Chapter I

A starter on measure theory and random variables

In this chapter, we present in Section I.1 a basic tool kit in measure theory with in mind the applications to probability theory, and in Section I.2 we develop the corresponding integration and expectation. The presentation of this chapter follows closely [6], see also [8].

I.1 Measures and measurable functions

I.1.1 Measurable space

Let Ω be a set also called a space. A measure on a set Ω is function which gives the "size" of subsets of Ω. We shall see that, if one asks the measure to satisfy some natural additive property, 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 I.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) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$; (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 $\mathcal{F}$-measurable.

When there is no ambiguity on the σ-field we shall simply say that A is measurable instead of $\mathcal{F}$-measurable. In a probability setting a measurable set is also called an event. Properties 1 and 2 implies that $\emptyset$ is measurable. Notice that $\mathcal{P}(\Omega)$ and $\{\emptyset, \Omega\}$ are a 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 $\mathcal{P}(\emptyset)$.

Proposition I.2. *Let $\mathcal{C} \subset \mathcal{P}(\Omega)$. There exists a smallest σ-field on Ω which contains $\mathcal{C}$. It is denoted by $\sigma(\mathcal{C})$.*

Proof. Let $(\mathcal{F}_j, j \in J)$ be the collection of σ-fields 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 $\mathcal{C}$. $\square$

Remark I.3. Let $\mathcal{C} = \{A_1, \ldots, A_n\}$, with $n \in \mathbb{N}^*$, be a finite family 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 $A_i = \bigcup_{I\ni i} C_I$. 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 $\mathcal{G}$ are σ-fields, we denote by $\mathcal{F} \vee \mathcal{G}$ the smallest σ-field containing $\mathcal{F}$ and $\mathcal{G}$. More generally, if $(\mathcal{F}_i, i \in I)$ is a family of σ-fields, we denote by $\bigvee_{i\in I} \mathcal{F}_i$ the smallest σ-field containing $\bigcup_{i\in I} \mathcal{F}_i$.

Definition I.4. *If E is a topological space, then the Borel σ-field, $\mathcal{B}(E)$, on E is the smallest σ-field containing 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 $\mathcal{P}(E)$. Vitali[1] give an example of a sub-set of $\mathbb{R}$ which is not a Borel set.

We shall consider product measurable spaces: if $(A_i, i \in I)$ is a collection of set, 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 I.5. *Let $((\Omega_i, \mathcal{F}_i), i \in I)$ be a collection (possibly infinite) of measurable spaces. The corresponding product measurable space is $(\Omega, \mathcal{F})$, with the product space $\Omega = \prod_{i\in I} \Omega_i$ and the product σ-field $\mathcal{F} = \bigotimes_{i\in I} \mathcal{F}_i$ being the smallest σ-field on Ω containing all the product sets $\prod_{i\in I} A_i$ with $A_i \in \mathcal{F}_i$ for all $i \in I$ such that $A_i = \Omega_i$ but for a finite number of indices.*

Remark I.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 the smallest σ-field containing all the intervals (a, b) for $a < b$.

We deduce from the definition of the product σ-field, that the Borel σ-field on $\mathbb{R}^d$, $d \geq 1$, is the smallest σ-field containing $\prod_{i=1}^d (a_i, b_i)$ for all $a_i < b_i$ and $1 \leq i \leq d$. ♢

I.1.2 Measures

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

Definition I.7. *Let $(\Omega, \mathcal{F})$ be a measurable space.*

- *A function μ defined on $\mathcal{F}$ and taking values in $[0, +\infty]$ is σ-additive if for all finite or countable collection $(A_i, i \in I)$ of measurable sets mutually disjoint, 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:*

$$\mu\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mu(A_i). \tag{I.1}$$

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

- *A measure* μ *on* $(\Omega, \mathcal{F})$ *is a* σ*-additive function defined on* $\mathcal{F}$ *and taking values in* $[0, +\infty]$*. We call* $(\Omega, \mathcal{F}, \mu)$ *a measured space.*

- *A probability measure* $\mathbb{P}$ *on* $(\Omega, \mathcal{F})$ *is a measure on* $(\Omega, \mathcal{F})$ *with total mass 1:* $\mathbb{P}(\Omega) = 1$*. The measured space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is also called a probability space.*

- *A measurable set of measure* 0*, is said to be negligible.*

Example I.8. We give some examples of measure (check these are indeed measures). Let Ω be a space.

- The counting measure Card is defined by $A \mapsto \text{Card}\,(A)$ for $A \subset \Omega$, with $\text{Card}\,(A)$ the cardinal of A.

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

- The Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, which will be introduced in Remarks I.19 and I.23, is such that $\lambda([a, b]) = b - a$ for all $a < b$. In particular, any point has Lebesgue measure zero, as well as, by σ-additivity, any countable set.

Notice that the counting measure and the Dirac measures are measures on $(\Omega, \mathcal{F})$ for any σ-field $\mathcal{F}$ on Ω. △

Assuming only the additivity property (that is I is assumed to be finite in (I.1)), instead of the stronger σ-additivity property, for the definition of measures[2] leads to a substantially different and less efficient approach. We give elementary properties of measures.

Proposition I.9. *Let* μ *be a measure on* $(\Omega, \mathcal{F})$*. We have the following properties.*

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

2. *Let* $A, B \in \mathcal{F}$ *such that* $A \subset B$*. We have the monotony property:* $\mu(A) \leq \mu(B)$*.*

3. *Let* $(A_n, n \in \mathbb{N})$ *be a sequence of elements of* $\mathcal{F}$ *such that* $A_n \subset A_{n+1}$ *for all* $n \in \mathbb{N}$*. We have the monotone convergence property:*

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

4. *Let* $(A_i, i \in I)$ *be a finite or countable collection of measurable sets. 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 negligible sets is negligible.*

Proof. We prove property 1. The sets $A \cap B^c$, $A \cap B$ and $A^c \cap B$ are measurable and mutually disjoint. Using the additivity property tree times, we get:

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

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

We prove property 2. As $A^c \cap B \in \mathcal{F}$, we get by additivity that $\mu(B) = \mu(A) + \mu(A^c \cap B)$. Then use $\mu(A^c \cap B) \geq 0$, to conclude.

We prove property 3. We set $B_0 = A_0$ and $B_n = A_n \cap A_{n-1}^c$ for all $n \in \mathbb{N}^*$ so that $\bigcup_{n \leq N} B_n = A_N$ for all $N \in \mathbb{N}^*$ 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 disjoints 2 by 2. By σ-additivity, we get $\mu(A_N) = \mu(\bigcup_{n\leq N} B_n) = \sum_{n\leq N} \mu(B_n)$ 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 $\sum_{n\leq N} \mu(B_n)$, whose terms are non-negative, towards $\sum_{n\in\mathbb{N}} \mu(B_n)$ as N goes to infitiny to conclude.

Property 4 is a direct consequence of properties 1 and 3. $\square$

We end this section with an evident property for probability measures.

Corollary I.10. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For all $A \in \mathcal{F}$, we have $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.*

Proof. Let $A \in \mathcal{F}$. By Definition I.1, we get $A^c \in \mathcal{F}$. Since $\Omega = A \cup A^c$ and $A \cap A^c = \emptyset$, by addivity, we get $\mathbb{P}(A) + \mathbb{P}(A^c) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1$. This gives the result. $\square$

We end this section with the definition of independent events.

Definition I.11. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The events $(A_i, i \in I)$ are said to be independent if for all finite subset $J \subset I$, we have:*

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

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

I.1.3 Caracterisation 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. After introducing monotone classes, we prove the monotone class theorem.

Definition I.12. *A collection $\mathcal{A}$ of sub-sets of Ω is a monotone class if:*

1. *$\Omega \in \mathcal{A}$.*
2. *$A, B \in \mathcal{A}$ and $A \subset B$ imply $B \cap A^c \in \mathcal{A}$.*
3. *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 I.13 (Mononote class Theorem)**.** *Let $\mathcal{C}$ be a collection of sub-sets of Ω stable by finite intersection. All monotone class containing $\mathcal{C}$ also contains $\sigma(\mathcal{C})$.*

Proof. Let $\mathcal{A}$ be the intersection of all monotone classes containing $\mathcal{C}$. It is easy to check that $\mathcal{A}$ is the smallest monotone class containing $\mathcal{C}$. It is clear that $\mathcal{A}$ satisfies properties 1 and 2 from Definition I.1. To check that property 3 from Definition I.1 holds also, so that $\mathcal{A}$ is a σ-field, it is enough, according to property 3 from Definition I.12, to check that $\mathcal{A}$ is stable by finite intersection. Let $B \in \mathcal{C}$. It is easy to check that $\mathcal{A}_B = \{A \subset \Omega; A \cap B \in \mathcal{A}\}$ is a monotone class 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 $A \cap B \in \mathcal{A}$.

Let $B \in \mathcal{A}$. It is easy to check that $\mathcal{A}_B = \{A \subset \Omega; A \cap B \in \mathcal{A}\}$ is a monotone class. 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 I.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 $\mathcal{A} = \{A \in \mathcal{F}; \mathbb{P}(A) = \mathbb{P}'(A)\}$ is a monotone class. It contains $\mathcal{C}$. By the monotone class theorem, it contains $\sigma(\mathcal{C})$. □

The next corollary is an immediate consequence of Definition I.4 and Corollary I.14.

Corollary I.15. *Let E be a topological space. Two probability measures on $(E, \mathcal{B}(E))$ which coincide on the open sets are equal.*

Exercise I.1. 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 $\mathbb{P} = \mathbb{P}'$. △

I.1.4 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 I.16. *A collection, $\mathcal{A}$, of subsets of Ω is called a Boolean algebra if:*

1. $\Omega \in \mathcal{A}$.
2. *If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.*
3. *If $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$.*

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 I.7).

Definition I.17. *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:*

1. *Total mass equal to 1:* $\mathrm{P}(\Omega) = 1$.
2. *Additivity: for all $A, B \in \mathcal{A}$ disjoint,* $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$.

3. *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 [8] or in Section 3 of [1].

Theorem I.18 (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 I.19 (Lebesgue measure)**.** *There exists a 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* $0 \leq a \leq b \leq 1$*.*

Remark I.20. 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 $\mathbb{R}$, $\lambda(A) = \sum_{x\in\mathbb{Z}} \lambda_1((A+x) \cap [0,1))$, 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 I.7). Moreover, we have that $\lambda([a,b]) = \lambda((a,b)) = b - a$ for all $a < b$. $\Diamond$

Let $((\Omega_i, \mathcal{F}_i, \mathbb{P}_i), i \in I)$ be a collection of probability spaces. The Boolean algebra $\mathcal{A}$ of finite unions of sets of the form $\prod_{i\in I} A_i$, where $A_i \in \mathcal{F}_i$ for all $i \in I$ and $A_i = \Omega_i$ but for a finite number of indices, generate the product σ-field $\mathcal{F} = \bigotimes_{i\in I} \mathcal{F}_i$ on the space product $\Omega = \prod_{i\in I} \Omega_i$.

Proposition I.21. *The map* P *defined on sets* $\prod_{i\in I} A_i$*, where* $A_i \in \mathcal{F}_i$ *for all* $i \in I$ *and* $A_i = \Omega_i$ *but for a finite number of indices, by* $\mathrm{P}\left(\prod_{i\in I} A_i\right) = \prod_{i\in I} \mathbb{P}_i(A_i)$ *can be extended as a probability measure on* $(\Omega, \mathcal{A})$*.*

Using the extension theorem, we can prove that the probability measure in Proposition I.21 defined on $\mathcal{A}$ has a unique extension on $\mathcal{F}$, which we denote by $\mathbb{P}$, see [**?**].

Definition I.22. *The probability* $\mathbb{P}$ *defined on* $\mathcal{F} = \bigotimes_{i\in I} \mathcal{F}_i$*, such that* $\mathbb{P}\left(\prod_{i\in I} A_i\right) = \prod_{i\in I} \mathbb{P}_i(A_i)$*, where* $A_i \in \mathcal{F}_i$ *for all* $i \in I$ *and* $A_i = \Omega_i$ *but for a finite number of indices, is called the product probability measure. It is denoted by* $\mathbb{P} = \bigotimes_{i\in I} \mathbb{P}_i$*. The probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is called product probability space.*

Remark I.23. The Lebesgue measure $\lambda_1^{\otimes d}$ on $([0,1)^d, \mathcal{B}([0,1)^d))$, with $d \geq 2$, is the product probability measure $\bigotimes_{i=1}^d \lambda_1$, with λ_1 the Lebesgue measure on $([0,1), \mathcal{B}([0,1)))$. The Lebesgue measure $\lambda^{\otimes d}$ on $\mathbb{R}^d$ is defined by: for all Borel set $A \in \mathbb{R}^d$, $\lambda^{\otimes d}(A) = \sum_{x\in\mathbb{Z}^d} \lambda_1^{\otimes d}(A + x \cap [0,1)^d)$, where $A + x = \{z+x, z \in A\}$. It is clear that the Lebesgue measure is σ-additive and is thus a measure on $\mathbb{R}^d$. $\Diamond$

I.1.5 Measurable functions

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

$$f^{-1}(A^c) = f^{-1}(A)^c, \quad f^{-1}\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f^{-1}(A_i) \quad \text{and} \quad f^{-1}\left(\bigcap_{i \in I} A_i\right) = \bigcap_{i \in I} f^{-1}(A_i). \quad \text{(I.2)}$$

Let $\mathcal{S}$ and $\mathcal{E}$ be σ-fields respectively on S and E.

Definition I.24. *A function f defined on space S and taking values in a space E is measurable from the measurable space $(S, \mathcal{S})$ to the measurable space $(E, \mathcal{E})$ if for all $A \in \mathcal{E}$, $f^{-1}(A) \in \mathcal{S}$.*

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

Example I.25. Let $A \subset S$. The (real-valued) indicator function $\mathbf{1}_A$ is defined by:

$$\mathbf{1}_A(x) = 1 \quad \text{if } x \in A \text{ and} \quad \mathbf{1}_A(x) = 0 \quad \text{otherwise.} \quad \text{(I.3)}$$

If A is measurable (with respect to the σ-field $\mathcal{S}$ on S) then $\mathbf{1}_A$ is measurable (with respect to $(S, \mathcal{S})$ and $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$). △

We deduce from the properties (I.2) and from the definition of a σ-field the following elementary corollary.

Corollary I.26. *Let f be a function from S to E and $\mathcal{E}$ a σ-field on E. The collection $(f^{-1}(A); A \in \mathcal{E})$ is a σ-field on S. It is denoted by $\sigma(f)$. The function f is measurable from $(S, \mathcal{S})$ to $(E, \mathcal{E})$ if and only if $\sigma(f) \subset \mathcal{S}$.*

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

Proposition I.27. *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 $\mathcal{I}$ the σ-field generated by $(f^{-1}(A), A \in \mathcal{C})$. We have $\mathcal{I} \subset \sigma(f)$. It is easy to check that the collection $(A \in E; f^{-1}(A) \in \mathcal{I})$ is a σ-field on E. It contains $\mathcal{C}$ and thus $\mathcal{E}$. This implies that $\sigma(f) \subset \mathcal{I}$ and thus $\sigma(f) = \mathcal{I}$. We conclude using Corollary I.26. □

We deduce the following important result.

Corollary I.28. *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 I.29. *Let $(S,\mathcal{S})$ and $((E_i,\mathcal{E}_i), i\in I)$ be measurable spaces. For all $i\in I$, let f_i 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 and let i_0 denote the only index, if any, such that $A_{i_0}\neq E_{i_0}$. Set $g=(f_i, i\in I)$. We have $g^{-1}\left(\prod_{i\in I} A_i\right)=f_{i_0}^{-1}(A_{i_0})\in\mathcal{S}$. Thus if g is measurable so is f_{i_0}. The converse is a consequence of Proposition I.27. $\square$

The proof of the next proposition is immediate.

Proposition I.30. *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 $\overline{\mathbb{R}}=\mathbb{R}\bigcup\{\pm\infty\}=[-\infty,+\infty]$. 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 $\{-\infty\}$. We say a function is real-valued if it takes values in $\mathbb{R}$ or $\overline{\mathbb{R}}$. 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 I.31. *Let f and g be measurable real-valued functions defined on the same space. The functions fg, $\max(f,g)$ 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 $(x,y)\mapsto xy$, $(x,y)\mapsto\max(x,y)$ 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 I.28. Thus, they are also measurable on $\overline{\mathbb{R}}^2$. The corollary is thus a consequence of Proposition I.30. $\square$

We recall that if $(a_n, n\in\mathbb{N})$ is an $\overline{\mathbb{R}}$-valued sequence then:

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

are well defined and belong to $\overline{\mathbb{R}}$. Furthermore, the sequence $(a_n, n\in\mathbb{N})$ is said to converge in $\mathbb{R}$ (resp. $\overline{\mathbb{R}}$) 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 $\mathbb{R}$ (resp. $\overline{\mathbb{R}}$). The next proposition asserts that the limit of measurable function is measurable.

Proposition I.32. *Let $(f_n, n\in\mathbb{N})$ be a sequence of real-valued measurable functions defined on 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. It is easy to check the σ-field $\mathcal{B}(\overline{\mathbb{R}})$ is generated by $[-\infty, a)$ for $a \in \mathbb{R}$. Notice then that for $a \in \mathbb{R}$:

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

Since the functions f_n are measurable, we deduce that $\{x \in S; \limsup_{n\to+\infty} f_n(x) < a\}$ is also measurable for all $a \in \mathbb{R}$. According to Proposition I.27, we deduce 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 I.31. 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 I.30.

Proposition I.33. *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 measurable function g defined on S taking values in $\overline{\mathbb{R}}$ 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 disjoints 2 by 2 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 function $g_n = 2^{-n} \sum_{k\in\mathbb{Z}} k \mathbf{1}_{B_{n,k}}$ is measurable from S to $\overline{\mathbb{R}}$, and we have $g_n \circ f \leq \varphi \leq g_n \circ f + 2^{-n}$. The function $g = \limsup_{n\to+\infty} g_n$ is measurable according to Proposition I.32, and we have $g \circ f \leq \varphi \leq 2^{-n} + g \circ f$ for all $n \in \mathbb{N}$. This implies that $g \circ f = \varphi$. □

I.1.6 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 I.34. *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.*

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 I.35. *Let $(E, \mathcal{E})$ be a measurable space. A random variable X defined on Ω and taking values in E is a measurable function from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$. Its probability distribution or law is the image probability measure $\mathbb{P}_X$.*

We recall that the events $\{\omega; X(\omega) \in A\} = X^{-1}(A)$ are usually written as $\{X \in A\}$.

We say that two E-valued random variables X and Y are equal in distribution, and we write $X \overset{(d)}{=} Y$, if $\mathbb{P}_X = \mathbb{P}_Y$. We say a random variable is real-valued if it takes values in $\mathbb{R}$ or $\overline{\mathbb{R}}$. If X is 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 I.1 that if X and Y are real-valued random variables, then X and Y are equal in distribution if and only if $F_X = F_Y$.

We say that 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 $\mathbb{P}(X = Y) = 1$.

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. According to Proposition I.29, X_j is an E_j-valued random variable for all $j \in I$. Its marginal probability distribution can be recovered from the distribution of X as

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

We now give the definition of independent random variables.

Definition I.36. *The random variables $(X_i, i \in I)$ are said independent if the σ-fields $(\sigma(X_i),\ i \in I)$ are independent. Equivalently, the random variables $(X_i, i \in I)$ are said independent if for all finite subset $J \subset I$, all $A_j \in \mathcal{E}_j$ with $j \in J$, we have:*

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

We deduce from this definition that if the marginal distributions P_i of all the random variable X_i for $i \in I$ are known and if the $(X_i, i \in I)$ are independent, then the distribution of X is the product probability $\bigotimes_{i\in I} \mathrm{P}_i$ introduced in Definition I.22.

We end this section with the Bernoulli scheme.

Theorem I.37. *Let P be a probability distribution on a measurable space $(E, \mathcal{E})$. Let I be a set of indices. Then, there exists a probability space and an sequence $(X_i, i \in I)$ of random variables defined on this spaces which are independent and with 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 $\mathbb{P}_i = \mathrm{P}$. Consider the product space $\Omega = \prod_{i\in I} \Omega_i$ with the product σ-field and the product probability. For all $i \in I$, we consider the random variable: $X_i(\omega) = \omega_i$ where $\omega = (\omega_i, i \in I)$. We deduce that the random variables $(X_i, i \in I)$ are independent with probability distribution P. $\square$

I.2 Integration, Expectation

Using the results from the integration theory of Sections I.2.1 and I.2.2, we introduce the in Section I.2.4 the expectation of real-valued or $\mathbb{R}^d$-valued random variables and give some well known inequalities. In Section I.2.5 we collect some further results on independence.

I.2.1 Integration: construction and properties

Let $(S, \mathcal{S}, \mu)$ be a measured space. The set $\overline{\mathbb{R}}$ is endowed with the 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 a representation $n \in \mathbb{N}^*$, $\alpha_1, \ldots, \alpha_n \in [0, +\infty]$, $A_1, \ldots, A_n \in \mathcal{S}$ such that $f = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$. The integral of f with respect to μ, denoted by $\int f \, d\mu$ or $\mu(f)$ or $\int f(x)\mu(dx)$, is defined by:

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

Lemma I.38. *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 I.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:

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

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. $\square$

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

$$\mu(af + bg) = a\mu(f) + b\mu(g) \quad \text{for } a, b \in [0, +\infty[\text{ (linearity)}, \tag{I.4}$$

$$f \leq g \Rightarrow \mu(f) \leq \mu(g) \quad \text{(monotony)}. \tag{I.5}$$

Definition I.39. *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:*

$$\mu(f) = \sup\{\mu(g);\, 0 \leq g \leq f \quad g \text{ simple}\}.$$

The next lemma give a representation of $\mu(f)$ using that 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 $\lfloor +\infty \rfloor = +\infty$. The functions $(f_n, n \in \mathbb{N}^*)$ are measurable and their non-decreasing limit is f.

Lemma I.40. *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 $\lim_{n\to+\infty} f_n \geq g$, we have $\lim_{n\to+\infty} \mu(f_n) \geq \mu(g)$. We deduce from the proof of Lemma I.38 that there exists a representation of g such that $g = \sum_{k=1}^N \alpha_k \mathbf{1}_{A_k}$ and the measurable sets $(A_k, 1 \leq k \leq N)$ are mutually 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 monotony, 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 set $B_n = \{x \in A; f_n(x) \geq \alpha'\}$. The sequence $(B_n, n \in \mathbb{N})$ is non-decreasing with A as limit because $\lim_{n\to+\infty} f_n \geq g$. The monotone property for measure, see property 3 from Proposition I.9, implies that $\lim_{n\to+\infty} \mu(B_n) = \mu(A)$. As $\mu(f_n) \geq \alpha' \mu(B_n)$, we deduce that $\lim_{n\to+\infty} \mu(f_n) \geq \alpha' \mu(A)$ and that $\lim_{n\to+\infty} \mu(f_n) \geq \mu(g)$ as $\alpha' \in [0, \alpha[$ is arbitrary. $\square$

Corollary I.41. *The linearity and monotony properties, see (I.4) and (I.5), 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 $f \leq g$. The non-decreasing sequence $(\max(f_n, g_n), n \in \mathbb{N})$ of simple functions converges towards g. By monotony, we get:

$$\mu(f) = \lim_{n\to+\infty} \mu(f_n) \leq \lim_{n\to+\infty} \mu(\max(f_n, g_n)) = \mu(g).$$

$\square$

Definition I.42. *Let $(S, \mathcal{S}, \mu)$ be a measured space. A measurable real-valued function f defined on S is μ-integrable if $\mu(|f|) < +\infty$. The integral of f with respect to the measure μ, denoted by $\mu(f)$ or $\int f \, d\mu$ or $\int f(x) \, \mu(dx)$ is defined by:*

$$\mu(f) = \mu(f^+) - \mu(f^-) \quad \text{with } f^+ = \max(f, 0) \text{ and } f^- = \max(-f, 0).$$

We directly get the following corollary.

Corollary I.43. *The linearity property, see (I.4) with $a, b \in \mathbb{R}$, and the monotony property (I.5) hold for real-valued measurable μ-integrable functions f and g defined on S.*

We deduce that the set $\mathcal{L}^1((S, \mathcal{S}, \mu))$ of real-valued μ-integrable functions defined on S is a vector space. When there is no ambiguity on the state S, we shall write instead $\mathcal{L}^1(\mu)$ and even $\mathcal{L}^1$ if there is no ambiguity on the measure μ. Let $\mathcal{L}^1_{(+)}$, resp. $\mathcal{L}^1_{(-)}$, denote the set of real-valued functions f such that $\mu(f^-) < +\infty$, resp. $\mu(f^+) < +\infty$. The linearity and monotony properties (I.4) and (I.5) hold on $\mathcal{L}^1_{(+)}$ and $\mathcal{L}^1_{(-)}$. We shall say that $\mu(f)$ is well defined if $f \in \mathcal{L}^1_{(+)} \bigcup \mathcal{L}^1_{(-)}$ that is if $\min(\mu(f^+), \mu(f^-))$ is finite. Notice that f is μ-integrable if $\max(\mu(f^+), \mu(f^-))$ is finite.

A property is said to hold 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 I.44. *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 $\{x; f(x) \neq 0\}$ is the non-decreasing limit of the measurable sets $\{x; f_n(x) \neq 0\}$, $n \in \mathbb{N}$, we deduce from the monotony property from Proposition I.9, that $f = 0$ a.e. if and only if $f_n = 0$ a.e. for all $n \in \mathbb{N}$. We get 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 deduce 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 $\mu(f) = 0$. □

We deduce the next corollary which asserts that it is enough to know f a.e. to compute its integral.

Corollary I.45. *Let f and g be two $[0, +\infty]$-valued measurable functions defined on S. If a.e. $f = g$, then we have $\mu(f) = \mu(g)$.*

Proof. By hypothesis the measurable set $A = \{f \neq g\}$ is such that $\mu(A) = 0$. 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, we get:

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

□

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 $\{g,\ f = g \text{ a.e.}\}$.

I.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 I.46. *Let $(f_n, n \in \mathbb{N})$ be a sequence of real-valued measurable functions defined on S. The sequence is said to converge 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 I.32 assures indeed that $\liminf_{n\to+\infty}$ is measurable. We thus deduce the following corollary.

Corollary I.47. *If a sequence of real-valued measurable functions defined on S converges a.s., then its limit is measurable.*

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

Theorem I.48 (Monotone convergence theorem)**.** *Let $(f_n, n \in \mathbb{N})$ be a sequence of real-valued measurable functions defined on S. We assume that, for all $n \in \mathbb{N}$, a.e. $0 \leq f_n \leq f_{n+1}$. We have:*

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

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 of 0 measure as countable union of measurable sets of 0 measure. Thus, we get a.e. $f_n = f_n \mathbf{1}_{A^c}$ for all $n \in \mathbb{N}$. Corollary I.45 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 limit of simple functions which converges towards f_n. We set $g_n = \max\{f_{i,n}; 1 \leq i \leq n\}$. The non-decreasing sequence $(g_n, n \in \mathbb{N})$ converges to f and thus $\lim_{n\to+\infty} \int g_n \, d\mu = \int f \, d\mu$. By monotony, $g_n \leq f_n \leq f$ implies $\int g_n \, d\mu \leq \int f_n \, d\mu \leq \int f \, d\mu$. Taking the limit, we get $\lim_{n\to+\infty} \int f_n \, d\mu = \int f \, d\mu$. □

Exercise I.2. Let $f_n(x) = n^{-1} |x|$ for $n \in \mathbb{N}^*$ and $x \in \mathbb{R}$. Check that $\lim_{n\to+\infty} \int_{\mathbb{R}} f_n(x)dx \neq \int_{\mathbb{R}} \lim_{n\to+\infty} f_n(x)dx$. Does this contradict the monotone convergence theorem? △

It is easy to check the following corollary using the monotone convergence theorem to get the σ-additivity.

Corollary I.49. *Let f be a real-valued measurable functions defined on S. If $f \geq 0$ a.e., then the function $f\mu$ defined on $\mathcal{S}$ by $f\mu(A) = \int \mathbf{1}_A f \, d\mu$ is a measure on $(S, \mathcal{S})$.*

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

Lemma I.50 (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 $n \in \mathbb{N}$. We have the semi-continuity property:*

$$\int \liminf_{n\to+\infty} f_n \, d\mu \leq \liminf_{n\to+\infty} \int f_n \, d\mu.$$

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:

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

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

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

Theorem I.51 (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 $\lim_{n\to+\infty} \int g_n \, d\mu = \int g \, d\mu$ and $\int g \, d\mu < +\infty$. Then, we have:*

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

When $g_n = g$ for all $n \in \mathbb{N}$, the above theorem is called the Lebesgue's theorem.

Proof. As a.e. $|f| \leq g$ and $\int g \, d\mu < +\infty$, we get 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:

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

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

Exercise I.3. Let $f_n(x) = \mathbf{1}_{[n,n+1]}(x)$ for $n \in \mathbb{N}$ and $x \in \mathbb{R}$. Check that $\lim_{n\to+\infty} \int_{\mathbb{R}} f_n(x)dx \neq \int_{\mathbb{R}} \lim_{n\to+\infty} f_n(x)dx$. Does this contradict the monotone convergence theorem, Fatou's lemman or the dominated convergence theorem? △

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

Corollary I.52. *Let $f, g, (f_n, n \in \mathbb{N})$ be real-valued measurable functions defined on S. We assume that $\mu(f_n)$ is well defined and a.e.: $f_n^+ \leq g$ for all $n \in \mathbb{N}$, $f = \lim_{n\to+\infty} f_n$ and that $\int g \, d\mu < +\infty$. Then, we have that $\mu(f)$ is well defined and:*

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

I.2.3 Integration: Fubini theorem and inequalities

We start with the Fubini's 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.

Theorem I.53 (Fubini's theorem)**.** *Let f be a $[0, +\infty]$-valued measurable function defined on $E \times S$.*

1. *For all $x \in E$, the function $y \mapsto f(x, y)$ defined on S is measurable.*
2. *The function $x \mapsto \int f(x, y)\, \mu(dy)$ defined on E is measurable.*
3. *There exists a unique measure on $(E \times S, \mathcal{E} \otimes \mathcal{S})$, denoted by $\nu \otimes \mu$ and called product measure such that:*

$$\nu \otimes \mu(A \times B) = \nu(A)\mu(B) \quad \textit{for all } A \in \mathcal{E},\ B \in \mathcal{S}. \tag{I.6}$$

4. *We have:*

$$\int f(x, y)\, \nu \otimes \mu(dx, dy) = \int \left(\int f(x, y)\, \mu(dy) \right) \nu(dx) \tag{I.7}$$

$$= \int \left(\int f(x, y)\, \nu(dx) \right) \mu(dy). \tag{I.8}$$

Formulas (I.7) and (I.8) holds for any $\overline{\mathbb{R}}$-valued measurable function defined $E \times S$ which are integrable that is $\int |f(x, y)|\, \nu \otimes \mu(dx, dy) < +\infty$.

We shall write $\nu(dx)\mu(dy)$ for $\nu \otimes \mu(dx, dy)$. If the measures ν and μ are probabilities, then the definition of the product measure $\nu \otimes \mu$ coincide with the ones given in Definition I.22.

Proof. Properties 1 and 2 are immediate for the functions $f = \mathbf{1}_C$ with $C = A \times B$, $A \in \mathcal{E}$ and $B \in \mathcal{S}$. The monotone class theorem with Corollary I.31 and the Proposition I.32 imply the result holds for all $C \in \mathcal{E} \otimes \mathcal{S}$. Then the result also holds for every simple function thanks to Corollary I.31, and then for every $[0, +\infty]$-valued measurable functions thanks to Proposition I.32 and the dominated convergence theorem.

For all $C \in \mathcal{E} \otimes \mathcal{S}$, we set $\nu \otimes \mu(C) = \int \left(\int \mathbf{1}_C(x, y)\, \mu(dy) \right) \nu(dx)$. 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 (I.6) holds. The family of sets $A \times B$ where $A \in \mathcal{E}$, $B \in \mathcal{S}$ is stable by finite intersection and generates $\mathcal{E} \otimes \mathcal{S}$. The monotone class theorem implies there exists at most one measure such that (I.6) holds. This ends the proof of property 3.

Property 4 holds clearly for functions $f = \mathbf{1}_C$ with $C = A \times B$, $A \in \mathcal{E}$ and $B \in \mathcal{S}$. Monotone class theorem with Corollary I.31, Proposition I.32 and the monotone convergence theorem imply that the results holds also for all $C \in \mathcal{E} \otimes \mathcal{S}$. We deduce the result for all simple functions thanks to Corollary I.31, and then for all $[0, +\infty]$-valued measurable functions thanks to Proposition I.32 and the monotone convergence theorem.

Let f be a $\overline{\mathbb{R}}$-valued measurable function defined $E \times S$ which is integrable with respect to $\nu \otimes \mu$. We deduce from (I.7) and then (I.8) with f replaced by $|f|$ that $N_E = \{x \in E; \int |f(x,y)| \, \mu(dy) = +\infty\}$ has 0 ν-measure, and then that $N_S = \{y \in S; \int |f(x,y)| \, \nu(dx) = +\infty\}$ has 0 μ-measure. We set $g = f \mathbf{1}_{N_E^c \times N_S^c}$. We can subtract (I.7) with f replaced by $\max(-g, 0)$ to (I.7) with f replaced by $\max(g, 0)$ and we get (I.7) with f replaced by g. Since $\nu \otimes \mu$-a.e. $f = g$, Lemma I.44 implies that (I.7) holds. Equality (I.8) is deduced by symmetry. □

Exercise I.4. Prove that:

$$\int_{]0,1[} \left(\int_{]0,1[} \frac{x^2 - y^2}{(x^2 + y^2)^2} dy \right) dx = \frac{\pi}{4} \cdot$$

Deduce that the function $f(x,y) = \dfrac{x^2 - y^2}{(x^2 + y^2)^2}$ is not integrable with respect to the Lebesque measure on $]0,1[^2$. (Hint: compute the derivative with respect to y of $y/(x^2 + y^2)$.) △

We end this section with very useful inequalities.

Proposition I.54. *Let f and g be two real-valued measurable functions.*

- ***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:*

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

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. $c|f|^p = c'|g|^q$.

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

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

Furthermore, we have $\int fg \, d\mu = \left(\int f^2 \, d\mu \right)^{1/2} \left(\int g^2 \, d\mu \right)^{1/2}$ 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:*

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

Proof. Hölder inequality. The Young inequality states that for $a, b \in [0, +\infty]$, $p, q \in]0, 1[$ such that $\frac{1}{p} + \frac{1}{q} = 1$, we get $ab \leq \frac{1}{p} \, a^p + \frac{1}{q} \, b^q$ (with equality if and only if $a = b$). Indeed, this inequality is obvious if a or b belongs to $\{0, +\infty\}$. For $a, b \in (0, +\infty)$, using the convexity to 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 is trivially true as a.e. $fg = 0$ thanks to Lemma I.44. If this is not the case, then integrating with respect to μ in the Young inequality with $a = |f|/\mu(|f|^p)$ and $b = |g|/\mu(|g|^q)$ 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 and b 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 the equality holds then we get $c|f| = c'|g|$ for some $c, c' \in [0, +\infty)$ such that $(c, c') \neq (0, 0)$, and $\int(|fg| - fg)\, d\mu = 0$. Use Lemma I.44 to conclude that a.e. $|fg| = fg$ and thus a.e. $c\, f = c'\, g$.

The case $p = 1$ of the Minkowski inequality comes from the triangular inequality in $\overline{\mathbb{R}}$. Let $p > 1$. We assume that $\int |f+g|^p\, d\mu > 0$, otherwise the inequality is trivial. Using Hölder inequality, we get:

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

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

I.2.4 Expectation 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 $\mathbb{E}[X]$. The expectation $\mathbb{E}[X]$ is well defined if X is $\mathbb{P}$-integrable that is if $\min(\mathbb{E}[X^+], \mathbb{E}[X^-])$ is finite, where $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$. Recall that X is integrable if by definition $\max(\mathbb{E}[X^+], \mathbb{E}[X^-])$ is finite.

Example I.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 $\mathbb{E}[\mathbf{1}] = 1$. △

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 I.35.

Lemma I.56. *Let X be an random variable taking values in a measured space $(E, \mathcal{E})$. Let φ be a real-valued function defined on $(E, \mathcal{E})$. If $\mathbb{E}[\varphi(X)]$ is well defined or equivalently if $\int \varphi\, \mathrm{P}_X$ is well defined, then we have $\mathbb{E}[\varphi(X)] = \int \varphi\, d\mathrm{P}_X$.*

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

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

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

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 I.57. We explicit the closed formula for the expectation when X is a discrete random variable taking values in $(E, \mathcal{E})$, that is the set $\Delta = \{x \in \overline{\mathbb{R}}, \ \mathbb{P}(X = x) > 0\}$, which is finite or countable, is measurable and $\mathbb{P}(X \in \Delta) = 1$. Let φ be a $[0, +\infty]$-valued function defined on E. Then we have:

$$\mathbb{E}[\varphi(X)] = \sum_{x \in \Delta} \varphi(x) \mathbb{P}(X = x). \tag{I.9}$$

Equation (I.9) also holds for φ a real-valued function as soon as $\mathbb{E}[\varphi(X)]$ is well defined (that is $\min(\mathbb{E}[\varphi^+(X)], \mathbb{E}[\varphi^-(X)]$ is finite). ◊

A real-valued random variable X is square-integrale if X^2 is integrable that is $\mathbb{E}[X^2]$ is finite. Since $2\,|x| \leq 1 + |x|^2$, we deduce from the monotony property of the expectation that if X is square integrable, then it is integrable.

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

We complete the inequalities given in Proposition I.54.

Proposition I.58.

- ***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 real-valued measurable convex function defined on $\mathbb{R}^d$. We have that $\mathbb{E}[\varphi(X)]$ is well defined and:*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]. \tag{I.10}$$

  Furthermore, if φ is strictly convex, the inequality in (I.10) is an equality if and only if X is a.s. constant.

Remark I.59. If X is a real-valued integrable random variable, we deduce from Cauchy-Schwarz inequality or Jensen inequality that $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$. ◊

Proof. Since $\mathbf{1}_{\{|X|\geq a\}} \leq X^2/a^2$, we deduce the Tchebychev inequality from the monotony property of the expectation and Example I.55.

Let $\langle \cdot, \cdot \rangle$ denote the scalar product of $\mathbb{R}^d$. Since the function φ is convex (and thus measurable), for all $a \in \mathbb{R}^d$ there exists $\lambda_a \in \mathbb{R}^d$ such that $\varphi(x) \geq \varphi(a) + \langle \lambda_a, x - a \rangle$ for all $x \in \mathbb{R}^d$. Taking $a = \mathbb{E}[X]$ and $x = X$, we obtain $\varphi(X) \geq \varphi(a) + \langle \lambda_a, X - a \rangle$ and thus $\varphi(X) \geq -|\varphi(a)| - |\lambda_a|(|a| + |X|)$. Since X is integrable, we deduce that $\mathbb{E}[\varphi(X)^-] < +\infty$, where $x^- = \max(-x, 0)$, and thus $\mathbb{E}[\varphi(X)]$ is well defined. Then, using the monotony of the expectation, take the expectation in $\varphi(x) \geq \varphi(a) + \langle \lambda_a, x - a \rangle$ with $a = \mathbb{E}[X]$ to get (I.10).

If φ is strictly convex, then $\varphi(X) > \varphi(a) + \langle \lambda_a, X - a \rangle$ on $\{X \neq a\}$. Taking the expectation in this inequality with $a = \mathbb{E}[X]$, we deduce that (I.10) is an equality if a.s. $X = \mathbb{E}[X]$. □

We give the definition of the variance and covariance. Let X be a real-valued square integrable (and thus integrable) random variable. Its variance $\mathrm{Var}(X)$ is defined by $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. By linearity, we get:

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

It is easy to check that for $a, b \in \mathbb{R}$, we have:

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

Using Lemme I.44 with $f = (X - \mathbb{E}[X])^2$, we deduce that $\mathrm{Var}(X) = 0$ implies there exists a constant a such that a.e. $X = a$.

Let X, Y be two real-valued square-integrable random variables. Thanks to Cauchy-Schwarz inequality, we get that XY is integrable. The covariance of X and Y, $\mathrm{Cov}(X, Y)$, is defined by:

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

Notice that $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$ and by linearity, we get:

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y). \tag{I.11}$$

The covariance can be defined for random vectors as follows.

Definition I.60. *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 random variables with $d, p \in \mathbb{N}^*$. We assume that X and Y are square-integrable. The covariance of X and Y, $\mathrm{Cov}(X, Y)$, is and $d \times p$ matrix defined by:*

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

I.2.5 Independence

Recall the independence of σ-fields given in Definition I.11 and of random variables given in Definition I.36.

Proposition I.61. *Let $n \geq 2$, $(E_1, \mathcal{E}_1), \ldots, (E_n, \mathcal{E}_n)$ be measurable spaces, and $(X_1, \ldots, X_n)$ be a random variable taking values in the product space $\prod_{i=1}^n E_i$ endowed with the product σ-field. The random variables $X_1, \ldots, X_n$ are independent if and only if for all bounded real-valued measurable function f_i defined on E_i, we have:*

$$\mathbb{E}\left[\prod_{i=1}^n f_i(X_i)\right] = \prod_{i=1}^n \mathbb{E}[f_i(X_i)]. \tag{I.12}$$

Proof. If (I.12) is true, then taking $f_i = \mathbf{1}_{A_i}$ with $A_i \in \mathcal{E}_i$, we deduce from Defintions I.11 and I.36 that $X_1, \ldots, X_n$ are independent.

If $X_1, \ldots, X_n$ are independent, then Definition I.11 implies that (I.12) holds for indicator functions. By linearity, we get (I.12) holds also for simple functions. Use monotone convergence theorem to get (I.12) holds also for $[0, +\infty]$-valued measurable functions. Use again linearity, to deduce (I.12) holds for bounded real-valued measurable functions. □

Exercise I.5.

- Extend (I.12) to functions f_i such that $f_i \geq 0$ for all $i \in [\![1, n]\!]$ or to functions f_i such that $f_i(X_i)$ is integrable for all $i \in [\![1, n]\!]$.
- Let X and Y be real-valued integrable random variable. 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 $\mathrm{Cov}(X, Y) = 0$.
- Prove that if $X_1, \ldots, X_n$ are independent real-valued integrable random variables, then $\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i)$.
- 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.

△

I.3 Convergence in distribution

TBD