_{k=0}^{n} B_k = \bigcup_{k=0}^{n} A_k$

and thus $\bigcup_{k\in\mathbb{N}} B_k = \bigcup_{k\in\mathbb{N}} A_k$. The sets $(B_n, n \in \mathbb{N})$ belongs to $\mathcal{A}$, are pairwise disjoint and such that $\bigcup_{n\in\mathbb{N}} B_n \in \mathcal{A}$. we deduce from the first part of the proof that:

$$\mathrm{P}\left(\bigcup_{k\in\mathbb{N}} A_k\right) = \mathrm{P}\left(\bigcup_{k\in\mathbb{N}} B_k\right) = \sum_{k\in\mathbb{N}} \mathrm{P}(B_k) \leq \sum_{k\in\mathbb{N}} \mathrm{P}(A_k).$$

□

Let $\mathcal{A}$ be a Boolean algebra on Ω and P a probability measure on $(\Omega, \mathcal{A})$. The outer probability measure P^* is a $[0,1]$-valued function defined on $\mathcal{P}(\Omega)$ by:

$$\mathrm{P}^*(A) = \inf\left\{\sum_{n\in\mathbb{N}} \mathrm{P}(B_n); A \subset \bigcup_{n\in\mathbb{N}} B_n \quad \text{and } B_n \in \mathcal{A} \text{ for all } n \in \mathbb{N}\right\}. \tag{7.1}$$

The next lemma states that the restriction of P^* to $\mathcal{A}$ coincide with P and that P^* is monotone and σ-sub-additive.

Lemma 7.9. *We have the following properties.*

(i) $\mathrm{P}^*(A) = \mathrm{P}(A)$ *for all* $A \in \mathcal{A}$.

(ii) Monotony. For all $A \subset B \subset \Omega$, *we have* $\mathrm{P}^*(A) \leq \mathrm{P}^*(B)$.

(iii) σ-sub-additivity. Let $(A_i, i \in I)$ be a family at most countable of subsets of Ω. We have:

$$\mathrm{P}^*\left(\bigcup_{i\in I} A_i\right) \leq \sum_{i\in I} \mathrm{P}^*(A_i).$$

Proof. Let $(B_n \in \mathbb{N})$ and A be elements $\mathcal{A}$ such that $A \subset \bigcup_{n\in\mathbb{N}} B_n$. We have $A \cap B_n \in \mathcal{A}$ and $\bigcup_{n\in\mathbb{N}}(A \cap B_n) = A \in \mathcal{A}$. We deduce from property (iii) of Proposition 7.8 that:

$$\mathrm{P}(A) = \mathrm{P}\left(\bigcup_{n\in\mathbb{N}} (A \cap B_n)\right) \leq \sum_{n\in\mathbb{N}} \mathrm{P}(A \cap B_n) \leq \sum_{n\in\mathbb{N}} \mathrm{P}(B_n),$$

which is an equality if for example $B_0 = A$ and $B_n = \emptyset$ for $n \in \mathbb{N}^*$. We deduce that $\mathrm{P}^*(A) = \mathrm{P}(A)$, that is property (i).

Property (ii) is a consequence of the definition of P^*. We now prove property (iii). Let $(A_n, n \in \mathbb{N})$ be sub-sets of Ω, $\varepsilon > 0$ and $(B_{n,k}; n, k \in \mathbb{N})$ be elements of $\mathcal{A}$ such that for all $n \in \mathbb{N}$, $A_n \subset \bigcup_{k\in\mathbb{N}} B_{n,k}$ and $\sum_{k\in\mathbb{N}} \mathrm{P}(B_{n,k}) \leq \mathrm{P}^*(A_n) + \varepsilon 2^{-n}$. As $\bigcup_{n\in\mathbb{N}} A_n \subset \bigcup_{n,k\in\mathbb{N}} B_{k,n}$, we deduce that:

$$\mathrm{P}^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n,k\in\mathbb{N}} \mathrm{P}(B_{n,k}) \leq \sum_{n\in\mathbb{N}} \mathrm{P}^*(A_n) + \varepsilon 2^{-n} = 2\varepsilon + \sum_{n\in\mathbb{N}} \mathrm{P}^*(A_n).$$

Since $\varepsilon > 0$ is arbitrary, we deduce that $\mathrm{P}^*\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mathrm{P}^*(A_n)$ and thus property (iii) holds. □

The sub-additivity of P^* implies that $\mathrm{P}^*(B) \leq \mathrm{P}^*(B \cap A) + \mathrm{P}^*(B \cap A^c)$ for all $A, B \subset \Omega$. We consider the family of measurable sets for P^* defined by:

$$\mathcal{G} = \{A \subset \Omega; \quad \mathrm{P}^*(B) = \mathrm{P}^*(B \cap A) + \mathrm{P}^*(B \cap A^c) \quad \text{pour tout } B \subset \Omega\}\,. \tag{7.2}$$

We first prove that $\mathcal{G}$ is a Boolean algebra which contains $\mathcal{A}$ (Lemma 7.10), then that $\mathcal{G}$ is a σ-field and that P^* is a probability measure on $\mathcal{G}$ (Lemma 7.11).

Lemma 7.10. *The set $\mathcal{G}$ is a Boolean algebra and it contains $\mathcal{A}$.*

Proof. As $\mathrm{P}^*(\emptyset) = \mathrm{P}(\emptyset) = 0$, we deduce that $\Omega \in \mathcal{G}$. By symmetry, if $A \in \mathcal{G}$ then we have $A^c \in \mathcal{G}$. Let $A_1, A_2 \in \mathcal{G}$ and $B \in \Omega$. We have:

$$\begin{aligned}
\mathrm{P}^*(B) &= \mathrm{P}^*(B \cap A_1) + \mathrm{P}^*(B \cap A_1^c) \\
&= \mathrm{P}^*(B \cap A_1 \cap A_2) + \mathrm{P}^*(B \cap A_1 \cap A_2^c) + \mathrm{P}^*(B \cap A_1^c) \\
&\geq \mathrm{P}^*(B \cap A_1 \cap A_2) + \mathrm{P}^*(B \cap (A_1 \cap A_2)^c),
\end{aligned} \tag{7.3}$$

where we used the sub-additivity of P^* for the inequality and $(A_1 \cap A_2^c) \bigcup A_1^c = (A_1 \cap A_2)^c$. As P^* is sub-additive, we deduce that the inequality (7.3) is in fact an equality, and thus that $A_1 \cap A_2 \in \mathcal{G}$. This implies that $\mathcal{G}$ is a Boolean algebra.

Let $A \in \mathcal{A}$, $B \in \Omega$ and $\varepsilon > 0$. There exists a sequence $(B_n, n \in \mathbb{N})$ of elements of $\mathcal{A}$ such that $B \subset \bigcup_{n\in\mathbb{N}} B_n$ and $\mathrm{P}^*(B) + \varepsilon \geq \sum_{n\in\mathbb{N}} \mathrm{P}(B_n)$. By additivity of P, we get:

$$\mathrm{P}^*(B) + \varepsilon \geq \sum_{n\in\mathbb{N}} \mathrm{P}(B_n) = \sum_{n\in\mathbb{N}} \mathrm{P}(B_n \cap A) + \sum_{n\in\mathbb{N}} \mathrm{P}(B_n \cap A^c) \geq \mathrm{P}^*(B \cap A) + \mathrm{P}^*(B \cap A^c),$$

where for the last equality we used the definition of P^* and that $B \cap A \subset \bigcup_{n\in\mathbb{N}} B_n \cap A$ as well as $B \cap A^c \subset \bigcup_{n\in\mathbb{N}} B_n \cap A^c$. Since $\varepsilon > 0$ is arbitrary, we deduce that $\mathrm{P}^*(B) \geq \mathrm{P}^*(B \cap A) + \mathrm{P}^*(B \cap A^c)$, and then that this inequality is an equality as P^* is sub-additive. Thus, we get that $A \in \mathcal{G}$ if $A \in \mathcal{A}$. □

Lemma 7.11. *The family $\mathcal{G}$ is a σ-field and the function* P^* *restricted to $\mathcal{G}$ is a probability measure.*

Proof. Notice that for $A \in \mathcal{G}$ and $B, C \in \Omega$ such that $A \cap C = \emptyset$, we deduce from the definition of $\mathcal{G}$ that:

$$\mathrm{P}^*\left(B \cap (A \cup C)\right) = \mathrm{P}^*(B \cap A) + \mathrm{P}^*(B \cap C). \tag{7.4}$$

Let $(A_n, n \in \mathbb{N})$ be elements of $\mathcal{G}$ pairwise disjoint and $B \in \Omega$. We set $A'_n = \bigcup_{k=0}^n A_k$ and $A' = \bigcup_{k\in\mathbb{N}} A_k$. We have $A'_n \in \mathcal{A}$. Using the monotonicity of P^* and then (7.4), we get:

$$\mathrm{P}^*(B \cap A') \geq \mathrm{P}^*(B \cap A'_n) = \sum_{k=0}^n \mathrm{P}^*(B \cap A_k). \tag{7.5}$$

We deduce that $\mathrm{P}^*(B \cap A') \geq \sum_{k\in\mathbb{N}} \mathrm{P}^*(B \cap A_k)$ and then, since P^* is σ-sub-additive, that:

$$\mathrm{P}^*(B \cap A') = \sum_{k\in\mathbb{N}} \mathrm{P}^*(B \cap A_k). \tag{7.6}$$

We deduce from the equality in (7.5) and the monotonicity of P^* that:

$$\mathrm{P}^*(B) = \mathrm{P}^*(B \cap A'_n) + \mathrm{P}^*(B \cap {A'_n}^c) \geq \sum_{k=0}^{n} \mathrm{P}^*(B \cap A_k) + \mathrm{P}^*(B \cap A'^c).$$

Letting n goes to infinity, we deduce from (7.6) that $\mathrm{P}^*(B) \geq \mathrm{P}^*(B \cap A') + \mathrm{P}^*(B \cap A'^c)$. Since P^* is sub-additive, this inequality is in fact an equality, and thus $A' = \bigcup_{k\in\mathbb{N}} A_n \in \mathcal{G}$. It is then immediate to check that $\mathcal{G}$ is stable by countable union. Thus, $\mathcal{G}$ is a σ-field.

For $B = \Omega$ in (7.6), we get that P^* is σ-additive on $\mathcal{G}$: $\mathrm{P}^*\left(\bigcup_{k\in\mathbb{N}} A_k\right) = \sum_{k\in\mathbb{N}} \mathrm{P}^*(A_k)$. The restriction of P^* to $\mathcal{G}$ is therefore a probability measure. □

Proof of Theorem 7.3. Let P be a probability measure on a Boolean algebra $\mathcal{A}$ of Ω. According to Lemma 7.11, the family of sets $\mathcal{G}$ defined by (7.2) is a σ-field and the restriction of P^*, defined by (7.1), on $\mathcal{G}$ is a probability measure. Since $\mathcal{G}$ is a σ-field, we deduce from Lemma 7.10 that $\sigma(\mathcal{A}) \subset \mathcal{G}$. According to Lemma 7.9 we have that P^* and P coincide on $\mathcal{A}$. We deduce that the restriction of P^* to $\sigma(\mathcal{A})$ is a probability measure on $\sigma(\mathcal{A})$ which coincide with P on $\mathcal{A}$.

Uniqueness of this extension on $\sigma(\mathcal{A})$ is a consequence of the monotone class theorem and more precisely Corollary 1.14. □

7.2 More on convergence for sequence of random variables

All the random variables of the section will be defined on a given probability space $(\Omega, \mathcal{G}, \mathbb{P})$.

7.2.1 Convergence in distribution

The convergence in distribution for random variables correspond to the weak convergence of their distribution. We shall investigate this subject only partially so that we can state the ergodic theorems for Markov chains.

Definition 7.12. *Let $(E, \mathcal{B})$ be a metric space with its Borel σ-field. A sequence $(\mu_n, n \in \mathbb{N})$ of probability measures on E converges weakly to a probability measure μ on E if for all bounded real-valued continuous function f defined on E, we have* $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$.

Let $(X_n, n \in \mathbb{N})$ and X be E-valued random variables. The sequence $(X_n, n \in \mathbb{N})$ converges in distribution towards X if the probability measures $(\mathrm{P}_{X_n}, n \in \mathbb{N})$ converges weakly towards P_X, that is $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ *for all bounded real-valued continuous function. And we write $X_n \xrightarrow[n\to\infty]{(d)} X$ (or some times $X_n \xrightarrow[n\to\infty]{(d)} \mathrm{P}_X$).*

We refer to [1] for further results on convergence in distribution. Since we shall be mainly interested by the convergence in distribution for random variables taking values in a discrete space we introduce the convergence for the distance in total variation. The distance in total variation d_{TV} between two finite measures μ and ν on $(S, \mathcal{S})$ is given by:

$$d_{\mathrm{TV}}(\mu, \nu) = \sup_{A\in\mathcal{S}} |\mu(A) - \nu(A)|.$$

It is elementary to check that d_{TV} is indeed a distance[1] on the set of finite measures on $(S, \mathcal{S})$.

Lemma 7.13. *The convergence for the distance in total variation for probability measures on a metric space implies the convergence in distribution.*

Proof. Let $(E, \mathcal{B}(E))$ be a metric space with its Borel σ-field. Let f be a real-valued measurable function defined on E taking values in $(0,1)$. By Fubini theorem, we have that $\int f \, \mathrm{d}\nu = \int_0^1 \nu(\{f > t\}) \, \mathrm{d}t$ for any probability measure ν on $(E, \mathcal{B}(E))$.

Let $(\mu_n, n \in \mathbb{N})$ be a sequence of probability measures which converges for the distance in total variation towards a probability measure μ. This implies that $\lim_{n\to\infty} \mu_n(\{f > t\}) = \mu(\{f > t\})$. By dominated convergence, we deduce from the comment at the beginning of the proof, that $\lim_{n\to\infty} \int f \, \mathrm{d}\mu_n = \int f \, \mathrm{d}\mu$. By linearity, we get this convergence holds for any bounded real-valued measurable function f. This implies that $(\mu_n, n \in \mathbb{N})$ converges weakly towards μ. $\square$

The discrete state space case

We now assume that E is a discrete space (and $\mathcal{E} = \mathcal{P}(E)$). Let λ denote the counting measure on $(E, \mathcal{E})$: $\lambda(A) = \text{Card }(A)$ for $A \in \mathcal{E}$. Notice that any measure μ on $(E, \mathcal{E})$ has a density with respect to the counting measure λ given by the function $(\mu(\{x\}), x \in E)$. We shall identify the density of μ with μ and thus write $\mu(x)$ for $\mu(\{x\})$. We shall consider the L^1 norm with respect to the counting measure so that for a real-valued function $f = (f(x), x \in E)$ we set $\|f\|_1 = \sum_{x\in E} |f(x)|$. It is left to the reader to check that for two finite measures μ and ν on $(E, \mathcal{E})$:

$$2\, d_{\mathrm{TV}}(\mu, \nu) = \|\mu - \nu\|_1 \,. \tag{7.7}$$

We now give a characterization of the weak convergence of probability measure on discrete space.

Lemma 7.14. *Let E be a discrete space. Let $(X_n, n \in \mathbb{N}^*)$ and X be E-valued random variables. The following conditions are equivalent.*

(i) $X_n \xrightarrow[n\to\infty]{\textbf{(d)}} X$.

(ii) For all $x \in E$, we have $\mathbb{P}(X_n = x) \xrightarrow[n\to\infty]{} \mathbb{P}(X = x)$.

(iii) $\lim_{n\to\infty} d_{TV}(\mathrm{P}_{X_n}, \mathrm{P}_X) = 0$.

Proof. Since $\{x\}$ is open and closed, the function $\mathbf{1}_{\{x\}}$ is continuous. Thus, property (i) implies property (ii). Property (iii) implies property (i) thanks to Lemma 7.13.

We now prove that property (ii) implies property (iii). Let $\varepsilon > 0$ and $K \subset E$ finite such that $\mathbb{P}(X \in K) \geq 1 - \varepsilon$. Since K is finite, we deduce from (ii) that $\lim_{n\to\infty} \mathbb{P}(X_n \in K) =$

[1] A non-negative finite function d defined on $S \times S$ is a distance on S if for all $x, y, z \in S$, we have: $d(x,y) = d(y,x)$ (symmetry); $d(x,y) \leq d(x,z) + d(z,y)$ (triangular inequality); $d(x,y) = 0$ implies $x = y$ (separation).

$\mathbb{P}(X \in K)$ and thus $\lim_{n\to\infty} \mathbb{P}(X_n \in K^c) = \mathbb{P}(X \in K^c) \leq \varepsilon$. So for n large enough, say $n \geq n_0$, we have $\mathbb{P}(X_n \in K^c) \leq 2\varepsilon$. We deduce from (7.7) that for $n \geq n_0$:

$$\begin{aligned} 2\, d_{\mathrm{TV}}(\mathrm{P}_{X_n}, \mathrm{P}_X) &= \sum_{x\in E} |\mathbb{P}(X_n = x) - \mathbb{P}(X = x)| \\ &\leq \sum_{x\in K} |\mathbb{P}(X_n = x) - \mathbb{P}(X = x)| + \mathbb{P}(X_n \in K^c) + \mathbb{P}(X \in K^c) \\ &\leq \sum_{x\in K} |\mathbb{P}(X_n = x) - \mathbb{P}(X = x)| + 3\varepsilon. \end{aligned}$$

This implies that $\limsup_{n\to\infty} 2\, d_{\mathrm{TV}}(\mathrm{P}_{X_n}, \mathrm{P}_X) \leq 3\varepsilon$. Conclude using that ε is arbitrary. □

7.2.2 Law of large number and central limit theorem

We refer to any introductory books in probability for a proof of the next results on the law of large number and central limit theorem (CLT). For a sequence $(X_n, n \in \mathbb{N}^*)$ of real-valued or $\mathbb{R}^d$-valued random variables, we define the empirical mean, when it is meaningful, by:

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

Theorem 7.15 (Law of large number)**.** *Let X be a real-valued random variable such that $\mathbb{E}[X]$ is well defined. Let $(X_n, n \in \mathbb{N}^*)$ be a sequence of independent real-valued random variables distributed as X. We have the following a.s. converge:*

$$\bar{X}_n \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}[X].$$

If $X \in L^1$, then the converge holds also in L^1.

The fluctuation are given by the CLT. We denote by $\mathcal{N}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ a symmetric non-negative matrix, the Gaussian distribution with mean μ and covariance matrix Σ.

Theorem 7.16 (Central Limit Theorem (CLT))**.** *Let X be a $\mathbb{R}^d$-valued random variable such that $X \in L^2$. Set $\mu = \mathbb{E}[X]$ and $\Sigma = \mathrm{Cov}(X, X)$. Let $(X_n, n \in \mathbb{N}^*)$ be a sequence of independent real-valued random variables distributed as X. We have the following convergences:*

$$\bar{X}_n \xrightarrow[n\to\infty]{\text{a.s.}} \mu \quad \text{and} \quad \sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow[n\to\infty]{\text{(d)}} \mathcal{N}(0, \Sigma).$$

7.2.3 Uniform integrability

Let $(\Omega, \mathcal{G}, \mathbb{P})$ be a probability space.

Definition 7.17. *We say a family of real-valued random variables $(X_i, i \in I)$ is uniformly integrable if for all $\varepsilon > 0$, there exists K finite such that for all $i \in I$:*

$$\mathbb{E}\left[|X_i| \mathbf{1}_{\{|X_i|\geq K\}}\right] \leq \varepsilon. \tag{7.8}$$

Notice that if (7.8) holds, then $\mathbb{E}[|X_i|] \leq \varepsilon + K$ and thus $\sup_{i\in I} \mathbb{E}[|X_i|]$ is finite.

We give some results related to the uniform integrability.

Proposition 7.18. *Let $(X_i, i \in I)$ be a family of real-valued random variables.*

(i) The family $(X_i, i \in I)$ is uniformly integrable if and only if the following two conditions are satisfied:

(a) For all $\varepsilon > 0$, there exists $\delta > 0$ such that for all events A with $\mathbb{P}(A) \leq \delta$, we have $\mathbb{E}[|X_i|\mathbf{1}_A] \leq \varepsilon$.

(b) $\sup_{i\in I} \mathbb{E}[|X_i|] < +\infty$.

(ii) Any single real-valued integrable random variable is uniformly integrable.

(iii) If there exists an integrable real-valued random variable Y such that $|X_i| \leq |Y|$ a.s. for all $i \in I$, then the family $(X_i, i \in I)$ is uniformly integrable. More generally, if there exists a family $(Y_i, i \in I)$ of uniformly integrable real-valued random variables such that $|X_i| \leq |Y_i|$ a.s. for all $i \in I$, then the family $(X_i, i \in I)$ is uniformly integrable.

(v) If there exists $r > 0$ such that $\sup_{i\in I} \mathbb{E}[|X_i|^{1+r}] < +\infty$, then the family $(X_i, i \in I)$ is uniformly integrable. More generally, if $\sup_{i\in I} \mathbb{E}[f(X_i)] < +\infty$, where f is a non-negative real-valued measurable function defined on $\overline{\mathbb{R}}$ such that $\lim_{x\to+\infty} f(x)/x = +\infty$, then the family $(X_i, i \in I)$ is uniformly integrable.

(vi) If $(X_n, n \in \mathbb{N})$ is a sequence of integrable real-valued random variables which converges in L^1 to zero, then it is uniformly integrable.

Proof. We first prove property (i). Assume that the family $(X_i, i \in I)$ is uniformly integrable. We have already noticed that (b) holds. Choose K such that (7.8) holds with ε replaced by $\varepsilon/2$. Set $\delta = \varepsilon/2K$ and let A be an event such that $\mathbb{P}(A) \leq \delta$. Using (7.8) and Markov inequality, we get:

$$\mathbb{E}[|X_i|\mathbf{1}_A] = \mathbb{E}\left[|X_i|\mathbf{1}_A\mathbf{1}_{\{|X_i|\geq K\}}\right] + \mathbb{E}\left[|X_i|\mathbf{1}_A\mathbf{1}_{\{|X_i|<K\}}\right] \leq \frac{\varepsilon}{2} + K\mathbb{P}(A) \leq \varepsilon.$$

This gives (a).

Assume that (a) and (b) hold. Set $C = \sup_{i\in I} \mathbb{E}[|X_i|]$ which is finite by (b). Let $\varepsilon > 0$ be fixed and δ given by (a). Set $K = C/\delta$ and $A_i = \{|X_i| \geq K\}$. Markov inequality gives that $\mathbb{P}(A_i) \leq \mathbb{E}[|X_i|]/K \leq C/K = \delta$. We deduce from (a), with A replaced by A_i, that (7.8) holds. This implies that the family $(X_i, i \in I)$ is uniformly integrable.

We prove property (ii). Let Y be an integrable real-valued random variable. In particular Y is a.s. finite. By dominated convergence, we get that $\lim_{K\to+\infty} \mathbb{E}\left[|Y|\mathbf{1}_{\{|Y|\geq K\}}\right] = 0$. Thus (7.8) holds and Y is uniformly integrable.

Thanks to property (i), to prove property (iii) it is enough to check (a) and (b). Notice that $\mathbb{E}[|X_i|] \leq \mathbb{E}[|Y|]$ for all $i \in I$ and thus (b) holds. We have $\mathbb{E}[|X_i|\mathbf{1}_A] \leq \mathbb{E}[|Y|\mathbf{1}_A]$. Then use that Y is uniformly integrable, thanks to (ii), to conclude that (a) holds. The proof of the more general case is similar.

We prove property (v). Let $\varepsilon > 0$. Use that $|X_i|\mathbf{1}_{\{|X_i|\geq K\}} \leq K^{-r}|X_i|^{1+r}$ to deduce that (7.8) holds when $K = (\sup_{i\in I} \mathbb{E}[|X_i|^{1+r}]/\varepsilon)^{1/r}$. The proof of the more general case is similar.

Thanks to property (i), to prove property (vi) it is enough to check (a) and (b). Let $(X_n, n \in \mathbb{N})$ be a sequence of integrable real-valued random variables which converges in L^1 towards zero. Condition (b) is immediate. Let us check (a). Fix $\varepsilon > 0$. There exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, we have $\mathbb{E}[|X_n|] \leq \varepsilon$. Then use (ii) and (i) to get there exists $\delta_i > 0$ such that if A is an event such that $\mathbb{P}(A) \leq \delta_i$ then $\mathbb{E}[|X_i|\mathbf{1}_A] \leq \varepsilon$ for all $i \leq n_0$. Take $\delta = \min_{0\leq i\leq n_0} \delta_i$ to deduce that (a) holds. □

We provide an interesting example of family of uniform integrable random variables.

Lemma 7.19. *Let X be an integrable real-valued random variable. The family $(X_{\mathcal{H}} = \mathbb{E}[X|\,\mathcal{H}]$; $\mathcal{H}$ is a σ-field and $\mathcal{H} \subset \mathcal{G})$ is uniformly integrable.*

Proof. We shall check (a) and (b) from property (i) of Proposition 7.18. Using Jensen inequality, we get that $\mathbb{E}[|X_{\mathcal{H}}|] \leq \mathbb{E}[|X|]$ for all σ-field $\mathcal{H} \subset \mathcal{G}$. We get (b) as X is integrable.

We prove (a). Let $\varepsilon > 0$. According to property (ii) from Proposition 7.18, we get that X is uniformly integrable. Thus, there exists K such that $\mathbb{E}\left[|X|\mathbf{1}_{\{|X|\geq K\}}\right] \leq \varepsilon/2$. Let $A \subset \mathcal{G}$ be such that $\mathbb{P}(A) \leq \varepsilon/2K$. For any σ-field $\mathcal{H} \subset \mathcal{G}$, we have using Jensen inequality:

$$\mathbb{E}\left[|X_{\mathcal{H}}|\mathbf{1}_A\right] = \mathbb{E}\left[|X_{\mathcal{H}}|\mathbb{E}\left[\mathbf{1}_A|\,\mathcal{H}\right]\right] \leq \mathbb{E}\Big[\mathbb{E}\left[|X|\,|\,\mathcal{H}\right]\mathbb{E}\left[\mathbf{1}_A|\,\mathcal{H}\right]\Big] = \mathbb{E}\Big[|X|\,\mathbb{E}\left[\mathbf{1}_A|\,\mathcal{H}\right]\Big].$$

Furthermore decomposing according to $\{|X| \geq K\}$ and $\{|X| < K\}$, and using that a.s. $0 \leq \mathbb{E}\left[\mathbf{1}_A|\,\mathcal{H}\right] \leq 1$, we get:

$$\mathbb{E}\Big[|X|\,\mathbb{E}\left[\mathbf{1}_A|\,\mathcal{H}\right]\Big] \leq \mathbb{E}\left[|X|\mathbf{1}_{\{|X|\geq K\}}\right] + K\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_A|\,\mathcal{H}\right]\right] \leq \frac{\varepsilon}{2} + K\mathbb{P}(A) \leq \varepsilon.$$

We have obtained that $\mathbb{E}\left[|X_{\mathcal{H}}|\mathbf{1}_A\right] \leq \varepsilon$ for all σ-field $\mathcal{H} \subset \mathcal{G}$ and all $A \subset \mathcal{G}$ such that $\mathbb{P}(A) \leq \varepsilon/2K$. This gives (a). □

7.2.4 Convergence in probability and in L^1

We recall that a sequence $(X_n, n \in \mathbb{N})$ of real-valued random variables converges in probability towards a real-valued random variable X_∞ if for all $\varepsilon > 0$, we have $\lim_{n\to\infty} \mathbb{P}(|X_n - X_\infty| \geq \varepsilon) = 0$. We also recall that the a.s. convergence implies the convergence in probability. The converse is false in general, but we have the following partial converse. We call $(a_{n_k}, n \in \mathbb{N})$ a sub-sequence of $(a_n, n \in \mathbb{N})$ if the $\mathbb{N}$-valued sequence $(n_k, k \in \mathbb{N})$ is increasing.

Lemma 7.20. *Let $(X_n, n \in \mathbb{N})$ be a sequence of real-valued random variables which converges in probability towards a real-valued random variable X_∞. Then, there is a sub-sequence $(X_{n_k},\, k \in \mathbb{N})$ which converges a.s. to X_∞.*

The proof of this lemma is a consequence of the Borel-Cantelli lemma, but we provide a direct short proof (see also the proof of Proposition 1.50 where similar arguments are used).

Proof. Let $n_0 = 0$, and for $k \in \mathbb{N}$, set $n_{k+1} = \inf\{n > n_k;\, \mathbb{P}(|X_n - X_\infty| \geq 2^{-k}) \leq 2^{-k}\}$. The sub-sequence $(n_k, k \in \mathbb{N})$ is well defined since $(X_n, n \in \mathbb{N})$ converges in probability towards X_∞. Since $\sum_{k\in\mathbb{N}} \mathbb{P}(|X_{n_k} - X_\infty| \geq 2^{-k}) < +\infty$, we get that $\sum_{k\in\mathbb{N}} \mathbf{1}_{\{|X_{n_k} - X_\infty| \geq 2^{-k}\}}$ is a.s. finite and thus a.s. $|X_{n_k} - X_\infty| \geq 2^{-k}$ for finitely many k. This implies that the sub-sequence $(X_{n_k},\, k \in \mathbb{N})$ converges a.s. to X_∞. □

The uniform integrability is the right concept to get the L^1 convergence from the a.s. convergence of real-valued random variables. This is a consequence of the following proposition.

Proposition 7.21. *Let $(X_n, n \in \mathbb{N})$ be a sequence of integrable real-valued random variables and X_∞ a real-valued random variable. The following properties are equivalent.*

(i) The random variables $(X_n, n \in \mathbb{N})$ are uniformly integrable and $(X_n, n \in \mathbb{N})$ converges in probability towards X_∞.

(ii) The sequence $(X_n, n \in \mathbb{N})$ converges in L^1 towards X_∞ which is integrable.

Proof. We first assume (i). Thanks to Lemma 7.20, there exists a sub-sequence $(X_{n_k},\, k \in \mathbb{N})$ which converges a.s. to X_∞. As $(|X_{n_k}|,\, k \in \mathbb{N})$ converges a.s. to $|X_\infty|$, we deduce from Fatou's lemma that $\mathbb{E}\left[|X_\infty|\right] \leq \liminf_{k\to\infty} \mathbb{E}\left[|X_{n_k}|\right] \leq \sup_{n\in\mathbb{N}} \mathbb{E}\left[|X_n|\right]$. Since the random variables $(X_n, n \in \mathbb{N})$ are uniformly integrable, we deduce from property (i)-(b) of Proposition 7.18 that X_∞ is integrable.

Let $\varepsilon > 0$. Since the random variables $(X_n, n \in \mathbb{N})$ are uniformly integrable as well as X_∞, thanks to property (ii) of Proposition 7.18, we deduce there exists $\delta > 0$ such that if A is an event with $\mathbb{P}(A) \leq \delta$, then $\mathbb{E}[|X_n|\mathbf{1}_A] \leq \varepsilon$ for all $n \in \overline{\mathbb{N}}$. Since $(X_n, n \in \mathbb{N})$ converges in probability towards X_∞, there exists n_0 such that for $n \geq n_0$, we have $\mathbb{P}(|X_n - X_\infty| > \varepsilon) \leq \delta$. This implies that for $n \geq n_0$:

$$\begin{aligned}\mathbb{E}\left[|X_n - X_\infty|\right] &= \mathbb{E}\left[|X_n - X_\infty|\mathbf{1}_{\{|X_n - X_\infty| \leq \varepsilon\}}\right] + \mathbb{E}\left[|X_n - X_\infty|\mathbf{1}_{\{|X_n - X_\infty| > \varepsilon\}}\right] \\ &\leq \varepsilon + \mathbb{E}\left[|X_n|\mathbf{1}_{\{|X_n - X_\infty| > \varepsilon\}}\right] + \mathbb{E}\left[|X_\infty|\mathbf{1}_{\{|X_n - X_\infty| > \varepsilon\}}\right] \leq 3\varepsilon.\end{aligned}$$

This gives (ii).

We now assume (ii). First recall the convergence in L^1 implies by Markov inequality the convergence in probability. Since $(X_n - X_\infty, n \in \mathbb{N})$ converges in L^1 towards 0, we deduce from property (vi) of Proposition 7.18 that this sequence is uniformly integrable. Since X_∞ is integrable, it is uniformly integrable according to property (ii) of Proposition 7.18. Using that $|X_n| \leq |X_n - X_\infty| + |X_\infty|$ and the characterization of the uniform integrability given by property (i) of Proposition 7.18, we easily get that $(X_n, n \in \mathbb{N})$ is uniformly integrable. □

Bibliography

[1] P. Billingsley. *Probability and measure.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, 2012.

[2] J. Neveu. *Bases mathématiques du calcul des probabilités.* Masson et Cie, Éditeurs, Paris, 1970.

Chapter 8

Exercises

8.1 Measure theory and random variables

Exercise 8.1 (Characterization of probability measures on $\mathbb{R}$). Let $\mathbb{P}$ and $\mathbb{P}'$ be two probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mathbb{P}((-\infty, a]) = \mathbb{P}'((-\infty, a])$ for all a in a dense subset of $\mathbb{R}$. Prove that $\mathbb{P} = \mathbb{P}'$. △

Exercise 8.2 (Characterization of σ-finite measures). Extend Corollary 1.14 (resp. Corollary 1.15) to σ-finite measures by assuming further that the sequence $(\Omega_n, n \in \mathbb{N})$ from (iii) of Definition 1.7 can be chosen to belong to $\mathcal{C}$ (resp. to $\mathcal{O}$). △

Exercise 8.3 (Limit of differences). Let $(a_n, n \in \mathbb{N})$ and $(b_n, n \in \mathbb{N})$ be real-valued sequence such that (a_n, b_n) for all $n \in \mathbb{N}$ as well as $(\limsup_{n\to\infty} a_n, \liminf_{n\to\infty} b_n)$ are not equal to $(+\infty, +\infty)$ nor $(-\infty, -\infty)$. Prove that:

$$\limsup_{n\to\infty}(a_n - b_n) \leq \limsup_{n\to\infty} a_n - \liminf_{n\to\infty} b_n.$$

If furthermore the sequence $(b_n, n \in \mathbb{N})$ converges, deduce that:

$$\limsup_{n\to\infty}(a_n - b_n) = \limsup_{n\to\infty} a_n - \lim_{n\to\infty} b_n.$$

△

Exercise 8.4 (Permutation of integrals). Prove that:

$$\int_{(0,1)} \left(\int_{(0,1)} \frac{x^2 - y^2}{(x^2 + y^2)^2} \, \lambda(\mathrm{d}y) \right) \mathrm{d}x = \frac{\pi}{4}.$$

Deduce that the function $f(x, y) = \dfrac{x^2 - y^2}{(x^2 + y^2)^2}$ is not integrable with respect to the Lebesgue measure on $(0, 1)^2$. (Hint: compute the derivative with respect to y of $y/(x^2 + y^2)$.) △

Exercise 8.5 (Independence). Extend (1.13) to functions f_j such that $f_j \geq 0$ for all $j \in J$ or to functions f_j such that $f_j(X_j)$ is integrable for all $j \in J$. And in the latter case $\prod_{j\in J} f_j(X_j)$ is also integrable. △

Exercise 8.6 (Independence and covariance). Let X and Y be real-valued integrable random variables. Prove that if X and Y are independent, then XY is integrable and $\mathrm{Cov}(X, Y) = 0$. Give an example such that X and Y are square-integrable not independent but with $\mathrm{Cov}(X, Y) = 0$. △

Exercise 8.7 (Independence). Let $(A_i, i \in I)$ be independent events. Prove that $(\mathbf{1}_{A_i}, i \in I)$ are independent random variables and deduce that $(A_i^c, i \in I)$ are also independents events. △

Exercise 8.8 (Independence). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{G} \subset \mathcal{F}$ be a σ-field and $\mathcal{C}$ a collection of events which are all independents of $\mathcal{G}$.

1. Prove by a counterexample that if $\mathcal{C}$ is not stable by finite intersection, then $\sigma(\mathcal{C})$ may not be independent of $\mathcal{G}$.

2. Using the monotone class theorem prove that if $\mathcal{C}$ is stable by finite intersection, then $\sigma(\mathcal{C})$ is independent of $\mathcal{G}$.

3. Let $\mathcal{C}$ and $\mathcal{C}'$ be two collections of events stable by finite intersection such that every $A \in \mathcal{C}$ and $A' \in \mathcal{C}'$ are independent. Prove that $\sigma(\mathcal{C})$ and $\sigma(\mathcal{C}')$ are independent.

△

8.2 Conditional expectation

Exercise 8.9 (Indicators). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{F}$ be two events. Describe $\sigma(\mathbf{1}_B)$ and then compute $\mathbb{E}[\mathbf{1}_A | \mathbf{1}_B]$. △

Exercise 8.10 (Random walk). Let $(X_n, n \in \mathbb{N}^*)$ be identically distributed independent real-valued random variables such that $\mathbb{E}[X_1]$ is well defined. Let $S_n = \sum_{k=1}^n X_k$ for $k \in \mathbb{N}^*$. Compute $\mathbb{E}[X_1 | S_2]$ and deduce $\mathbb{E}[X_1 | S_n]$ for $n \geq 2$. △

Exercise 8.11 (Symmetric random variable). Let X be a real-valued random variable integrable and symmetric, that is X and $-X$ have the same distribution. Compute $\mathbb{E}\left[X | X^2\right]$. △

Exercise 8.12 (X conditioned on $|X|$). Let X be an integrable real-valued random variable with density f with respect to the Lebesgue measure. Compute $\mathbb{E}[X | \, |X|]$. Compute also $\mathbb{E}\left[X | \, X^2\right]$. △

Exercise 8.13 (Variance). Let X be a real-valued random variable such that $\mathbb{E}[X^2] < +\infty$. Let $\mathcal{H}$ be a σ-field. Prove that $\mathbb{E}[X | \, \mathcal{H}]^2$ is integrable and $\mathrm{Var}(\mathbb{E}[X | \, \mathcal{H}]) \leq \mathrm{Var}(X)$. △

Exercise 8.14 (L^1 distance). Let X, Y be independent $\mathbb{R}$-valued integrable random variables such that $\mathbb{E}[Y] = 0$. Prove that $\mathbb{E}[|X - Y|] \geq \mathbb{E}[|X|]$. △

Exercise 8.15 (Kolmogorov's maximal inequality). Let $(X_n, n \in \mathbb{N}^*)$ be identically distributed independent real-valued random variables. We assume that $\mathbb{E}[X_1^2] < +\infty$ and $\mathbb{E}[X_1] = 0$. Let $x > 0$. We set $S_n = \sum_{k=1}^n X_k$ for $n \in \mathbb{N}^*$ and $T = \inf\{n \in \mathbb{N}^*; |S_n| \geq x\}$ with the convention that $\inf \emptyset = +\infty$.

1. Prove that $\mathbb{P}(T = k) \leq \frac{1}{x^2} \mathbb{E}\left[S_k^2 \mathbf{1}_{\{T=k\}}\right]$ for all $k \in \mathbb{N}^*$.

2. Check that $\sum_{k=1}^n \mathbb{P}(T=k) = \mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq x\right)$.

3. By noticing that $S_n^2 \geq S_k^2 + 2S_k(S_n - S_k)$, prove Kolmogorov's maximal inequality:

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k| \geq x\right) \leq \frac{\mathbb{E}[S_n^2]}{x^2} \qquad \text{for all } x > 0 \text{ and } n \in \mathbb{N}^*.$$

△

Exercise 8.16 (An application of Jensen inequality). Let X and Y be two integrable real-valued random variables such that a.s. $\mathbb{E}[X|\, Y] = Y$ and $\mathbb{E}[Y|\, X] = X$. Using Jensen inequality (twice) with a positive strictly convex function φ such that $\lim_{x\to+\infty} \varphi(x)/x$ and $\lim_{x\to-\infty} \varphi(x)/x$ are finite, prove that a.s. $X = Y$. △

Exercise 8.17 (Independence and conditional expectation). Let $\mathcal{H} \subset \mathcal{F}$ be a σ-field, Y and V random variables taking values in measurable spaces $(S, \mathcal{S})$ and $(E, \mathcal{E})$ such that Y is independent of $\mathcal{H}$ and V is $\mathcal{H}$-measurable. Let φ be a non-negative real-valued measurable function defined on $S \times E$ (endowed with the product σ-field). Prove that a.s.:

$$\mathbb{E}[\varphi(Y,V)|\, \mathcal{H}] = g(V) \quad \text{with} \quad g(v) = \mathbb{E}[\varphi(Y,v)]. \tag{8.1}$$

△

Exercise 8.18 (Conditional independence). Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{H}$ be σ-fields, subsets of $\mathcal{F}$. Assume that $\mathcal{H} \subset \mathcal{A} \cap \mathcal{B}$ and that conditionally on $\mathcal{H}$ the σ-fields $\mathcal{A}$ and $\mathcal{B}$ are independent, that is for all $A \in \mathcal{A}$, $B \in \mathcal{B}$, we have a.s. $\mathbb{P}(A \cap B|\, \mathcal{H}) = \mathbb{P}(A|\, \mathcal{H})\mathbb{P}(B|\, \mathcal{H})$.

1. Let $A \in \mathcal{A}$, $B \in \mathcal{B}$. Check that a.s. $\mathbb{E}\left[\mathbf{1}_A \mathbb{E}\left[\mathbf{1}_B|\, \mathcal{A}\right] |\, \mathcal{H}\right] = \mathbb{E}\left[\mathbf{1}_A \mathbb{E}\left[\mathbf{1}_B|\, \mathcal{H}\right] |\, \mathcal{H}\right]$.

2. Deduce that, for all $B \in \mathcal{B}$, a.s. $\mathbb{P}(B|\, \mathcal{A}) = \mathbb{P}(B|\, \mathcal{H})$.

△

Exercise 8.19 (Convergence of conditional expectation). Let $B = (B_n, n \in \mathbb{N}^*)$ and $B' = (B'_n, n \in \mathbb{N}^*)$ be independent sequences of independent random variables such that B_n and B'_n are Bernoulli with parameter $1/n$. Let $\mathcal{H} = \sigma(B)$. We set $X_n = nB_nB'_n$ for $n \in \mathbb{N}^*$.

1. Prove that $(X_n, n \in \mathbb{N}^*)$ converges a.s. and in L^1 towards 0.

2. Prove that $(\mathbb{E}\left[X_n|\, \mathcal{H}\right], n \in \mathbb{N}^*)$ converges in L^1 but not a.s. towards 0.

△

Exercise 8.20 (Conditional densities). Let (Y, V) be an $\mathbb{R}^2$-valued random variable whose law has density $f_{(Y,V)}(y, v) = \lambda v^{-1}\, \mathrm{e}^{-\lambda v}\, \mathbf{1}_{\{0<y<v\}}$ with respect to the Lebesgue measure on $\mathbb{R}^2$. Check that the law of Y conditionally on V is the uniform distribution on $[0, V]$. For a real-valued measurable bounded function φ defined on $\mathbb{R}$, deduce that $\mathbb{E}[\varphi(Y)|V] = V^{-1}\int_0^V \varphi(y)\, \mathrm{d}y$. △

Exercise 8.21 (Conditional distribution and independence). Let (Y, V) be an $S \times E$-valued random variable. Prove that Y and V are independent if and only if the conditional distribution of Y given V exists and is given by a kernel, say κ, such that $\kappa(v, \mathrm{d}y)$ does not depend on $v \in E$. In this case, check that $\kappa(v, \mathrm{d}y) = \mathrm{P}_Y(\mathrm{d}y)$. △

8.3 Discrete Markov chains

Exercise 8.22 (Markov chains built from a Markov chain-I). Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain on a finite or countable set E with transition matrix P. Set $Z = (Z_n = X_{2n}, n \in \mathbb{N})$.

1. Compute $\mathbb{P}(X_2 = y|X_0 = x)$ for $x, y \in E$. Prove that Z is a Markov chain and gives its transition matrix.

2. Prove that any invariant probability measure for X is also invariant for Z. Prove the converse is false in general.

$\triangle$

Exercise 8.23 (Markov chains built from a Markov chain-II). Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain on a finite or countable set E with transition matrix P. Set $Y = (Y_n, n \in \mathbb{N}^*)$ where $Y_n = (X_{n-1}, X_n)$.

1. Prove that Y is a Markov chain on E^2 and give its transition matrix.

2. Give an example with X irreducible and Y nor irreducible. If X is irreducible, change the state space of Y so that it is also irreducible.

3. Let π be an invariant probability distribution of X. Deduce an invariant probability distribution for Y.

$\triangle$

Exercise 8.24 (Labyrinth). A mouse is in the labyrinth depicted in figure 8.1 with 9 squares. We consider the three classes of squares: $A = \{1, 3, 7, 9\}$ (the corners), $B = \{5\}$ (the center) and $C = \{2, 4, 6, 8\}$ the other squares. At each step $n \in \mathbb{N}$, the mouse is in a square and we denote by X_n its number and Y_n its class.

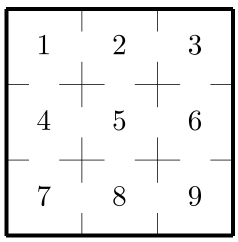

Figure 8.1: Labyrinth

.

1. At each step, the mouse choose an adjacent square at random (an uniformly). Prove that $X = (X_n, n \in \mathbb{N})$ is a Markov chain and represent its transition graph. Classify the states of X.

2. Prove that $Y = (Y_n, n \in \mathbb{N})$ is a Markov chain and represent its transition graph. Compute the invariant probability measure of Y and deduce the one of X.

$\triangle$

Exercise 8.25 (Skeleton Markov chains). Let $X = (X_n, n \in \mathbb{N})$ be a Markov chain on a countable space E with transition matrix P. We use the convention $\inf \emptyset = +\infty$. We define $\tau_1 = \inf\{k \geq 1; X_k \neq X_0\}$.

1. Let $x \in E$. Give the distribution of τ_1 conditionally on $\{X_0 = x\}$. Check that, conditionally on $\{X_0 = x\}$, $\tau_1 = +\infty$ a.s. if x is an absorbing state and otherwise a.s. τ_1 is finite.
2. Conditionally on $\{X_0 = x\}$, if x is not an absorbing state, give the distribution of X_{τ_1}.

We set $S_0 = 0$, $Y_0 = X_0$ and by recurrence for $n \geq 1$: $S_n = S_{n-1} + \tau_n$, and if $S_n < +\infty$: $Y_n = X_{S_n}$ as well as $\tau_{n+1} = \inf\{k \geq 1; X_{k+S_n} \neq X_{S_n}\}$. Let $R = \inf\{n \in \mathbb{N}^*; \tau_n = +\infty\} = \inf\{n \in \mathbb{N}^*; S_n = +\infty\}$.

3. Prove that if X does not have absorbing states, then a.s. $R = +\infty$.

We assume that X does not have absorbing states.

4. Prove that $Y = (Y_n, n \in \mathbb{N})$ is a Markov chain (it is called skeleton of X). Prove that its transition matrix, Q, is given by:
$$Q(x, y) = \frac{P(x, y)}{1 - P(x, x)} \mathbf{1}_{\{x \neq y\}} \quad \text{for } x, y \in E.$$
5. Let π be an invariant probability distribution of X. We define a measure ν on E by:
$$\nu(x) = \frac{\pi(x)(1 - P(x, x))}{\sum_{y \in E} \pi(y)(1 - P(y, y))}, \quad x \in E.$$
Check that ν is an invariant probability measure of Y.

△

Exercise 8.26 (Parameter estimation). Let $X = (X_n, n \in \mathbb{N})$ be an irreducible positive recurrent Markov chain on a countable state space E with transition matrix P and invariant probability π. The aim of this exercise is to give an estimation of the parameter π and P of the Markov chain X.

1. For $x \in E$ and $n \in \mathbb{N}^*$, we set $\hat{\pi}(x; n) = \frac{1}{n} \operatorname{Card} \{1 \leq k \leq n;\ X_k = x\}$. Prove that a.s. for all $x \in E$, $\lim_{n \to +\infty} \hat{\pi}(x; n) = \pi(x)$.

We set $Z = (Z_n, n \in \mathbb{N}^*)$ with $Z_n = (X_{n-1}, X_n)$.

2. Prove that Z is an irreducible Markov chain on $E_2 = \big\{(x, y) \in E^2;\ P(x, y) > 0\big\}$. And compute its transition matrix.
3. Compute the invariant probability distribution of Z and deduce that Z is recurrent positive.
4. For $x, y \in E$ and $n \in \mathbb{N}^*$, we set:
$$\hat{P}(x, y; n) = \frac{\operatorname{Card} \{1 \leq k \leq n;\ Z_k = (x, y)\}}{\operatorname{Card} \{0 \leq k \leq n-1;\ X_k = x\}},$$
with the convention that $\hat{P}(x, y; n) = 0$ if $\operatorname{Card} \{0 \leq k \leq n-1;\ X_k = x\} = 0$. Prove that a.s. for all $(x, y) \in E_2$, $\lim_{n \to +\infty} \hat{P}(x, y; n) = P(x, y)$.

△

8.4 Martingales

Exercise 8.27 (Exit time distribution). Let U be a random variable on $\{-1,1\}$ such that $\mathbb{P}(U=1) = 1-\mathbb{P}(U=-1) = p$ with $p \in (0,1)$. We consider the simple random walk $X = (X_n, n \in \mathbb{N})$ from Exercise 3.4 started at $X_0 = 0$ defined by $X_n = \sum_{k=1}^n U_k$, where $(U_n, n \in \mathbb{N}^*)$ are independent random variables distributed as U. Let $a \in \mathbb{N}^*$ and consider $\tau_a = \inf\{n \in \mathbb{N}^*;\, |X_n| \geq a\}$ the exit time of $(-a,a)$. We set $\varphi(\lambda) = \log\left(\mathbb{E}[\mathrm{e}^{\lambda U}]\right)$ for $\lambda \in \mathbb{R}$. Let $\lambda \in \mathbb{R}$ such that $\varphi(\lambda) \geq 0$.

1. Prove that τ_a is a stopping time. Using that X is an irreducible Markov chain, prove that a.s. τ_a is finite (but not bounded if $a \geq 2$).
2. Prove that $M^{(\lambda)} = (M_n^{(\lambda)} = \mathrm{e}^{\lambda X_n - n\varphi(\lambda)};\, n \in \mathbb{N})$ is a positive martingale.
3. Using the optional stopping theorem and that $\varphi(\lambda) \geq 0$, prove that $\mathbb{E}[M_{\tau_a}^{(\lambda)}] = 1$.
4. Assume that $p = 1/2$. Check that φ is non-negative. By considering $M_{\tau_a}^{(\pm\lambda)}$ for $\lambda \in \mathbb{R}$, prove that for all $r \geq 0$:

$$\mathbb{E}\left[\mathrm{e}^{-r\tau_a}\right] = \frac{1}{\cosh(a \cosh^{-1}(\mathrm{e}^r))} \cdot$$

△

Exercise 8.28 (Return time to 0). Let U be a random variable on $\{-1,1\}$ such that $\mathbb{P}(U = 1) = 1-\mathbb{P}(U=-1) = 1/2$. We consider the simple random walk $X = (X_n, n \in \mathbb{N})$ started at $X_0 = 1$ defined by $X_n = 1 + \sum_{k=1}^n U_k$, where $(U_n, n \in \mathbb{N}^*)$ are independent random variables distributed as U. Let $\tau = \inf\{n \in \mathbb{N}^*;\, X_n = 0\}$ be the return time to 0.

1. Check that the $\mathbb{Z}$-valued Markov chain X is irreducible.
2. Prove that $M = (M_n = X_{n\wedge\tau};\, n \in \mathbb{N})$ is a non-negative martingale.
3. Deduce that τ is a.s. finite and thus X is recurrent.
4. Check that $\mathbb{E}[M_\tau] \neq \mathbb{E}[M_0]$ (thus τ is not bounded and M is not uniformly integrable).
5. Prove that $N = (N_n = X_n^2 - n)$ is a martingale.
6. Deduce that $\mathbb{E}[\tau] = +\infty$ and prove that X is recurrent null.

△

Exercise 8.29 (Martingale not converging in L^1). Let $(X_n, n \in \mathbb{N}^*)$ be a sequence of independent Bernoulli random variables of parameter $\mathbb{E}[X_n] = (1+\mathrm{e})^{-1}$. We define $M_0 = 1$ and for $n \in \mathbb{N}^*$:

$$M_n = \mathrm{e}^{-n+2\sum_{i=1}^n X_i} .$$

1. Prove that $M = (M_n, n \in \mathbb{N})$ is a martingale and that a.s. $\lim_{n\to\infty} M_n = 0$.
2. Check that M doesn't converge in L^1.

△

Exercise 8.30 (Martingale not converging a.s.). Let $(Z_n, n \in \mathbb{N}^*)$ be independent random variables such that $\mathbb{P}(Z_n = 1) = \mathbb{P}(Z_n = -1) = 1/(2n)$ and $\mathbb{P}(Z_n = 0) = 1 - n^{-1}$. We set $X_1 = Z_1$ and for $n \geq 1$:

$$X_{n+1} = Z_{n+1}\mathbf{1}_{\{X_n=0\}} + (n+1)X_n|Z_{n+1}|\,\mathbf{1}_{\{X_n\neq 0\}}.$$

1. Check that $|X_n| \leq n!$ and that $X = (X_n, n \in \mathbb{N}^*)$ is a martingale.
2. Prove directly that X converge in probability towards 0.
3. Using Borel-Cantelli's lemma, prove that $\mathbb{P}(Z_n \neq 0 \text{ infinitely often}) = 1$. Deduce that $\mathbb{P}(\lim_{n\to\infty} X_n \text{ exists}) = 0$. In particular, the martingale does not converge a.s. towards 0.

△

Exercise 8.31 (Wright-Fisher model). We consider a population of constant size N. We assume that the reproduction is random: this corresponds in the end to each individual choosing his parent independently in the previous generation. The Wright-Fisher model study the evolution of the number of individuals carrying one of the two alleles A and a. For $n \in \mathbb{N}$, let X_n denote the number of alleles A at generation n in the population. We assume that $X_0 = i \in \{0, \ldots, N\}$ is given. We shall study the process $X = (X_n, n \geq 0)$.

1. Give the law of X_{n+1} conditionally on X_n.
2. Prove that X is a martingale (specify the filtration).
3. Prove that X converges to a limit, say X_∞, and give the type of convergence.
4. Prove that $M = (M_n = (\frac{N}{N-1})^n X_n(N - X_n), n \geq 0)$ is a martingale and compute $\mathbb{E}[X_\infty(N - X_\infty)]$.
5. Prove that one of the allele disappears a.s. in finite time. Compute the probability that allele A disappears.
6. Compute $\lim_{n\to\infty} M_n$ and deduce that M doesn't converge in L^1.

△

Exercise 8.32 (Waiting time of a given sequence). Let $X = (X_n, n \in \mathbb{N}^*)$ be a sequence of independent Bernoulli random variable with parameter $p \in (0,1)$: $\mathbb{P}(X_n = 1) = 1 - \mathbb{P}(X_n = 0) = p$. Let $\tau_{ijk} = \inf\{n \geq 3; (X_{n-2}, X_{n-1}, X_n) = (i,j,k)\}$ be the waiting time of the sequence $(i,j,k) \in \{0,1\}^3$. The aim of this exercise is to compute its expectation.

1. Prove that τ_{ijk} is a stopping time a.s. finite.
2. We set $S_0 = 0$ and $S_n = (S_{n-1} + 1)\dfrac{X_n}{p}$ for $n \geq 1$. Prove that $(S_n - n, n \geq 0)$ is a martingale. Deduce $\mathbb{E}[\tau_{111}]$.
3. Compute $\mathbb{P}(\tau_{110} < \tau_{111})$.

4. Compute $\mathbb{E}[\tau_{110}]$, using the sequence $(T_n, n \geq 2)$ defined by $T_2 = \dfrac{X_1X_2}{p^2} + \dfrac{X_2}{p}$ and $T_n = T_{n-1}\dfrac{1-X_n}{1-p} + \dfrac{X_{n-1}X_n}{p^2} - \dfrac{X_{n-1}(1-X_n)}{p(1-p)} + \dfrac{X_n}{p}$ for $n \geq 3$.

5. Using similar arguments, compute $\mathbb{E}[\tau_{100}]$ and $\mathbb{E}[\tau_{101}]$.

If $p = 1/2$, it can be proved[1] that for any sequence $(i, j, k) \in \{0, 1\}$, the sequence $(\bar{\jmath}, i, j)$, with $\bar{\jmath} = 1 - j$, appears earlier in probability, that is $\mathbb{P}(\tau_{\bar{\jmath}ij} < \tau_{ijk}) > 1/2$. △

Exercise 8.33 (When does an insurance companies goes bankrupt?). We consider the evolution of the capital of an insurance company. Let $S_0 = x > 0$ be the initial capital, $c > 0$ the fixed income per year and $X_n \geq 0$ the (random) cost of the damage for the year n. The capital at the end of year $n \geq 1$ is thus $S_n = x + nc - \sum_{k=1}^n X_k$. Bankruptcy happens if the capital becomes negative that is if the bankruptcy time $\tau = \inf\{k \in \mathbb{N}; S_k < 0\}$, with the convention $\inf \emptyset = \infty$, is finite. The goal of this exercise is to find an upper bound of the bankruptcy probability $\mathbb{P}(\tau < \infty)$.

We assume the real random variables $(X_k, k \geq 1)$ are independent, identically distributed, a.s. non constant, and have all its exponential moments (i.e. $\mathbb{E}[\mathrm{e}^{\lambda X_1}] < \infty$ for all $\lambda \in \mathbb{R}$).

1. Check that $\mathbb{E}[X_1] > c$ implies $\mathbb{P}(\tau < \infty) = 1$, and that $\mathbb{P}(X_1 > c) = 0$ implies $\mathbb{P}(\tau < \infty) = 0$.

We assume that $\mathbb{E}[X_1] < c$ and $\mathbb{P}(X_1 > c) > 0$.

2. Check that if $\mathbb{E}[\mathrm{e}^{\lambda X_k}] \geq \mathrm{e}^{\lambda c}$, then $V = (V_n = \mathrm{e}^{-\lambda S_n + \lambda x}, n \geq 0)$ is a non-negative sub-martingale.

3. Let $N \geq 1$. Prove that $\{\tau \leq N\}$ is the disjoint union of the events $F_k = \{S_r \geq 0 \text{ for } r < k, S_k < 0\} = \{\tau = k\}$ for $k \in \{1, \ldots, N\}$. Deduce that:

$$\mathbb{E}[V_N \mathbf{1}_{\{\tau \leq N\}}] \geq \sum_{k=1}^{N} \mathbb{E}[V_k \mathbf{1}_{\{\tau = k\}}] \geq \mathrm{e}^{\lambda x}\, \mathbb{P}(\tau \leq N).$$

(You can check we recover the maximal inequality for the positive sub-martingale.)

4. Deduce that $\mathbb{P}(\tau < \infty) \leq \mathrm{e}^{-\lambda_0 x}$, where $\lambda_0 \in (0, \infty)$ is the unique root of $\mathbb{E}[\mathrm{e}^{\lambda X_1}] = \mathrm{e}^{\lambda c}$.

△

Exercise 8.34 (A.s. convergence and convergence in distribution). Let $(X_n, n \geq 1)$ be a sequence of independent real random variables. We set $S_n = \sum_{k=1}^n X_k$ for $n \geq 1$. The goal of this exercise is to prove that if the sequence $(S_n, n \geq 1)$ converges in distribution, then it converges a.s. also.

For $t \in \mathbb{R}$, we set $\psi_n(t) = \mathbb{E}[\mathrm{e}^{itX_n}]$ and $M_n(t) = \dfrac{\mathrm{e}^{itS_n}}{\prod_{k=1}^n \psi_k(t)}$ for $n \geq 1$ if $\prod_{k=1}^n \psi_k(t) \neq 0$.

1. Let $t \in \mathbb{R}$ be such that $\prod_{k=1}^n \psi_k(t) \neq 0$. Prove that $(M_k(t), 1 \leq k \leq n)$ is a martingale.

[1] R. Graham, D. Knuth and O. Patashnik. *Concrete mathematics: a foundation of computer science*, 2nd Edition. Addison-Wesley Publishing Company, 1994. (See Section 8.4.)

We assume that $(S_n, n \geq 1)$ converges in distribution towards S.

2. Prove there exists $\varepsilon > 0$ such that for all $t \in [-\varepsilon, \varepsilon]$, a.s. the sequence $(\mathrm{e}^{itS_n}, n \geq 1)$ converges.

3. We recall that if there exists $\varepsilon > 0$ s.t., for almost all $t \in [-\varepsilon, \varepsilon]$, the sequence $(\mathrm{e}^{its_n}, n \geq 1)$ converges, then the sequence $(s_n, n \geq 1)$ converges. Prove that $(S_n, n \geq 1)$ converges a.s. towards a random variable distributed as S.

△

Exercise 8.35 (Non-negative martingale and stopping times). Let $\mathbb{F} = (\mathcal{F}_n, n \in \bar{\mathbb{N}})$ be a filtration with $\mathcal{F}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$, $M = (M_n, n \in \mathbb{N})$ a non-negative martingale, and $\tau' \geq \tau$ two stopping times. We shal prove that:

$$\mathbb{E}\left[M_\tau\right] \geq \mathbb{E}\left[M_{\tau'}\right]. \tag{8.2}$$

1. After explaining why M_∞ is well defined, check that $\mathbb{E}[M_0] \geq \mathbb{E}[M_\infty]$.

2. Deduce that $\mathbb{E}[M_0] \geq \mathbb{E}[M_\tau]$.

We consider the process $N = (N_n = M_{\tau+n}, n \in \mathbb{N})$ and the filtration $\mathbb{G} = (\mathcal{G}_n = \mathcal{F}_{\tau+n}, n \in \bar{\mathbb{N}})$.

3. Prove that N is a martingale with respect to the filtration $\mathbb{G}$.

4. Prove that $\tau' - \tau$, with the convention $\infty - \infty = 0$, is a stopping time with respect to the filtration $\mathbb{G}$.

5. Prove (8.2).

△

8.5 Optimal stopping

Exercise 8.36 (Dice). In a game, you are allowed to roll a dice at most three times. When you roll the dice, you can either stop and gain the result of the dice or, unless it is your third roll, go on. What is the strategy which maximizes your gain and the corresponding average gain. △

8.6 Brownian motion

Exercise 8.37 (Transformations of Brownian motion). Let $B = (B_t, t \in \mathbb{R}_+)$ be a standard Brownian motion.

1. Let $t_0 \in \mathbb{R}_+$. Prove that $(B_{t+t_0} - B_{t_0},\, t \in \mathbb{R}_+)$ is a Brownian motion.

2. Let $\lambda > 0$. Prove that $(\lambda^{-1/2} B_{\lambda t},\, t \in \mathbb{R}_+)$ is a Brownian motion.

3. Prove that $(tB_{1/t},\, t \in (0, +\infty))$ is distributed as $(B_t,\, t \in (0, +\infty))$. Deduce that a.s. $\lim_{t\to+\infty} B_t/t = 0$.

△

Exercise 8.38 (Simulation of Brownian motion). We present a recursive algorithm due to Lévy to simulate the Brownian motion on the interval $[0, T]$ with $T > 0$.

1. Prove that the Brownian bridge W^T is a centered Gaussian process with covariance kernel $K = (K(s,t);\ s,t \in [0,T])$ given by $K(s,t) = t \wedge s(T - t \vee s)/T$.

2. Prove that $\mathbb{E}[W_t^T B_{T+s}] = 0$ for all $t \in [0,T]$ and $s \in \mathbb{R}_+$. Deduce that W^T is independent of $(B_{T+s}, s \in \mathbb{R}_+)$.

Let $s \geq r \geq 0$ be fixed. We define the process $\tilde{W} = (\tilde{W}_t, t \in [r,s])$ by:

$$\tilde{W}_t = B_t - B_r - \frac{t-r}{s-r}(B_s - B_r).$$

3. Prove that $\tilde{W}$ is a Gaussian process distributed as W^{s-r}. And deduce the variance of $\tilde{W}_t$ for $t \in [r,s]$.

4. Using that $(\tilde{W}, B)$ is a Gaussian process, prove that $\tilde{W}$ is independent of $(B_u, u \in [0,r] \bigcup [s, +\infty))$.

5. Let $t \in [r,s]$. Deduce that conditionally on $(B_u, u \in [0,r] \bigcup [s,+\infty))$, B_t is distributed as:
$$\sqrt{\frac{(t-r)(s-t)}{s-r}}G + \frac{s-t}{s-r}B_r + \frac{t-r}{s-r}B_s,$$
where $G \sim \mathcal{N}(0,1)$ is independent of $(B_u, u \in [0,r] \bigcup [s,+\infty))$.

6. Let $n \geq 1$. Deduce a recursive algorithm to simulate a standard Brownian motion at times $0 = t_0 \leq t_1 \leq \cdots \leq t_{2^n} = T$, by first simulating the standard Brownian motion at time 0 and T.

△

Exercise 8.39 (Ornstein-Uhlenbeck process). Let $V = (V_t, t \in \mathbb{R}_+)$ be the solution of the Langevin equation (6.10) with initial condition V_0. Let U be a centered Gaussian random variable with variance $\sigma^2/(2a)$ and independent of the Brownian motion B.

1. Prove that $(V_t, t \in \mathbb{R}_+)$ converges in distribution towards U as t goes to infinity.

2. Assume that V_0 is deterministic. Prove that V is a Gaussian process and a Markov process. Prove that V_t is distributed as $\mathcal{N}\left(V_0 \,\mathrm{e}^{-at}, \sigma^2(2a)^{-1}(1 - \mathrm{e}^{-2at})\right)$.

3. Assume that $V_0 = U$. Prove that V_t is distributed as U for all $t \in \mathbb{R}_+$.

4. Assume that $V_0 = U$. Prove that V is distributed as $(Z_t, t \in \mathbb{R}_+)$ with:
$$Z_t = \frac{\sigma}{\sqrt{2a}}\,\mathrm{e}^{-at}\,B_{\mathrm{e}^{2at}}. \tag{8.3}$$

△

Chapter 9

Solutions

9.1 Measure theory and random variables

Exercise 8.1 The family $\mathcal{C} = \{]-\infty, a], a \in \mathbb{R}\}$ is stable by finite intersection (this is a π-system). In particular, all the intervals belong to the σ-field $\sigma(\mathcal{C})$. Thus it is the Borel σ-field on $\mathbb{R}$.

Exercise 8.2 Let μ and μ' be two σ-finite measures on $(\Omega, \mathcal{F})$ which coincide on a collection of events $\mathcal{C}$ stable by finite intersection such that $\Omega_n \in \mathcal{C}$ for all $n \in \mathbb{N}$, where $\mu(\Omega_n) < +\infty$ and $\bigcup_{n\in\mathbb{N}} \Omega_n = \Omega$. By replacing Ω_n by $\cup_{0\leq k\leq n}\Omega_k$ for $n \in \mathbb{N}$, we can assume that the sequence $(\Omega_n, n \in \mathbb{N})$ is non-decreasing. For $n \in \mathbb{N}$, we can define $\mathbb{P}_n = \mu(\Omega_n)^{-1}\mu$ and $\mathbb{P}'_n = \mu'(\Omega_n)^{-1}\mu'$. Those two probability measures coincide on $\mathcal{C}_n = \{A \cap \Omega_n,\, A \in \mathcal{C}\} \subset \mathcal{C}$ which is also stable by finite intersection, and thus thanks to Corollary 1.14, they coincide on $\sigma(\mathcal{C}_n)$. As $\mu(\Omega_n) = \mu'(\Omega_n)$, we deduce that μ and μ' coincide on $\sigma(\mathcal{C}_n)$ for all $n \in \mathbb{N}$.

Let $\mathcal{G} = \{A \in \mathcal{F},\, A \cap \Omega_n \in \sigma(\mathcal{C}_n) \text{ for all } n \in \mathbb{N}\}$. It is elementary to check that $\mathcal{G}$ is a σ-field. Since $\mathcal{C}$ is stable by finite intersection, we have $\mathcal{C} \subset \mathcal{G}$ and thus $\sigma(\mathcal{C}) \subset \mathcal{G}$. If $A \in \mathcal{G}$, we get $A = \bigcup_{n\in\mathbb{N}}(A \cap \Omega_n)$. As $A \cap \Omega_n \in \sigma(\mathcal{C}_n) \subset \sigma(\mathcal{C})$, we deduce that $A \in \sigma(\mathcal{C})$. This implies that $\mathcal{G} = \sigma(\mathcal{C})$. By monotone convergence, we get that for $A \in \sigma(\mathcal{C})$, that is $A \in \mathcal{G}$:

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap \Omega_n) = \lim_{n\to\infty} \mu'(A \cap \Omega_n) = \mu'(A),$$

where we used that μ and μ' coincide on $\sigma(\mathcal{C}_n)$ for the second equality. We deduce that $\mu = \mu'$ on $\sigma(\mathcal{C})$.

The extension of Corollary 1.15 is immediate.

Exercise 8.3 We have:

$$\limsup_{n\to\infty}(a_n - b_n) = \lim_{n\to\infty} \sup_{0\leq k\leq n} (a_k - b_k) \leq \lim_{n\to\infty} (\sup_{0\leq k\leq n} a_k - \inf_{0\leq k\leq n} b_k) = \limsup_{n\to\infty} a_n - \liminf_{n\to\infty} b_n.$$

We also have:

$$\limsup_{n\to\infty}(a_n - b_n) = \lim_{n\to\infty} \sup_{0\leq k\leq n} (a_k - b_k) \geq \lim_{n\to\infty} (\sup_{0\leq k\leq n} a_k - \sup_{0\leq k\leq n} b_k) = \limsup_{n\to\infty} a_n - \limsup_{n\to\infty} b_n.$$

If furthermore the sequence $(b_n, n \in \mathbb{N})$ converges, we have $\limsup_{n\to\infty} b_n = \liminf_{n\to\infty} b_n$, which allows to conclude.

Exercise 8.4 By successive integration, we get:

$$I_1 = \int_{(0,1)} \left(\int_{(0,1)} f(x,y)dy \right) dx = \int_{(0,1)} \left[\frac{y}{(x^2+y^2)^2} \right]_0^1 dx = \int_{(0,1)} \frac{1}{x^2+1}\, dx = \frac{\pi}{4} \cdot$$

Similarly, we get $I_2 = \int_{(0,1)} \left(\int_{(0,1)} f(x,y)dx \right) dy = -\frac{\pi}{4} \cdot$; If f were integrable on $(0,1)^2$, we could use Fubini's theorem and get that $I_1 = I_2$. Since this equality doesn't hold, we deduce that f is not integrable over $(0,1)^2$.

Exercise 8.5 For the case $f_j \geq 0$ for $j \in J$, only the last sentence of the proof of Proposition 1.62 need to be changed. Use monotone convergence theorem, to get (1.13) holds if the function f_j are non-negative.

For the other case, according to the above argument, we get that:

$$\mathbb{E}\left[\prod_{j \in J} f_j^{\varepsilon_j}(X_j) \right] = \prod_{j \in J} \mathbb{E}[f_j^{\varepsilon_j}(X_j)],$$

where $\varepsilon_j \in \{-,+\}$ for $j \in J$. Those quantities being all finite as $f_j(X_j)$ is integrable, we obtain (1.13) using the linearity of the expectation in $L^1(\mathbb{P})$.

Exercise 8.6 The fact that XY is integrable and that $\mathrm{Cov}(X,Y) = 0$ is a consequence of Exercise 8.5. Let X be a non-negative square-integrable random variable with non-zero variance. Let $Y = \varepsilon X$, with ε independent of X and such that $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$. We have $\mathbb{E}[Y] = 0$ and $\mathbb{E}[XY] = 0$ so that $\mathrm{Cov}(X,Y) = 0$. However, we have $\mathrm{Cov}(X, |Y|) = \mathrm{Var}(X) > 0$. This implies that X and Y are not independent.

Exercise 8.7 Use that $f(\mathbf{1}_A) = 1 + (f(1) - f(0))\mathbf{1}_A$ and Proposition 1.62 to deduce that if the events $(A_i, i \in I)$ are independent then the random variables $(\mathbf{1}_{A_i}, i \in I)$ are independent.

By Definition 1.31, if the random variables $(X_i, i \in I)$ are independent so are the random variables $(f_i(X_i), i \in I)$ for any measurable functions $(f_i, i \in I)$. Take $X_i = \mathbf{1}_{A_i}$ and $f_i(x) = 1 - x$ to deduce that $(\mathbf{1}_{A_i^c}, i \in I)$ are independent random variables, and thus thanks to (1.13), deduce that $(A_i^c, i \in I)$ are also independents events.

Exercise 8.8

1. Let X_1 and X_2 be two independent random variables distributed as $2X - 1$ where X has Bernoulli distribution with parameter $1/2$. Set $\mathcal{G} = \sigma(X_2)$ and $\mathcal{C} = \{\{X_1 = 1\}, \{X_1 X_2 = 1\}\}$. It is elementary to check that if $A \in \mathcal{C}$ then the set A is independent of $\mathcal{G}$. Notice that $\sigma(\mathcal{C}) = \sigma(X_1, X_2)$ which is not independent of $\mathcal{G} = \sigma(X_2)$.

2. Let $\mathcal{B} \subset \mathcal{F}$ be a collection of events. Consider the collection $\mathcal{A} \subset \mathcal{F}$ of events A which are independents of $\mathcal{B}$, that is $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $B \in \mathcal{B}$. It is easy to check that this collection is a monotone class. It contains $\mathcal{C}$ which is stable by finite intersection. By the monotone class theorem, we get that $\sigma(\mathcal{C}) \subset \mathcal{A}$, and thus $\sigma(\mathcal{C})$ is independent of $\mathcal{B}$. Taking $\mathcal{B} = \mathcal{G}$ gives the results.

3. Take $\mathcal{B} = \mathcal{C}'$ in the proof of the previous question to deduce that $\sigma(\mathcal{C})$ and $\mathcal{C}'$ are independent, and then use the previous question to deduce that $\sigma(\mathcal{C})$ and $\sigma(\mathcal{C}')$ are independent.

9.2 Conditional expectation

Exercise 8.9 We have $\sigma(\mathbf{1}_B) = \{\emptyset, \Omega, B, B^c\}$. We get using the characterization (2.1) of the conditional expectation that:

$$\mathbb{E}\left[\mathbf{1}_A | \mathbf{1}_B\right] = \mathbb{P}(A|B)\mathbf{1}_B + \mathbb{P}(A|B^c)\mathbf{1}_{B^c}.$$

Exercise 8.10 Since (X_1, X_2) has the same distribution as (X_2, X_1), we deduce that (X_1, S_2) has the same distribution as (X_2, S_2). This implies that $\mathbb{E}[X_1 | S_2] = \mathbb{E}[X_2 | S_2]$. By linearity, we have:

$$\mathbb{E}[X_1 | S_2] + \mathbb{E}[X_2 | S_2] = \mathbb{E}[S_2 | S_2] = S_2.$$

We deduce that $\mathbb{E}[X_1 | S_2] = S_2/2$. Similarly, we get $\mathbb{E}[X_1 | S_n] = S_n/n$.

Exercise 8.11 Notice that (X, X^2) and $(-X, X^2)$ have the same distribution. We deduce that $\mathbb{E}[X | X^2] = \mathbb{E}[-X | X^2] = -\mathbb{E}[X | X^2]$. This implies that $\mathbb{E}[X | X^2] = 0$.

Exercise 8.12 Let h be measurable bounded. We want to write

$$\mathbb{E}[X\, h(|X|)] = \mathbb{E}[g(|X|)\ h(|X|)] \tag{9.1}$$

for some measurable function g, such that thanks to (2.2) (with h given by $\mathbf{1}_A = h(|X|)$ for $A \in \sigma(|X|)$), we will deduce that a.s. $g(|X|) = \mathbb{E}[X | \, |X|]$. On one hand, we have:

$$\begin{aligned}\mathbb{E}[X\, h(|X|)] = \int_{\mathbb{R}} xh(|x|)f(x)\,\mathrm{d}x &= \int_{\mathbb{R}_+} xh(|x|)f(x)\,\mathrm{d}x + \int_{\mathbb{R}_-} xh(|x|)f(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}_+} xh(x)\big(f(x) - f(-x)\big)\,\mathrm{d}x,\end{aligned}$$

and on the other hand we have:

$$\mathbb{E}[g(|X|)\ h(|X|)] = \int_{\mathbb{R}} g(|x|)h(|x|)f(x)\,\mathrm{d}x = \int_{\mathbb{R}_+} g(x)h(x)\big(f(x) + f(-x)\big)\,\mathrm{d}x.$$

We deduce that for $x \in \mathbb{R}_+$:

$$g(x) = x\,\frac{f(x) - f(-x)}{f(x) + f(-x)}$$

satisfies (9.1) and thus a.s.

$$\mathbb{E}[X | \, |X|] = |X|\,\frac{f(|X|) - f(-|X|)}{f(X) + f(-X)}.$$

Notice that $\sigma(|X|) = \sigma(X^2)$ (as $|x| = \sqrt{x^2}$ and $x^2 = |x|^2$) so that $\mathbb{E}[X | X^2] = \mathbb{E}[X | \, |X|]$.

Exercise 8.13 By Jensen's inequality, we have $\mathbb{E}[X | \mathcal{H}]^2 \leq \mathbb{E}[X^2 | \mathcal{H}]$. Since $\mathbb{E}[\mathbb{E}[X^2 | \mathcal{H}]] = \mathbb{E}[X^2] < +\infty$, we deduce that $\mathbb{E}[X | \mathcal{H}]^2$ is integrable. Using Jensen's inequality, we get that:

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

Exercise 8.14 Set $\varphi(x) = \mathbb{E}[|x - Y|]$ for $x \in \mathbb{R}$. Using Jensen's inequality, we get that $\varphi(x) = \mathbb{E}[|x - Y|] \geq |x - \mathbb{E}[Y]| = |x|$ for all $x \in \mathbb{R}$. As X and Y are independent, we also have that $\mathbb{E}[|X - Y|] = \mathbb{E}[\varphi(X)]$. This gives the result: $\mathbb{E}[|X - Y|] \geq \mathbb{E}[|X|]$.

Exercise 8.15

Exercise 8.16 Let φ be a positive strictly convex function on $\mathbb{R}$ such that $\lim_{x\to+\infty} \varphi(x)/x$ and $\lim_{x\to-\infty} \varphi(x)/x$ are finite. This implies in particular that $\varphi(X)$ and $\varphi(Y)$ are integrable. We deduce from Jensen's inequality that $\mathbb{E}[\varphi(X)|\, Y] \geq \varphi(\mathbb{E}[X|\, Y]) = \varphi(Y)$ and thus $\mathbb{E}[\varphi(X)] \geq \mathbb{E}[\varphi(Y)]$. By symmetry, we get $\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)]$ and thus the Jensen's inequality is an equality: a.s. $\mathbb{E}[\varphi(X)|\, Y] = \varphi(\mathbb{E}[X|\, Y])$. Since φ is strictly convex, this implies that a.s. $\varphi(X) = \varphi(\mathbb{E}[X|\, Y]) = \varphi(Y)$, and thus, as φ is general, a.s. $X = Y$.

Exercise 8.17 Let $A \in \mathcal{H}$ and consider the random variable $X = (V, \mathbf{1}_A)$ which takes values in $E \times \{0, 1\}$. Since Y and X are independent, we deduce from Lemma 1.56 that for measurable sets $B \in \mathcal{S}$ and $C \in \mathcal{E}$:

$$\mathrm{P}_{(Y,X)}(B \times C) = \mathbb{P}(Y \in B,\, X \in C) = \mathbb{P}(Y \in B)\mathbb{P}(X \in C) = \mathrm{P}_Y(B)\mathrm{P}_X(C).$$

We deduce from Fubini's theorem that $\mathrm{P}_{(Y,X)}(\mathrm{d}y, \mathrm{d}x) = \mathrm{P}_Y(\mathrm{d}y)\mathrm{P}_X(\mathrm{d}x)$, and, with $x = (x_1, x_2) \in E \times \{0, 1\}$ and $f(x, y) = \varphi(y, x_1)x_2$, from Equation (1.6) that:

$$\mathbb{E}[\varphi(Y, V)\mathbf{1}_A] = \int f(y, x)\; \mathrm{P}_{(Y,X)}(\mathrm{d}y, \mathrm{d}x) = \int \left(\int f(x, y)\; \mathrm{P}_Y(\mathrm{d}y) \right)\; \mathrm{P}_X(\mathrm{d}x).$$

Set $g(v) = \mathbb{E}[\varphi(v, Y)] = \int \varphi(v, y)\; \mathrm{P}_Y(\mathrm{d}y)$ so that:

$$\mathbb{E}[\varphi(Y, V)\mathbf{1}_A] = \int g(x_1)x_2 \mathrm{P}_X(\mathrm{d}x) = \mathbb{E}[g(V)\mathbf{1}_A].$$

This directly implies that a.s. $\mathbb{E}[\varphi(Y, V)|\, \mathcal{H}] = g(V)$.

Exercise 8.18 1. We have:

$$\begin{aligned} \mathbb{E}\left[\mathbf{1}_A\, \mathbb{E}\left[\mathbf{1}_B|\, \mathcal{A}\right] |\, \mathcal{H}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_A \mathbf{1}_B|\, \mathcal{A}\right] |\, \mathcal{H}\right] \\ &= \mathbb{P}(A \cap B|\, \mathcal{H}) \\ &= \mathbb{P}(A|\, \mathcal{H})\mathbb{P}(B|\, \mathcal{H}) \\ &= \mathbb{E}\left[\mathbf{1}_A\, \mathbb{E}\left[\mathbf{1}_B|\, \mathcal{H}\right] |\, \mathcal{H}\right], \end{aligned}$$

where we used that $A \in \mathcal{A}$ for the first equality, that A and B are independent conditionally on $\mathcal{H}$ for the third, and that $\mathbb{E}\left[\mathbf{1}_B|\, \mathcal{H}\right]$ is $\mathcal{H}$-measurable for the last.

2. We have for any $A \in \mathcal{A}$:

$$\begin{aligned} \mathbb{E}\left[\mathbf{1}_A\, \mathbb{E}\left[\mathbf{1}_B|\, \mathcal{A}\right]\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_A\, \mathbb{E}\left[\mathbf{1}_B|\, \mathcal{A}\right] |\, \mathcal{H}\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_A\, \mathbb{E}\left[\mathbf{1}_B|\, \mathcal{H}\right] |\, \mathcal{H}\right]\right] \\ &= \mathbb{E}\left[\mathbf{1}_A\, \mathbb{E}\left[\mathbf{1}_B|\, \mathcal{H}\right]\right], \end{aligned}$$

where we used the first question for the second equality. Since $\mathcal{H} \subset \mathcal{A}$, we deduce that $\mathbb{E}\left[\mathbf{1}_B|\, \mathcal{H}\right]$ is $\mathcal{A}$-measurable, and by uniqueness of the conditional expectation, we get that a.s. $\mathbb{E}\left[\mathbf{1}_B|\, \mathcal{A}\right] = \mathbb{E}\left[\mathbf{1}_B|\, \mathcal{H}\right]$.

Exercise 8.19 1. We have $\mathbb{E}[|X_n|] = 1/n$, which implies that the sequence $X = (X_n, n \in \mathbb{N}^*)$ converges to 0 in L^1. Notice that $\mathbb{P}(X_n > 1/n) = \mathbb{P}(B_n = B'_n = 1) = n^{-2}$ and thus the non-negative random variable $\sum_{n\in\mathbb{N}^*} \mathbf{1}_{\{X_n>1/n\}}$ is integrable and thus finite. Since the terms in the sum are either 1 or 0, we deduce that $X_n \leq 1/n$ for n large enough, that is X converges to a.s. 0.

2. Set $Y = (Y_n = \mathbb{E}[X_n | \mathcal{H}], n \in \mathbb{N}^*)$. As Y_n is nonnegative, we get $\mathbb{E}[|Y_n|] = \mathbb{E}[Y_n] = \mathbb{E}[X_n]$, and thus Y converges to 0 in L^1.

   We also have $Y_n = B'_n \in \{0, 1\}$. The events $(\{Y_n = 1\}, n \in \mathbb{N}^*)$ are independent and $\sum_{n\in\mathbb{N}^*} \mathbb{P}(Y_n = 1) = \sum_{n\in\mathbb{N}^*} n^{-1} = +\infty$. We deduce from the Borel-Cantelli lemma that a.s. $\limsup_{n\to\infty} Y_n = 1$, and thus the sequence Y does not converges a.s. to 0.

Exercise 8.20 The computations are elementary, see Sections 2.3.2 and 2.3.3. The density of the probability distribution of V is $f_V(v) = \int f_{Y,V}(y, v)\, \mathrm{d}y = \lambda\, \mathrm{e}^{-\lambda v}\, \mathbf{1}_{\{v>0\}}$. We deduce that for $y > 0$, $f_{Y|V}(y|v) = v^{-1}\mathbf{1}_{\{0<y<v\}}$, which is the density of the uniform distribution on $[0, v]$. The last formula is then clear.

Exercise 8.21 We first assume that Y and V are independent. We deduce from Exercise 8.17 with $\mathcal{H} = \sigma(V)$ that, for all nonnegative measurable function φ, we have $\mathbb{E}[\varphi(Y, V)| \mathcal{H}] = g(V)$ with $g(v) = \mathbb{E}[\varphi(Y, v)] = \int \varphi(y, v)\, \mathrm{P}_Y(\mathrm{d}y)$. We deduce from Definition 2.17 that $\mathbb{P}(Y \in A|V) = \mathrm{P}_Y(A)$, and thus the conditional distribution of Y given V exists and is given by the kernel $\kappa(v, \mathrm{d}y) = \mathrm{P}_Y(\mathrm{d}y)$ which does not depend on $v \in E$.

We now assume that the conditional distribution of Y given V exists and is given by a kernel which does not depend on $v \in E$, say $\kappa(\mathrm{d}y)$. By Definition 2.17, we have $\mathbb{P}(Y \in A|V) = \kappa(V, A) = \kappa(A)$, and taking the expectation, we get $\mathbb{P}(Y \in A) = \kappa(A)$, that is $\kappa = \mathrm{P}_Y$. We also get $\mathbb{P}(Y \in A, V \in B) = \mathbb{E}[\mathbf{1}_B(V)\mathbb{P}(Y \in A| V)] = \mathrm{P}_Y(A)\mathbb{E}[\mathbf{1}_B(V)] = \mathbb{P}(Y \in A)\mathbb{P}(V \in B)$, which means that Y and V are independent.

9.3 Discrete Markov chains

Exercise 8.22 1. We have $\mathbb{P}(X_2 = y|X_0 = x) = P^2(x, y)$ as:

$$\begin{aligned}
\mathbb{P}(X_2 = y|X_0 = x) &= \sum_{z\in E} \mathbb{P}(X_2 = y, X_1 = z|X_0 = x)\\
&= \sum_{z\in E} \mathbb{P}(X_2 = y|X_1 = z, X_0 = x)\mathbb{P}(X_1 = z|X_0 = x)\\
&= \sum_{z\in E} \mathbb{P}(X_2 = y|X_1 = z)\mathbb{P}(X_1 = z|X_0 = x)\\
&= \sum_{z\in E} P(x, z)P(z, y)\\
&= P^2(x, y),
\end{aligned}$$

where we used the Markov property for the third equality.

Assume that X is a stochastic dynamical system: $X_{n+1} = f(X_n, U_{n+1})$ for some measurable function f and $(U_n,\, n \in \mathbb{N}^*)$ independent identically distributed random variables independent of X_0. Then, we have:

$$Z_{n+1} = X_{2n+2} = f(f(X_{2n}, U_{2n+1}), U_{2n+2}) = g(Z_n, V_{n+1}),$$

with $V_{n+1} = (U_{2n+1}, U_{2n+2})$ and $g(x, (v_1, v_2)) = f(f(x, v_1), v_2)$. Since the random variables $(V_n, n \in \mathbb{N}^*)$ are independent, identically distributed and independent of $Z_0 = X_0$, we deduce that Z is a stochastic dynamical system and thus a Markov chain.

In general, X is distributed as a stochastic dynamical system, say $\tilde{X}$. The process Z is a functional of X, say $Z = F(X)$. We deduce that Z is distributed as $\tilde{Z} = F(\tilde{X})$, which is a stochastic dynamical system, according to the previous argument. Hence, Z is a Markov chain.

Notice that Z has transition matrix $Q = P^2$.

2. On one hand, if π is an invariant probability measure for P, then we have $\pi P^2 = (\pi P)P = \pi P = \pi$. Hence it is also an invariant probability measure for Q.

   On the other hand, for the state space $E = \{a, b\}$ (with $a \neq b$):

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

   and the unique invariant probability measure is $\pi = (1/2, 1/2)^{\mathrm{t}}$. As $Q = P^2$ is the idendity matrix we get that any probability measure is invariant for Q.

Exercise 8.23 1. Assume that X is a stochastic dynamical system: $X_{n+1} = f(X_n, U_{n+1})$ for some measurable function f and $(U_n,\, n \in \mathbb{N}^*)$ independent identically distributed random variables independent of X_0. Then, we have:

$$Y_{n+1} = (X_n, X_{n+1}) = (X_n, f(X_n, U_{n+1})) = g(Y_n, U_{n+1}),$$

with $g((y_1, y_2), u) = (y_2, f(y_2, u))$. Since the random variable $(U_n,\, n \geq 2)$ is independent of $Y_1 = (X_0, X_1)$, we deduce that Y is a stochastic dynamical system and thus a Markov chain.

In general, X is distributed as a stochastic dynamical system, say $\tilde{X}$. The process Y is a functional of X, say $Y = F(X)$. We deduce that Y is distributed as $\tilde{Y} = F(\tilde{X})$, which is a stochastic dynamical system, according to the previous argument. Hence, Y is a Markov chain.

The transition matrix Q of Y on E^2 is given by:

$$Q((x_1, x_2), (y_1, y_2)) = \mathbf{1}_{\{x_2 = y_1\}} P(y_1, y_2).$$

2. Take $E = \{0, 1\}$ and $P(1, 1) = 1$ so that 1 is an absorbing state for X. Then $Q(x, (1, 0)) = 0$ for all $x \in \mathbb{E}^2$. Thus Y is not irreducible.

   Assume that X is irreducible. Notice that Y takes values in $\tilde{E} = \{(x, y) \in \mathbb{E},\, P(x, y) > 0\}$. Let $x = (x_1, x_2), y = (y_1, y_2) \in \tilde{E}$. Since X is irreducible, there exists $n \geq 3$, $x_3, \ldots, x_n = y_1 \in E$ such that $\prod_{i=3}^{n} P(x_{i-1}, x_i) > 0$. This clearly implies that $\prod_{i=3}^{n+1} Q(z_{i-1}, z_i) > 0$, with $z_i = (x_{i-1}, x_i)$, $z_2 = x$ and $z_{n+1} = y$ (notice we used that $Q(z_n, z_{n+1}) = Q(z_n, y) > 0$ as as $y \in \tilde{E}$). Thus Y is irreducible (on $\tilde{E}$).

3. It is easy to check that $\nu = (\nu(z), z \in E^2)$, with $\nu(z) = \pi(x)P(x,y)$ for $z = (x,y) \in E^2$ is an invariant measure for Y. Indeed, we have for $z = (v,w) \in E^2$:

$$\begin{aligned}\nu Q(z) &= \sum_{x,y\in E} \nu((x,y))Q((x,y),(v,w))\\ &= \sum_{x,y\in E} \pi(x)P(x,y)\mathbf{1}_{\{y=v\}}P(v,w) = \pi(v)P(v,w) = \nu(z).\end{aligned}$$

This gives that $\nu Q = \nu$, that is ν is invariant for Q.

Exercise 8.24 1. Since each new step is chosen uniformly random on the available neighbors, the next position depends on the past only through the current position. This implies that X is a Markov chain. Clearly it is irreducible (so all the states belong to the same closed class). Since the state space is finite, the Markov chain is positive recurrent (so all the states are positive recurrent).

2. For all $n \in \mathbb{N}$, set $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$. We shall check that Y is a Markov chain with respect to the filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$. We first compute $\mathbb{P}(Y_{n+1} = A | \mathcal{F}_n)$. Let P be the transition matrix of the Markov chain X. We have:

$$\begin{aligned}\mathbb{P}(Y_{n+1} = A| \mathcal{F}_n) &= \sum_{y\in A} \mathbb{P}(X_{n+1} = y| \mathcal{F}_n)\\ &= \sum_{y\in A} \mathbb{P}(X_{n+1} = y| X_n)\\ &= \sum_{y\in A} P(X_n, y)\\ &= \sum_{y\in A}\sum_{x\in C} P(x,y)\mathbf{1}_{\{X_n=x\}}\\ &= \frac{2}{3}\sum_{x\in C} \mathbf{1}_{\{X_n=x\}}\\ &= \frac{2}{3}\mathbf{1}_{\{Y_n=C\}},\end{aligned}$$

where we used that X is a Markov chain for the second equality, that $P(x,y) = 0$ for $x \notin C$ and $y \in A$ for the fourth, and that $\sum_{y\in A} P(x,y) = 2/3$ for all $x \in C$ for the fifth. Since the last right hand-side term is $\sigma(Y_n)$-measurable, this implies that $\mathbb{P}(Y_{n+1} = A| \mathcal{F}_n) = \mathbb{P}(Y_{n+1} = A| Y_n)$. Similarly, we obtain $\mathbb{P}(Y_{n+1} = B| \mathcal{F}_n) = \frac{1}{3}\mathbf{1}_{\{Y_n=C\}} = \mathbb{P}(Y_{n+1} = B| X_n)$ and $\mathbb{P}(Y_{n+1} = C| \mathcal{F}_n) = \mathbf{1}_{\{Y_n\neq C\}} = \mathbb{P}(Y_{n+1} = C| X_n)$. Since $\mathbb{P}(Y_{n+1} = \bullet | \mathcal{F}_n) = \mathbb{P}(Y_{n+1} = \bullet | Y_n)$, we deduce that Y is a Markov chain. Its transition matrix (on $E = A, B, C$) is given by:

$$Q = \begin{pmatrix} 0 & 0 & 1\\ 0 & 0 & 1\\ 2/3 & 1/3 & 0 \end{pmatrix}.$$

The invariant probability for Q is given by $\pi_Q = (1/3, 1/6, 1/2)$. When Y_n is in a given state $D \in E$, then X_n can be in any state $x \in D$, so intuitively the invariant probability for P could be given by $\pi_P(x) = 1/12$ for $x \in A$, $\pi_P(x) = 1/6$ for $x \in B$ and $\pi_P(x) = 1/8$ for $x \in C$. It is indeed elementary to check that $\pi_P P = \pi_P$. Since the Markov chain X is irreducible on a finite state, the invariant probability exists and is unique, and thus is given by π_P. (One can check that the invariant probability for a uniform random walk on a general finite graph (*i.e.* the next state is chosen uniformly at random among the closest neighbors) is proportional to the degree of the nodes: $\pi(x) = \deg(x)/\sum_y \deg(y)$ for all nodes x of the finite graph.)

Exercise 8.25

1. Conditionally on $\{X_0 = x\}$, τ_1 has a geometric probability distribution with parameter $p = P(x,x) \in [0,1]$. In particular, $\tau_1 = +\infty$ a.s. if $P(x,x) = 1$, that is x is an absorbing state, and $\tau_1 < +\infty$ a.s. otherwise.

2. If $P(x,x) < 1$, we have $\mathbb{P}(X_{\tau_1} = x | X_0 = x) = 0$ and for $y \neq x$:

$$\begin{aligned}\mathbb{P}(X_{\tau_1} = y | X_0 = x) &= \sum_{k=1}^{\infty} \mathbb{P}(X_k = y, \tau_1 = k | X_0 = x)\\ &= \sum_{k=1}^{\infty} P(x,y)P(x,x)^{k-1}\\ &= \frac{P(x,y)}{1 - P(x,x)}.\end{aligned}$$

3. Conditionally on $\{X_{S_n} = x\}$, we deduce from the first question that τ_{n+1} is finite for all $x \in E$, as there is no absorbing state.

4. Let $(\mathcal{F}_n, n \in \mathbb{N})$ be the natural filtration of the process X. Set $\tau_0 = 0$. As a direct consequence of the strong Markov property of the Markov chain X, we get that Y_{n+1} conditionally on $\mathcal{F}_{\tau_n}$ is distributed as Y_{n+1} conditionally on X_{τ_n}. Since $X_{\tau_n} = Y_n$, we deduce that the process Y is a Markov chain with respect to the filtration $(\mathcal{F}_{\tau_n}, n \in \mathbb{N})$, and thus also a Markov chain with respect to its own natural filtration. Since X is homogeneous, we get that Y is also homogeneous, with transition matrix given in the second question.

5. Clearly ν is a probability measure. It is enough to check that $\nu Q = \nu$. Let $y \in E$. We have:

$$\begin{aligned}\nu Q(y) = \sum_{x \in E} \nu(x) Q(x,y) &= \sum_{x \neq y} \frac{\pi(x)(1 - P(x,x))}{\sum_{y \in E} \pi(y)(1 - P(y,y))} \frac{P(x,y)}{1 - P(x,x)}\\ &= \frac{1}{\sum_{y \in E} \pi(y)(1 - P(y,y))} \sum_{x \neq y} \pi(x) P(x,y)\\ &= \frac{1}{\sum_{y \in E} \pi(y)(1 - P(y,y))} (\pi(y) - \pi(y)P(y,y)).\end{aligned}$$

This gives the result.

Exercise 8.26

9.4 Martingales

Exercise 8.27 1. We have $\{\tau_a \leq n\} = \bigcap_{k=1}^n \{|X_n| < a\}$. This implies that τ_a is a stopping time. Since X is an irreducible Markov chain, it is either transient and the time spent in $(-a, a)$ is finite, or recurrent and the number of visit of a is infinite. In both case, we have that X leaves $(-a, a)$ in finite time. That is τ_a is finite. The event $\bigcap_{k=1}^n \{U_{2k} = 1, U_{2k+1} = -1\}$ has positive probability, and, if $a \geq 2$, we have on this event that $\tau_a \geq 2n$. Therefore, τ_a is not bounded.

2. $M^{(\lambda)}$ is clearly adapted. Since $|X_n| \leq n$, we deduce that for fixed n, $M^{(\lambda)}$ is bounded thus integrable. We have, using that U_{n+1} is independent of $\mathcal{F}_n$:

$$\begin{aligned}
\mathbb{E}[M_{n+1}^{(\lambda)} | \mathcal{F}_n] &= \mathrm{e}^{\lambda X_n - (n+1)\varphi(\lambda)}\, \mathbb{E}\left[\mathrm{e}^{\lambda U_{n+1}} \,|\, \mathcal{F}_n\right] \\
&= \mathrm{e}^{\lambda X_n - (n+1)\varphi(\lambda)}\, \mathbb{E}\left[\mathrm{e}^{\lambda U_{n+1}}\right] \\
&= \mathrm{e}^{\lambda X_n - n\varphi(\lambda)} = M_n^{(\lambda)}.
\end{aligned}$$

This implies that $M^{(\lambda)}$ is a martingale.

3. Using the optional stopping theorem, we get that $\mathbb{E}[M_{\tau_a \wedge n}^{(\lambda)}] = \mathbb{E}[M_0^{(\lambda)}] = 1$ for all $n \in \mathbb{N}$. We have that $\lim_{n\to\infty} \tau_a \wedge n = \tau_a$ and as τ_a is finite a.s., that a.s. $\lim_{n\to\infty} X_{\tau_a \wedge n} = X_{\tau_a}$. This implies that a.s. $\lim_{n\to\infty} M_{\tau_a \wedge n}^{(\lambda)} = M_{\tau_a}^{(\lambda)}$.

Since $\varphi(\lambda) \geq 0$ and $|X_{\tau_a \wedge n}| \leq a$, we deduce that $0 \leq M_{\tau_a \wedge n} \leq \mathrm{e}^{|\lambda| a}$. By dominated convergence we get that:

$$\mathbb{E}[M_{\tau_a}^{(\lambda)}] = 1.$$

4. we have $\varphi(\lambda) = \log(\cosh(\lambda))$. Since $\cosh(\lambda) \geq 1$, we deduce that $\varphi \geq 0$. As τ_a is finite, we obtain that $X_{\tau_a} \in \{-a, a\}$. We have, considering first λ and then $-\lambda$, and with $r = \varphi(\lambda) = \varphi(-\lambda)$:

$$\begin{aligned}
\mathrm{e}^{\lambda a}\, \mathbb{E}\left[\mathrm{e}^{-r\tau_a}\, \mathbf{1}_{\{X_{\tau_a}=a\}}\right] + \mathrm{e}^{-\lambda a}\, \mathbb{E}\left[\mathrm{e}^{-r\tau_a}\, \mathbf{1}_{\{X_{\tau_a}=-a\}}\right] = 1, \\
\mathrm{e}^{-\lambda a}\, \mathbb{E}\left[\mathrm{e}^{-r\tau_a}\, \mathbf{1}_{\{X_{\tau_a}=a\}}\right] + \mathrm{e}^{\lambda a}\, \mathbb{E}\left[\mathrm{e}^{-r\tau_a}\, \mathbf{1}_{\{X_{\tau_a}=-a\}}\right] = 1.
\end{aligned}$$

We deduce that:

$$\mathbb{E}\left[\mathrm{e}^{-r\tau_a}\, \mathbf{1}_{\{X_{\tau_a}=a\}}\right] = \frac{\sinh(\lambda a)}{\sinh(2\lambda a)} \quad \text{and} \quad \mathbb{E}\left[\mathrm{e}^{-r\tau_a}\, \mathbf{1}_{\{X_{\tau_a}=-a\}}\right] = \frac{\sinh(\lambda a)}{\sinh(2\lambda a)}.$$

Then use that $2\sinh(x)\cosh(x) = \sinh(2x)$ to deduce that:

$$\mathbb{E}\left[\mathrm{e}^{-r\tau_a}\right] = 2\frac{\sinh(\lambda a)}{\sinh(2\lambda a)} = \frac{1}{\cosh(\lambda a)},$$

and conclude using that $\lambda = \cosh^{-1}(\mathrm{e}^r)$.

Exercise 8.28 1. We have for $y \neq x$, and $r = |y - x|$ that $\mathbb{P}(X_r = y | X_0 = x) = 2^{-r} > 0$. Hence X is irreducible.

2. We consider the natural filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$ where $\mathcal{F}_n = \sigma(U_1, \dots, U_n) = \sigma(X_0, \dots, X_n)$. Recall the return time τ is a stopping time. Since X is a martingale, we get that M is the stopped martingale X at the stopping time τ.

3. Since M is a non-negative martingale, it a.s. converges. On the event $\{\tau = +\infty\}$, we have $|M_{n+1} - M_n| = |U_{n+1}| = 1$. Thus on $\{\tau = +\infty\}$ the martingale M is not converging. We deduce that a.s. τ is finite. This implies that X is recurrent.

4. Since τ is finite a.s., we get $X_\tau = 0$. In particular $0 = \mathbb{E}[X_\tau] = \mathbb{E}[M_\infty] < \mathbb{E}[M_0] = \mathbb{E}[X_0] = 1$. Thus, the optional stopping time theorem does not applied; this implies that τ is not bounded. This also implies that M is not a closed martingale, that is M is not uniformly integrable.

5. Notice that N is $\mathbb{F}$-adapted and integrable (as for all $n \in \mathbb{N}$, $|X_n| \leq n$ and thus $|N_n| \leq n^2$). We have:
$$\begin{aligned}\mathbb{E}[X_{n+1}^2 |\, \mathcal{F}_n] = \mathbb{E}[(X_n + U_{n+1})^2 |\, \mathcal{F}_n] &= X_n^2 + \mathbb{E}[U_{n+1} |\, \mathcal{F}_n] + 2X_n \mathbb{E}[U_{n+1} |\, \mathcal{F}_n] \\ &= X_n^2 + \mathbb{E}[U_{n+1}] + 2X_n \mathbb{E}[U_{n+1}] \\ &= X_n^2 + 1.\end{aligned}$$
This implies that N is a martingale.

6. By the optional stopping theorem, we have that $\mathbb{E}[N_{\tau\wedge n}] = 0$. This gives $\mathbb{E}[M_n^2] = \mathbb{E}[X_{\tau\wedge n}^2] = \mathbb{E}[\tau \wedge n] \leq \mathbb{E}[\tau]$. Since M is not uniformly integrable, it is not bounded in L^2. Thus we have $\sup_{n\in\mathbb{N}} \mathbb{E}[M_n^2] = +\infty$. This implies that $\mathbb{E}[\tau] = +\infty$. Let T the first return time to 0. By decomposing the random walk started at 0 with respect to the first step, and considering only the case where it goes first at 1, we get $T \geq \mathbf{1}_{\{U_1=1\}}(1 + \tau')$, where τ' is distributed as τ and independent of U_1. Therefore we have $\mathbb{E}[T] \geq (1 + \mathbb{E}[\tau])/2 = +\infty$. This implies that X is recurrent null.

Exercise 8.29 1. Let $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$ be the natural filtration associated to M, so that M is $\mathbb{F}$-adapted. Notice that $0 \leq M_n \leq \mathrm{e}^n$, so that M_n is integrable. We also have that:
$$\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n \mathbb{E}[\mathrm{e}^{2X_{n+1}-1}] = M_n,$$
where we used that X_{n+1} is independent of $\mathcal{F}_n$ for the first equality. Thus M is a martingale.

By the strong law of large numbers, we have that a.s. $\lim_{n\to\infty} n^{-1} \sum_{k=1}^n X_k = (1+\mathrm{e})^{-1}$ and, thus a.s. $\lim_{n\to\infty} -n + 2\sum_{k=1}^n X_k = -\infty$ ans a.s. $\lim_{n\to\infty} M_n = 0$.

2. Since $\lim_{n\to\infty} \mathbb{E}[M_n] = 1 > 0 = \mathbb{E}[\lim_{n\to\infty} M_n]$, this implies that M doesn't converge in L^1.

Exercise 8.30 1. An elementary induction gives that $|X_n| \leq n!$ and thus X_n is integrable. The process X is adapted with respect to the natural filtration of the process $(Z_n, n \in \mathbb{N}^*)$. We have:
$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] = \mathbb{E}[Z_{n+1}]\mathbf{1}_{\{X_n=0\}} + (n+1)X_n\mathbb{E}[|Z_{n+1}|]\,\mathbf{1}_{\{X_n\neq 0\}} = X_n\mathbf{1}_{\{X_n\neq 0\}} = X_n,$$
where we used that Z_{n+1} is independent of $\mathcal{F}_n$ for the first equality. We deduce that X is a martingale.

2. We have $\mathbb{P}(X_n \neq 0) = \mathbb{P}(Z_n \neq 0) = n^{-1}$. This implies that X converges in probability towards 0.

3. Set $A_n = \{Z_n \neq 0\}$. The events $(A_n, n \in \mathbb{N}^*)$ are independent and $\sum_{n\in\mathbb{N}^*} \mathbb{P}(A_n) = +\infty$. Borel-Cantelli's lemma implies that the set of $\omega \in \Omega$ such that Card $\{n \in \mathbb{N}^*, \omega \in A_n\} = \infty$ is of probability 1, that is $\mathbb{P}(Z_n \neq 0$ infinitely often$) = 1$. Since $\{X_n \neq 0\} = \{Z_n \neq 0\}$ and X_n belongs to $\mathbb{Z}$, we deduce that $\mathbb{P}(|X_n| \geq 1$ infinitely often$) = 1$. Since X_n converges in probability towards 0, we deduce that $\mathbb{P}(\lim_{n\to\infty} X_n$ exists$) = \mathbb{P}(\lim_{n\to\infty} X_n = 0)$. But this latter quantity is 0 as $\mathbb{P}(|X_n| \geq 1$ infinitely often$) = 1$.

Exercise 8.31 1. Let $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$ be the natural filtration of the process X. We have that, conditionally on $\mathcal{F}_n$, X_{n+1} has a binomial distribution with parameter $(N, X_n/N)$. Notice this proves that X is an homogeneous Markov chain.

2. By definition X is $\mathbb{F}$-adapted. We have that $X_n \in [0, N]$, thus X_n is integrable. We have, using that a binomial (N, p) random variable has mean Np, that:

$$\mathbb{E}[X_{n+1}|\, \mathcal{F}_n] = \mathbb{E}[X_{n+1}|\, X_n] = N\frac{X_n}{N} = X_n.$$

Thus X is a martingale.

3. Since X is a non-negative bounded martingale, it converges a.s. and in L^p for any $p \geq 1$ towards a limit say X_∞.

4. As M_n is a measurable function of X_n, we deduce that M is $\mathbb{F}$-adapted. For $n \in \mathbb{N}$ fixed, since $X_n \in [0, N]$, we deduce that M_n is bounded and thus integrable. Let Y be a binomial (N, p) random variable, then we have $\mathbb{E}[Y] = Np$ and $\mathbb{E}[Y^2] = Np(1-p)+N^2p^2$ and thus:

$$\mathbb{E}[Y(N-Y)] = N^2p - Np(1-p) - N^2p^2 = N(N-1)p(1-p).$$

We deduce that:

$$\begin{aligned}\mathbb{E}[M_{n+1}|\, \mathcal{F}_n] &= \left(\frac{N}{N-1}\right)^{n+1} \mathbb{E}[X_{n+1}(N - X_{n+1})|\, X_n] \\ &= \left(\frac{N}{N-1}\right)^{n+1} N(N-1)\frac{X_n}{N}\left(1 - \frac{X_n}{N}\right) \\ &= M_n.\end{aligned}$$

Thus M is a martingale. By dominated convergence, we have:

$$\mathbb{E}[X_\infty(N - X_\infty)] = \lim_{n\to\infty} \mathbb{E}[X_n(N - X_n)] = \lim_{n\to\infty} \left(\frac{N-1}{N}\right)^n \mathbb{E}[M_n] = 0.$$

5. Since $\mathbb{E}[X_\infty(N - X_\infty)] = 0$, we deduce that a.s. $X_\infty \in \{0, N\}$, that is a.s. an allele disappears. Since the sequence $(X_n, n \in \mathbb{N})$ is $\mathbb{N}$-valued and a.s. converging, it is a.s. constant for n large enough, *i.e.* for $n \geq N_0$, with N_0 a finite random variable. In particular one of the allele has disappeared at time N_0. We have $X_0 = \mathbb{E}[X_\infty] = N\mathbb{P}(X_\infty = N)$, which gives $\mathbb{P}(A$ disappears$) = (N - X_0)/N$.

6. Since $X_n(N - X_n) = 0$ for $n \geq N_0$, we deduce that $M_n = 0$ for $n \geq N_0$. This implies that a.s. $M_\infty = 0$. Since $\mathbb{E}[M_\infty] < M_0$ (unless $X_0 \in \{0, N\}$), we deduce that M does not converge in L^1.

Notice that 0 and N are absorbing states of the Markov chain X, and that $\{1, \dots, N-1\}$ are transient states.

Exercise 8.32 1. Set $Y_n = (X_{n-2}, X_{n-1}, X_n)$ for $n \geq 3$. Since $Y = (Y_n, n \geq 3)$ is a stochastic dynamical system, it is a Markov chain on the finite state space $\{0,1\}^3$. Since $p \in (0,1)$, Y is irreducible. An irreducible Markov chain on a finite state space is positive recurrent, and thus visits infinitely often all the states. In particular the hitting time τ_{ijk} is a.s. finite.

2. Let $\mathbb{F}_n = (\mathcal{F}_n, n \in \mathbb{N})$ be the natural filtration of X, with $\mathcal{F}_0 = \{\emptyset, \Omega\}$ the trivial σ-field. The process $M = (M_n = S_n - n, n \in \mathbb{N})$ is $\mathbb{F}$-adapted. We have $0 \leq S_n \leq \sum_{k=1}^n p^{-k}$. This implies that S_n and M_n are integrable. We have:

$$\mathbb{E}[S_{n+1}|\, \mathcal{F}_n] = (S_n + 1)p^{-1}\mathbb{E}[X_n|\, \mathcal{F}_n] = S_n + 1,$$

where we used that S_n is $\mathcal{F}_n$-measurable for the first equality, and that X_{n+1} is independent of $\mathcal{F}_n$ for the second. We deduce that M is a martingale.

Write τ for τ_{111}. Using the optional stopping theorem, we get that $\mathbb{E}[M_{n\wedge\tau}] = \mathbb{E}[M_0] = 0$. This implies that for all $n \in \mathbb{N}$:

$$\mathbb{E}[n \wedge \tau] = \mathbb{E}[S_{n\wedge\tau}].$$

By monotone convergence, we get that $\lim_{n\to\infty} \mathbb{E}[n \wedge \tau] = \mathbb{E}[\tau]$. Since τ is finite a.s., we have that a.s. $\lim_{n\to\infty} S_{n\wedge\tau} = S_\tau$. It is clear from the dynamic of S that:

$$0 \leq S_{n\wedge\tau} \leq S_\tau = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} \cdot$$

thus, by dominated convergence, we deduce that $\lim_{n\to\infty} \mathbb{E}[S_{n\wedge\tau}] = \mathbb{E}[S_\tau]$. This gives:

$$\mathbb{E}[\tau] = \mathbb{E}[S_\tau] = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} \cdot$$

3. Using the strong Markov property at time τ_{11} for the Markov chain X, we deduce that $(X_{\tau_{11}+n}, n \in \mathbb{N}^*)$ is independent of $(X_k, 1 \leq k \leq \tau_{11})$ and distributed as X. We deduce that:

$$\mathbb{P}(\tau_{111} > \tau_{110}) = \mathbb{P}(X_{\tau_{11}+1} = 0) = \mathbb{P}(X_1 = 0) = 1 - p.$$

4. Arguing as in Question 2, we get that $M = (M_n = T_n - n, n \geq 2)$ is a martingale, with $\mathbb{E}[M_0] = 0$; that for $n < \tau_{110}$, $T_n = 0$ if $X_n = 0$, $T_n = p^{-1}$ if $(X_{n-1}, X_n) = (0,1)$, $T_n = p^{-1} + p^{-2}$ if $(X_{n-1}, X_n) = (1,1)$, and $T_{\tau_{110}} = 1/(p^2(1-p))$; and then that:

$$\mathbb{E}[\tau_{110}] = \mathbb{E}[T_{\tau_{110}}] = \frac{1}{p^2(1-p)} = \frac{1}{p} + \frac{1}{p^2} + \frac{1}{1-p} \cdot$$

5. Consider $U_1 = \dfrac{X_1}{p}$ and $U_n = U_{n-1}\dfrac{1-X_n}{1-p} + \dfrac{X_n}{p}$ for $n \geq 2$, to get:

$$\mathbb{E}[\tau_{100}] = \mathbb{E}[U_{\tau_{100}}] = \frac{1}{p(1-p)^2}\cdot$$

Consider $V_2 = \dfrac{X_1(1-X_2)}{p(1-p)} + \dfrac{X_2}{p}$ and $V_n = V_{n-1}\dfrac{X_n}{p} + \dfrac{X_{n-1}(1-X_n)}{p(1-p)} - \dfrac{X_{n-1}X_n}{p^2} + \dfrac{X_n}{p}$ for $n \geq 3$, to get:

$$\mathbb{E}[\tau_{101}] = \frac{1}{p^2(1-p)}\cdot$$

Exercise 8.33 1. Assume that $\mathbb{E}[X_1] > c$. By the strong law of large numbers, we get that a.s. $\lim_{n\to\infty} S_n/n = c - \mathbb{E}[X_1] < 0$. This implies that a.s. $\lim_{n\to\infty} S_n = -\infty$ and thus τ is a.s. finite.

If $\mathbb{P}(X_1 > c) = 0$, then a.s. $X_k \leq c$ and then $S_n \geq x$ a.s. for all $n \in \mathbb{N}$. This implies that a.s. τ is infinite.

2. The process V is adapted to the natural filtration $\mathbb{F} = (\mathcal{F}_n, n \in \mathbb{N})$ of the process $(X_n, n \in \mathbb{N}^*)$. We have, by independence, that $\mathbb{E}[V_n] = \prod_{k=1}^n \mathbb{E}[\mathrm{e}^{\lambda X_k - \lambda c}] < +\infty$. We also get for $n \in \mathbb{N}$:

$$\mathbb{E}[V_{n+1}|\, \mathcal{F}_n] = V_n\, \mathbb{E}[\mathrm{e}^{\lambda X_{n+1} - \lambda c}\,|\, \mathcal{F}_n] = V_n\, \mathbb{E}[\mathrm{e}^{\lambda X_{n+1} - \lambda c}] \geq V_n.$$

This gives that V is a non-negative sub-martingale.

3. We have:

$$\begin{aligned}
\mathbb{E}[V_N \mathbf{1}_{\{\tau \leq N\}}] &= \sum_{k=1}^N \mathbb{E}[V_N \mathbf{1}_{\{\tau = k\}}] \\
&= \sum_{k=1}^N \mathbb{E}[\mathbb{E}[V_N \mathbf{1}_{\{\tau = k\}}|\, \mathcal{F}_k]] \\
&= \sum_{k=1}^N \mathbb{E}[\mathbf{1}_{\{\tau = k\}} \mathbb{E}[V_N|\, \mathcal{F}_k]] \\
&\geq \sum_{k=1}^N \mathbb{E}[V_k \mathbf{1}_{\{\tau = k\}}] \\
&\geq \mathrm{e}^{\lambda x}\, \mathbb{P}(\tau \leq N),
\end{aligned}$$

where we used that V is a sub-martingale for the first inequality, and that $V_k \geq \mathrm{e}^{\lambda x}$ on $\{\tau = k\}$ for thesecond inequality.

4. The function φ defined on $\mathbb{R}_+$ by $\varphi(\lambda) = \mathbb{E}[\mathrm{e}^{\lambda(X_1 - c)}]$ belongs to $\mathcal{C}^\infty(\mathbb{R}_+)$ (use dominated convergence to prove the continuity and Fubini to prove recursively that $\varphi^{(n)}$ is derivable). Since $\varphi''(\lambda) = \mathbb{E}[(X_1 - c)^2\, \mathrm{e}^{\lambda(X_1 - c)}] > 0$, we deduce that φ is strictly convex. We have $\varphi(0) = 1$ and $\varphi'(0) = \mathbb{E}[X_1 - c] < 0$. As $\mathbb{P}(X_1 > c) > 0$, there exists $a > c$ such

that $p = \mathbb{P}(X_1 \geq a) > 0$. We deduce that $\varphi(\lambda) \geq \mathbb{E}[\mathbf{1}_{\{X_1 \geq a\}} \, \mathrm{e}^{\lambda(X_1 - c)}] \geq p \, \mathrm{e}^{\lambda(a-c)}$, so that $\lim_{\lambda \to \infty} \varphi(\lambda) = +\infty$. Thus, there exists a unique root λ_0 of $\varphi(\lambda) = 1$ on $(0, +\infty)$.

Taking $\lambda = \lambda_0$ in Question 2 , we deduce that V is a martingale. Then, using Question 3 for the inequality, we get that:

$$1 = \mathbb{E}[V_0] = \mathbb{E}[V_N] \geq \mathrm{e}^{\lambda_0 x} \, \mathbb{P}(\tau \leq N).$$

This gives $\mathbb{P}(\tau \leq N) \leq \mathrm{e}^{-\lambda_0 x}$. Then let N goes to infinity to get the result.

Exercise 8.34

Exercise 8.35 1. Since M is a non-negative martingale, it converges a.s. to a limit, say $M_\infty \in L^1$. By Fatou's lemma, we get that:

$$\mathbb{E}[M_0] = \lim_{n \to \infty} \mathbb{E}[M_n] \geq \mathbb{E}[\lim_{n \to \infty} M_n] = \mathbb{E}[M_\infty].$$

2. This is a direct consequence of the first question with the stopped martingale $M^\tau = (M_{\tau \wedge n}, n \in \mathbb{N})$ whose limit is $M^\tau_\infty = M_\tau$.

3. Since $\tau + n$ is a stopping time and the process M is adapted with respect to the filtration $\mathbb{F}$, we deduce that $M_{\tau+n}$ is $\mathcal{F}_{\tau+n}$-measurable. From the previous question, we deduce that $M_{\tau+n}$ is non-negative and integrable. Let $n \in \mathbb{N}$. We have:

$$\begin{aligned} \mathbb{E}\left[M_{\tau+n+1} \,|\, \mathcal{F}_{\tau+n}\right] = \sum_{k \in \bar{\mathbb{N}}} \mathbb{E}\left[M_{k+n+1} \,|\, \mathcal{F}_{k+n}\right] \mathbf{1}_{\{\tau+n=k\}} &= \sum_{k \in \bar{\mathbb{N}}} M_{k+n} \mathbf{1}_{\{\tau+n=k\}} \\ &= M_{\tau+n}. \end{aligned}$$

where for the second equality we used that M is a martingale for $k \in \mathbb{N}$ and that M_∞ is $\mathcal{F}_\infty$-measurable for $k = \infty$.

4. By construction $\tau' - \tau$ is a $\bar{\mathbb{N}}$-valued random variable which is $\mathcal{G}_\infty = \mathcal{F}_\infty$ measurable. We check that $\{\tau' - \tau = n\}$ belongs to $\mathcal{G}_n$ for $n \in \mathbb{N}$. This amounts to check that $\{\tau' - \tau = n\} \cap \{\tau = k\}$ belongs to $\mathcal{F}_{n+k}$ for $n, k \in \mathbb{N}$. This is immediate as:

$$\{\tau' - \tau = n\} \cap \{\tau = k\} = \{\tau' = n + k\} \cap \{\tau = k\}.$$

Thus $\tau' - \tau$ is a stopping time with respect to the filtration $\mathbb{G}$.

5. Using the second question for the martingale N and the stopping time $\tau' - \tau$ (and the convention $\infty - \infty = 0$ so that $N_{\tau' - \tau} = M_{\tau'}$), we obtain (8.2).

9.5 Optimal stopping

Exercise 8.36 Using the optimal equations, see (5.4) and Proposition 5.6, we get that at the first roll of the dice you stop only if you get 5 or 6, at the second you stop only if you get 4, 5 or 6. The average gain of this strategy is 14/3.

9.6 Brownian motion

Exercise 8.37

Exercise 8.38

Exercise 8.39

Chapter 10

Vocabulary

english	*français*
$\mathbb{N}$	$\mathbb{N}$ (but $\mathbb{N}^*$ in some books)
$(0, 1)$	$]0, 1[$
positive	*strictement positif*
countable	*dénombrable*
pairwise disjoint sets	*ensembles disjoints 2 à 2*
a σ-field	*une tribu* ou *σ-algèbre*
a λ-sytem	*une classe monotone*
nested	*emboité(es)*
non-negative	*positif ou nul*
one-to-one (or injective)	*injectif*
onto (or surjective)	*surjectif*
convergence in distribution	*convergence en loi*
pointwise convergence	*convergence simple*
irreducible	*irréductible*
super-martingale	*sur-martingale*
sub-martingale	*sous-martingale*
predictable	*prévisible*
optional stopping theorem	*théorème d'arrêt*
optimal stopping	*arrêt optimal*

Index