A review of probability theory foundations

Nicolas Forien, based on notes by Paul Gassiat and Jean-François Le Gall

September 11, 2025

The content of this class is covered in numerous textbooks and lecture notes, see for example [3, 1, 2]. In particular, the lecture notes of Jean-François Le Gall present in full details the proofs of almost all the results stated here. These references should be accessible for free if you are using the wifi network of the university. A French version of the lecture notes of Le Gall is also available on his webpage.

1 Basics of measure theory and integration

Definition 1. *A set $\mathcal{E} \subset \mathcal{P}(E)$ is a **σ-algebra** (also called a **σ-field**) on a set E if it satisfies the three following properties:*

- *The whole set belongs to it: $E \in \mathcal{E}$;*
- *For every $A \in \mathcal{E}$, we have $E \setminus A \in \mathcal{E}$;*
- *If $(A_n)_{n\in\mathbb{N}}$ is a countable family of elements of $\mathcal{E}$, then $\cup_{n\in\mathbb{N}} A_n \in E$.*

*The sets in $\mathcal{E}$ are called the **$\mathcal{E}$-measurable sets** (or simply the measurable sets, when there is no ambiguity on the σ-field considered), and the pair $(E, \mathcal{E})$ is called a **measurable space**.*

Example 1. *If E is any set, then $\{\varnothing, E\}$ and $\mathcal{P}(E)$ are σ-algebras on E.*

Proposition 1. *An intersection of σ-algebras is also a σ-algebra.*

Definition 2. *If $\mathcal{F} \subset \mathcal{P}(E)$, then the **σ-algebra generated by** $\mathcal{F}$ is defined as*

$$\sigma(\mathcal{F}) = \bigcap_{\mathcal{E} \ \sigma\text{-algebra}\ :\ \mathcal{E}\supset\mathcal{F}} \mathcal{E}\,.$$

It is the smallest σ-algebra which contains $\mathcal{F}$.

Definition 3. *If E is a topological space, with family of open sets $\mathcal{O}$, the **Borel σ-algebra**, denoted $\mathcal{B}(E)$, is $\sigma(\mathcal{O})$. In the sequel, when we work on $\mathbb{R}$ or more generally $\mathbb{R}^d$, they will always be equipped with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$.*

Definition 4. *Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. A map $f : E \to F$ is called **measurable** if for every set $A \in \mathcal{F}$ we have $f^{-1}(A) \in \mathcal{E}$.*

Definition 5. *A **measure** on a measurable space $(E, \mathcal{E})$ is a map $\mu : \mathcal{E} \to [0, +\infty]$ such that, if $(A_n)_{n\in\mathbb{N}}$ is a countable family of pairwise disjoint elements of $\mathcal{E}$, then*

$$\mu\bigg(\bigcup_{n\in\mathbb{N}} A_n\bigg) = \sum_{n\in\mathbb{N}} \mu(A_n)\,.$$

*The triplet $(E, \mathcal{E}, \mu)$ is then called a **measured space**.*

Example 2. *On any measurable space $(E, \mathcal{E})$, we can define the counting measure $\mu : A \in \mathcal{E} \mapsto \mathrm{card}(A)$.*

If $x \in E$, the Dirac mass at x is the measure $\delta_x : A \in \mathcal{E} \mapsto 1_{x\in A}$.

*A more interesting example is the **Lebesgue measure** on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, which is the unique measure λ on $\mathcal{B}(\mathbb{R})$ such that $\lambda([a, b]) = b - a$ for any $a \le b$ (its existence is not a trivial fact !).*

Definition 6. *Let μ be a measure on a measurable space $(E, \mathcal{E})$. We say that the measure μ is **finite** if $\mu(E) < \infty$. We say that μ is **sigma-finite** if there exists a countable family $(E_n)_{n\in\mathbb{N}} \in \mathcal{E}^{\mathbb{N}}$ such that $\mu(E_n) < \infty$ for every $n \in \mathbb{N}$ and $E = \cup_{n\in\mathbb{N}} E_n$.*

If $(E, \mathcal{E}, \mu)$ is a measured space, we say that a property holds for μ-almost every $x \in E$ if there exists a set $N \in \mathcal{E}$ such that $\mu(N) = 0$ and the property holds for every $x \in E \setminus N$.

A useful technical tool to prove equality of two measures is Proposition 2 below, which is a consequence of the monotone class lemma, that we present now.

Definition 7. *A **monotone class** is a subset $\mathcal{M} \subset \mathcal{P}(E)$ such that:*

- *The whole set belongs to it: $E \in \mathcal{M}$;*
- *For every $A, B \in \mathcal{M}$, if $A \subset B$ then $B \setminus A \in \mathcal{M}$;*
- *It is stable by non-decreasing countable union, that is to say, if $(A_n)_{n\in\mathbb{N}}$ is a non-decreasing sequence of events in $\mathcal{M}$ (i.e., such that $A_n \subset A_{n+1}$ for every $n \in \mathbb{N}$), then $\cup_{n\in\mathbb{N}} A_n \in \mathcal{M}$.*

Note that an intersection of monotone classes is also a monotone class, which allows to define:

Definition 8. *If $\mathcal{F} \subset \mathcal{P}(E)$, the **monotone class generated by** $\mathcal{F}$ is defined as*

$$\mathcal{M}(\mathcal{F}) = \bigcap_{\mathcal{M} \text{ monotone class } : \ \mathcal{M} \supset \mathcal{F}} \mathcal{M}\,.$$

It is the smallest monotone class which contains $\mathcal{F}$.

We can now state the monotone class lemma:

Theorem 1 (Monotone class lemma)**.** *If $\mathcal{C} \subset \mathcal{P}(E)$ is stable by finite intersections (i.e., if for every $A, B \in \mathcal{C}$ we have $A \cap B \in \mathcal{C}$), then $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.*

Note that a σ-algebra is also a monotone class, so $\sigma(\mathcal{C})$ is a monotone class, so the inclusion $\mathcal{M}(\mathcal{C}) \subset \sigma(\mathcal{C})$ comes for free from the definition. Hence, the interesting part of this theorem is that if $\mathcal{C}$ is stable by finite intersections, then $\mathcal{M}(\mathcal{C})$ is also a σ-field, whence the inclusion $\sigma(\mathcal{C}) \subset \mathcal{M}(\mathcal{C})$.

We will rarely use directly the monotone class theorem, but we often use the following important consequence, to show equality of two measures.

Proposition 2. *Let μ, ν be measures on a measurable space $(E, \mathcal{E})$. Assume that $\mathcal{C} \subset \mathcal{E}$ is stable under finite intersections, satisfies $\sigma(\mathcal{C}) = \mathcal{E}$ and that $\mu(A) = \nu(A)$ for all $A \in \mathcal{C}$.*

Assume moreover that $\mu(E) = \nu(E) < \infty$ (note that this condition is always satisfied if μ and ν are two probability measures) or, more generally, that there exists a sequence $(E_n)_{n\in\mathbb{N}} \in \mathcal{C}^{\mathbb{N}}$ such that $\cup_{n\in\mathbb{N}} E_n = E$ and $\mu(E_n) = \nu(E_n) < \infty$ for every $n \in \mathbb{N}$. Then, we have $\mu = \nu$.

This allows for instance to prove uniqueness of the Lebesgue measure, or that two finite measures on $\mathbb{R}$ coincide if and only if they agree on sets of the form $(-\infty, a]$, for $a \in \mathbb{R}$.

Exercise 1. *Is it true in general that if μ and ν are two σ-finite measures defined on the same measurable space $(E, \mathcal{E})$ which agree on a set $\mathcal{C}$ which is stable by finite intersections and which generates all the σ-field $\mathcal{E}$, then $\mu = \nu$?*

The main use of measures is that they allow to define integrals of measurable functions $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$:

Proposition/Definition 1. *Let $(E, \mathcal{E}, \mu)$ be a measured space. Then there exists a unique way of defining, for every measurable and non-negative function $f : E \to \mathbb{R}_+$, a quantity $\int_E f d\mu \in [0, +\infty]$ (also written $\int_E f(x) d\mu(x)$ or $\int_E f(x)\mu(dx)$ and omitting the set E if there is no ambiguity), such that:*

- *If $A \in \mathcal{E}$, then $\int 1_A d\mu = \mu(A)$;*
- *Linearity: for any measurable functions f, $g \geq 0$ and a, $b \in \mathbb{R}_+$, it holds $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$;*
- *Monotonicity: for any measurable functions f, g with $0 \leq f \leq g$, we have $\int f d\mu \leq \int g d\mu$ (note that in fact this point follows from linearity combined with the fact that $\int f d\mu \geq 0$ for every $f \geq 0$).*

More generally, if f is a measurable function taking values in $\mathbb{R}$, we then say that f is integrable if $\int |f| d\mu < +\infty$, and we then define $\int f d\mu = \int f_+ d\mu - \int f_- d\mu$, where $f_+ = \max(f, 0)$ and $f_- = \max(-f, 0)$. Similarly, if f takes values in $\mathbb{C}$, we say that f is integrable if its real part and its imaginary part are integrable, and we define the integral of f as $\int f d\mu = \int (\operatorname{Re} f) d\mu + i \int (\operatorname{Im} f) d\mu$.

The three following limit theorems are very useful.

Theorem 2. ***(Monotone convergence)*** *Let $(f_n)_{n \in \mathbb{N}}$ be a non-decreasing sequence of measurable and non-negative functions, and let $f = \lim_{n \to \infty} f_n$. Then it holds that*

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

(Fatou's lemma) *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of measurable and non-negative functions, then it holds that*

$$\int \left(\liminf_{n \to \infty} f_n \right) d\mu \;\leq\; \liminf_{n \to \infty} \int f_n d\mu \,.$$

To recall the direction of the inequality, you can have in mind the example of $f_n = 1_{[n, n+1)}$, with the Lebesgue measure on $\mathbb{R}$.

(Dominated convergence) *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of measurable functions and f be a measurable function such that for μ-almost every $x \in E$, we have $f_n(x) \to f(x)$, and assume that there exists an integrable function g such that for every $n \in \mathbb{N}$, for μ-almost every $x \in E$, we have $|f_n(x)| \leq g(x)$. Then, we have*

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

Fatou's lemma can be deduced from the monotone convergence theorem, and the dominated convergence theorem can be proved using Fatou's lemma.

The two following results follow from the dominated convergence theorem, and are useful to study functions defined using an integral.

Proposition 3 (Continuity theorem)**.** *Let $(E, \mathcal{E}, \mu)$ be a measured space, let (F, d) be a metric space and let $f : E \times F \to \mathbb{R}$ be such that:*

- *For every $y \in F$, the map $x \in E \mapsto f(x, y)$ is $\mathcal{E}$-measurable and μ-integrable (so that $\int_E f(x, y) d\mu(x)$ is well defined);*
- *For μ-almost every $x \in E$, the map $y \in F \mapsto f(x, y)$ is continuous on F;*
- *There exists a μ-integrable function $g : E \to \mathbb{R}$ such that for μ-almost every $x \in E$, for all $y \in F$, we have $|f(x, y)| \leq g(x, y)$.*

Then the function $h : y \in F \mapsto \int_E f(x, y) d\mu(x)$ is continuous on F.

Proposition 4 (Derivation under the integral)**.** *Let $(E,\, \mathcal{E},\, \mu)$ be a measured space, let I be an interval of $\mathbb{R}$ and let $f : (x,\, y) \in E \times I \mapsto f(x,\, y) \in \mathbb{R}$ be such that:*

- *For every $y \in I$, the map $x \in E \mapsto f(x,\, y)$ is $\mathcal{E}$-measurable and μ-integrable;*
- *For μ-almost every $x \in E$, the map $y \in I \mapsto f(x,\, y)$ is differentiable on I (we denote by $\partial f / \partial y$ its derivative);*
- *There exists a μ-integrable function $g : E \to \mathbb{R}$ such that for μ-almost every $x \in E$, for all $y \in I$, we have $|\frac{\partial f}{\partial y}(x,\, y)| \leq g(x,\, y)$.*

Then the function $h : y \in I \mapsto \int_E f(x,\, y) d\mu(x)$ is differentiable on I and for every $y \in I$ we have

$$h'(y) \;=\; \int_E \frac{\partial f}{\partial y}(x,\, y)\, d\mu(x)\,.$$

Definition 9. *Let $(E,\, \mathcal{E})$ and $(F,\, \mathcal{F})$ be two measurable spaces. We can define the **product σ-algebra** $\mathcal{E} \otimes \mathcal{F}$, which is a σ-algebra on $E \times F$ given by*

$$\mathcal{E} \otimes \mathcal{F} \;=\; \sigma(\mathcal{E} \times \mathcal{F}) \;=\; \sigma\big(\{A \times B\,,\; A \in \mathcal{E},\, B \in \mathcal{F}\}\big)\,.$$

Definition 10. *Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two measured spaces such that μ and ν are sigma-finite. The **product measure** $\mu \otimes \nu$ is the unique measure on $(E \times F,\, \mathcal{E} \otimes \mathcal{F})$ such that for all $A \in \mathcal{E}$ and $B \in \mathcal{F}$,*

$$(\mu \otimes \nu)(A \times B) \;=\; \mu(A)\nu(B)\,.$$

Theorem 3 (Fubini for non-negative functions)**.** *Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two measured spaces such that both μ and ν are sigma-finite. Let $f : E \times F \to \mathbb{R}$ be a measurable function with respect to $\mathcal{E} \otimes \mathcal{F}$. Then the maps $x \in E \mapsto \int_F f(x,\, y) d\nu(y)$ and $y \in F \mapsto \int_E f(x,\, y) d\mu(x)$ are measurable and*

$$\int_{E\times F} f\, d(\mu \otimes \nu) \;=\; \int_F \left(\int_E f(x,y)\mu(dx) \right) \nu(dy) \;=\; \int_E \left(\int_F f(x,y)\nu(dy) \right) \mu(dx)\,. \tag{1}$$

Theorem 4 (Fubini for integrable functions)**.** *Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two measured spaces such that both μ and ν are sigma-finite. Let $f : E \times F \to \mathbb{R}$ be a measurable function with respect to $\mathcal{E} \otimes \mathcal{F}$. Then the following are equivalent:*

1. *f is integrable with respect to $\mu \otimes \nu$;*
2. *$\int \big(\int |f(x,y)|\, \mu(dx) \big)\, \nu(dy) < \infty$;*
3. *$\int \big(\int |f(x,y)|\, \nu(dy) \big)\, \mu(dx) < \infty$,*

and if this holds, then the maps $x \in E \mapsto \int_F f(x,\, y) d\nu(y)$ and $y \in F \mapsto \int_E f(x,\, y) d\mu(x)$ are integrable and (1) *holds.*

Given a measured space $(E,\, \mathcal{E},\, \mu)$ and a measurable $f \geq 0$, we can always define a new measure ν on $(E,\, \mathcal{E})$ by

$$\nu(A) \;=\; \int_E f 1_A\, d\mu\,.$$

We say that f is the density of ν with respect to μ, also written $f = d\nu/d\nu$.

A measure ν is said to be **absolutely continuous** with respect to μ, (written $\nu \ll \mu$) if for all $A \in \mathcal{E}$, we have the implication $\mu(A) = 0 \implies \nu(A) = 0$. It is easy to check that if ν admits a density with respect to μ, then $\nu \ll \mu$. The converse turns out to be also true.

Theorem 5 (Radon-Nikodym)**.** *Let ν and μ be two σ-finite measures on a measured space $(E,\, \mathcal{E})$, such that $\nu \ll \mu$. Then ν admits a density with respect to μ.*

2 Probability: random variables, independence,

We now fix a **probability space**, namely a measured space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathbb{P}$ is a **probability measure**, i.e., it satisfies $\mathbb{P}(\Omega) = 1$.

In this context, familiar objects from measure theory are given new names:

Definition 11. *An **event** is a measurable set $A \in \mathcal{F}$.*

*An event A holds **almost surely** (abbreviated **a.s.**) if $\mathbb{P}(A) = 1$.*

*A (E-valued) **random variable** X is a measurable map from $(\Omega, \mathcal{F})$ to a measurable space $(E, \mathcal{E})$. (When the arrival set E is not specified, this usually means $E = \mathbb{R}$).*

*The **expectation** or **mean** of a non-negative or integrable random variable X is $\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(d\omega)$.*

*An integrable variable is said to be **centered** if its mean is 0.*

*The **variance** of an integrable variable X is*

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

*The **law** of a E-valued random variable X is the measure image on E, defined for $A \in \mathcal{E}$ by*

$$\mathcal{L}_X(A) \;=\; \mathbb{P} \circ X^{-1}(A) \;=\; \mathbb{P}\big(\{\omega \in \Omega \,:\, X(\omega) \in A\}\big) .$$

*The **cumulant distribution function** of a scalar random variable X is the function $x \in \mathbb{R} \mapsto \mathbb{P}(X \leq x)$. By a remark above, it fully characterizes the law of X.*

Note that in probability theory, the underlying set Ω is typically unimportant (and is often not specified). The important objects are random variables and their properties (such as their laws).

Proposition 5. *Let X be a scalar random variable, let $h : \mathbb{R} \to \mathbb{R}$ be measurable and such that $h(X)$ is integrable. Then*

$$\mathbb{E}\left[h(X)\right] = \int_{\mathbb{R}} h(x)\, \mathcal{L}_X(dx) .$$

For instance, if the law of X admits a density f with respect to Lebesgue measure, then

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}} h(x) f(x)\, dx .$$

Given a random variable X taking values in $(E, \mathcal{E})$, the **σ-algebra generated by** X, denoted $\sigma(X)$, is

$$\sigma(X) \;=\; \big\{X^{-1}(A),\, A \in \mathcal{E}\big\} .$$

It is the smallest sub-σ-algebra of $\mathcal{F}$ for which X is measurable.

Proposition 6. *Let X be a random variable taking values in some measurable space $(E, \mathcal{E})$ and let Y be a $\sigma(X)$-measurable real random variable, then there exists a measurable map $\psi : E \to \mathbb{R}$ such that $Y = \psi(X)$.*

Definition 12 (L^p spaces)**.** *Fix $1 \leq p < \infty$. Given a random variable X, its L^p norm is defined by*

$$\|X\|_{L^p} \;=\; \mathbb{E}\big[|X|^p\big]^{1/p} .$$

We then define the set

$$\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P}) \;=\; \big\{X \,:\, \Omega \to \mathbb{R} \text{ measurable such that } \|X\|_{L^p} < \infty\big\} .$$

The L^p norm is not a norm on this set because $\|X\|_{L^p} = 0$ only implies that $X = 0$ almost surely, so we define the space $L^p(\Omega, \mathcal{F}, \mathbb{P})$ which is obtained by taking the quotient by the equivalence relation $X \sim Y$ if and only if $X = Y$ almost surely. Equipped with the L^p norm, it is a Banach space.

We can also define the L^∞ norm (and the corresponding L^∞ space)

$$\|X\|_{L^\infty} \;=\; \operatorname{ess\,sup} |X| \;=\; \inf\big\{c \in \mathbb{R} \;:\; \mathbb{P}(|X| \leq c) = 1\big\} .$$

We record the following important inequalities for expectations of random variables.

Proposition 7. ***(Jensen)*** *If $\phi : \mathbb{R} \to \mathbb{R}$ is convex and X is a real random variable, then*

$$\phi\left(\mathbb{E}[X]\right) \;\le\; \mathbb{E}\left[\phi(X)\right]$$

(as long as the two expectations above are well defined).

Cauchy-Schwarz) *For any two real random variables X and Y,*

$$\mathbb{E}\big[|XY|\big] \;\le\; \|X\|_{L^2}\,\|Y\|_{L^2}\ .$$

(Hölder) *Fix $1 \le p,\, q \le \infty$ such that $1/p + 1/q = 1$. Then for any two real random variables X and Y,*

$$\mathbb{E}\big[|XY|\big] \;\le\; \|X\|_{L^p}\,\|Y\|_{L^q}\ .$$

Note that it follows from Jensen's inequality that $\|\cdot\|_{L^p} \le \|\cdot\|_{L^q}$ if $p \le q$.

A simple but often very efficient way of measuring probabilities is given by the following proposition.

Proposition 8. *Let $U : \mathbb{R} \to \mathbb{R}$ be a non-decreasing positive function. Then for any real random variable X such that the below expectation make sense, for any $a \in \mathbb{R}$, it holds that*

$$\mathbb{P}\left(X \ge a\right) \;\le\; \frac{\mathbb{E}[U(X)]}{U(a)}\ .$$

Proof. The result follows from taking the expectation in the inequality $U(a)1_{\{X \ge a\}} \le U(X)$. □

This implies the following classical special cases:

$$\forall a > 0 \qquad \mathbb{P}(|X| \ge a) \;\le\; \frac{\mathbb{E}[|X|]}{a} \qquad \text{(Markov)}\ ,$$

$$\forall a > 0 \qquad \mathbb{P}\big(|X - \mathbb{E}[X]| \ge a\big) \;\le\; \frac{\mathrm{Var}(X)}{a^2}\ \text{(Bienaymé-Tchebychev)}\ ,$$

$$\forall \lambda > 0 \quad \forall a \in \mathbb{R} \qquad \mathbb{P}(X \ge a) \;\le\; \mathbb{E}\big[e^{\lambda X}\big]\, e^{-\lambda a}\ \text{(Chernoff)}\ .$$

Definition 13. *Two events $A, B \in \mathcal{F}$ are **independent** if*

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

A family $(A_i)_{i\in I}$ of events is independent if for every $n \ge 1$, for any choice of indices $i_1, \dots, i_n \in I$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_n}) \;=\; \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_n})\,.$$

Similarly, two random variables X and Y taking values respectively in $(E,\, \mathcal{E})$ and $(E',\, \mathcal{E}')$ are independent if for every $A \in \mathcal{E}$ and every $B \in \mathcal{E}'$ we have

$$\mathbb{P}(X \in A,\, Y \in B) \;=\; \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B)\,.$$

A family $(X_i)_{i\in I}$ of random variables taking values in whatever measurable spaces $(E_i,\, \mathcal{E}_i)$ is independent if for every $n \ge 1$, for any choice of indices $i_1, \dots, i_n \in I$ and of measurable sets $A_1 \in \mathcal{E}_{i_1}, \dots, A_n \in \mathcal{E}_{i_n}$, it holds that

$$\mathbb{P}(X_{i_1} \in A_1, \dots, X_{i_n} \in A_n) \;=\; \mathbb{P}(X_{i_1} \in A_1) \cdots \mathbb{P}(X_{i_n} \in A_n)\,.$$

A family $(\mathcal{G}_i)_{i\in I}$ of σ-algebras is independent if any family $(A_i)_{i\in I}$ of events with $A_i \in \mathcal{G}_i$ for every $i \in I$ is independent.

Note that independence of a family is stronger than pairwise independence of its elements.

Note that it follows from the definition that two variables X and Y are independent if and only if the law of (X, Y) is the product measure $\mathcal{L}_X \otimes \mathcal{L}_Y$. A similar result is true for families of random variables. In particular, in conjunction with Proposition 5 and Fubini's theorem, this implies that if X and Y are independent and f and g are two measurable functions such that the expectations below make sense, then

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

Given a sequence $(A_n)_{n\in\mathbb{N}}$ of events, we define

$$\limsup_n A_n \;=\; \bigcap_{n\in\mathbb{N}}\Big(\bigcup_{k\geq n} A_k\Big) \;=\; \big\{\omega\in\Omega \;:\; \omega \text{ belongs to infinitely many events } A_n\big\}\,.$$

Proposition 9 (Borel-Cantelli's Lemma)**.** *(1) Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events such that $\sum_{n\in\mathbb{N}}\mathbb{P}(A_n) < \infty$. Then $\mathbb{P}(\limsup A_n) = 0$.*

(2) Let $(A_n)_{n\in\mathbb{N}}$ be an independent sequence of events such that $\sum_{n\in\mathbb{N}}\mathbb{P}(A_n) = \infty$. Then $\mathbb{P}(\limsup A_n) = 1$.

Proof. (1) By Fubini,

$$\mathbb{E}\left[\sum_{n=0}^{\infty} 1_{A_n}\right] \;=\; \sum_{n=0}^{\infty}\mathbb{P}(A_n) \;<\; \infty\,,$$

which implies that $\mathbb{P}(\sum_n 1_{A_n} < \infty) = 1$, which is the claim.

(2) Note that

$$\Omega\setminus(\limsup_n A_n) \;=\; \bigcup_{n\in\mathbb{N}}\Big(\bigcap_{k\geq n}(\Omega\setminus A_k)\Big)\,,$$

and since this is an increasing union, we have

$$\mathbb{P}\big(\Omega\setminus(\limsup_n A_n)\big) \;=\; \lim_{n\to\infty}\mathbb{P}\Big(\bigcap_{k\geq n}(\Omega\setminus A_k)\Big)\,.$$

Let $n\in\mathbb{N}$ be fixed. For every $p\in\mathbb{N}$, using independence we can write

$$\mathbb{P}\Big(\bigcap_{k\geq n}(\Omega\setminus A_k)\Big) \;\leq\; \mathbb{P}\Big(\bigcap_{k=n}^{n+p}(\Omega\setminus A_k)\Big) \;=\; \prod_{k=n}^{n+p}(1-\mathbb{P}(A_k)) \;\leq\; \exp\Big(-\sum_{k=n}^{n+p}\mathbb{P}(A_k)\Big)\,,$$

using the inequality $1-x\leq e^{-x}$. This goes to 0 when $p\to\infty$ by assumption, which shows that for every $n\in\mathbb{N}$,

$$\mathbb{P}\Big(\bigcap_{k\geq n}(\Omega\setminus A_k)\Big) \;=\; 0\,,$$

which implies the result. □

In fact, (2) above only holds under pairwise independence. Let us give this as an exercise.

We first record the important fact:

Lemma 1. *Let $X_1,\ldots,X_n$ be pairwise independent random variables in $L^2(\Omega,\mathcal{F},\mathbb{P})$. Then* $\mathrm{Var}(X_1+\cdots+X_n) = \mathrm{Var}(X_1)+\cdots+\mathrm{Var}(X_n)$.

Exercise 2. *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of pairwise independent events such that $\sum_n\mathbb{P}(A_n) = \infty$. Let $S_n = \sum_{k=0}^{n} 1_{A_k}$, $S = \lim_{n\to\infty} S_n$ and $m_n = \mathbb{E}[S_n]$ (which converges to $+\infty$ by assumption). Show that* $\mathrm{Var}(S_n)\leq m_n$. *Deduce from Chebychev's inequality that $\mathbb{P}(S\leq m_n/2)\leq 4/m_n$, and conclude that $\mathbb{P}(S=\infty) = 1$.*

3 Convergence of random variables

Let X, and X_n, $n \in \mathbb{N}$ be some random variables defined on the same probability space.

Definition 14. *We say that X_n converges to X:*

in probability *if for every $\varepsilon > 0$, we have $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$.*

in *L^p (for a given $p \geq 1$) if $\|X_n - X\|_{L^p} \to 0$.*

almost surely *if $\mathbb{P}(X_n \to X) = 1$.*

in law *or* ***in distribution*** *if for every bounded and continuous function ϕ, we have $\mathbb{E}[\phi(X_n)] \to \mathbb{E}[\phi(X)]$.*

Note that convergence in law is really a property of the laws of the random variables, not of the random variables themselves, unlike the other modes of convergence.

Proposition 10. *We have the following implications between the different modes of convergence:*

1. *If $X_n \to X$ almost surely then $X_n \to X$ in probability.*
2. *If $X_n \to X$ in L^p and $q \leq p$ then $X_n \to X$ in L^q.*
3. *If $X_n \to X$ in L^1 then $X_n \to X$ in probability.*
4. *If $X_n \to X$ in probability then $X_n \to X$ in law.*
5. *If $X_n \to X$ in law and the variable X is constant, then $X_n \to X$ in probability.*
6. *If $X_n \to X$ in probability, then there exists a subsequence $(X_{\varphi(n)})_{n\in\mathbb{N}}$ which converges almost surely to X.*
7. *(Dominated convergence theorem) If $X_n \to X$ almost surely and there exists an integrable variable Z such that $|X_n| \leq Z$ for every $n \in \mathbb{N}$, then $X_n \to Z$ in L^1.*

Note that Theorem 18 below gives yet another partial implication.

Proof. 1. The first implication follows from the dominated convergence theorem applied to the functions $f_n = 1_{\{|X_n - X| > \varepsilon\}}$.

2. The second implication follows from Jensen's inequality.

3. The third implication follows from Markov's inequality, which entails that $\mathbb{P}(|X_n - X| > \varepsilon) \leq \|X_n - X\|_{L^1}/\varepsilon$.

5. If $X_n \to c \in \mathbb{R}$ in law and $\varepsilon > 0$ then, taking ϕ a continuous and bounded function such that $\phi(c) = 0$ and $\phi(x) = 1$ if $|x - c| \geq \varepsilon$, we have $\mathbb{P}(|X_n - c| \geq \varepsilon) \leq \mathbb{E}[\phi(X_n)] \to \phi(c) = 0$.

6. Using that $X_n \to X$ in probability we can construct an extraction φ (i.e., $\varphi : \mathbb{N} \to \mathbb{N}$ is strictly increasing) such that for every $n \in \mathbb{N}$,
$$\mathbb{P}\left(|X_{\varphi(n)} - X| \geq \frac{1}{n}\right) \leq \frac{1}{n^2}.$$
By the Borel-Cantelli lemma, this implies that almost surely, for n large enough, $|X_{\varphi(n)} - X| < n/n$, which implies that $X_{\varphi(n)} \to X$.

4. Assume that $X_n \to X$ in probability, and let f be a continuous and bounded function. Since the sequence $(\mathbb{E}[f(X_n)])_{n\in\mathbb{N}}$ is bounded, to show that it converges to $\mathbb{E}[f(X)]$, it is enough to show that all its converging subsequences have this limit. Thus, we consider a subsequence $(X_{\varphi(n)})_{n\in\mathbb{N}}$, we assume that the sequence $(\mathbb{E}[f(X_{\varphi(n)})])_{n\in\mathbb{N}}$ converges to some limit $\ell \in \mathbb{R}$ when $n \to \infty$. Since $X_{\varphi(n)} \to X$ in probability, using the implication 6 we can find a subsequence $(X_{\varphi(\psi(n))})_{n\in\mathbb{N}}$ which converges almost surely to X. Then, we have
$$\ell = \lim_{n\to\infty} \mathbb{E}\big[f(X_{\varphi(n)})\big] = \lim_{n\to\infty} \mathbb{E}\big[f(X_{\varphi(\psi(n))})\big] = \mathbb{E}[f(X)].$$
where the last equality follows from the dominated convergence theorem.

□

In general, there are no other implications between the various notions of convergence (exercise: find counterexamples).

We will now spend more time on the convergence in law.

The **characteristic function** of (the law of) a real random variable X is defined by

$$\phi_X \;:\; t \in \mathbb{R} \;\mapsto\; \mathbb{E}[e^{itX}]\,.$$

Note that for any random variable X, it is a continuous and bounded function on $\mathbb{R}$.

It can be linked to moments of X in the following way:

Proposition 11. *Assume that $\mathbb{E}|X|^k < \infty$. Then ϕ_X is C^k on $\mathbb{R}$, and $\phi^{(k)}(0) = i^k \mathbb{E}[X^k]$.*

Proof. Exercise (use differentiation under $\mathbb{E}$) □

The main utility of characteristic functions comes from the following result.

Theorem 6 ((Lévy's Continuity Theorem))**.** *The following are equivalent:*

(1) $X_n \to X$ in law,

(2) For every $t \in \mathbb{R}$, $\phi_{X_n}(t) \to \phi_X(t)$.

Proof. The implication (1) ⇒ (2) follows from the definitions. We give the fact that (2) ⇒ (1) as an exercise, with main steps sketched:

- Show that
$$\mathbb{P}(|X| \geq r) \leq \frac{r}{2} \int_{-2/r}^{2/r} (1 - \Phi_X(t))\, dt$$
(hint: use Fubini's theorem to evaluate $\int_{-c}^{c} (1 - \Phi_X(t)) dt$ and use that $|\sin(x)| \leq |x|/2$ for $|x| \geq 2$).
- Deduce that if (2) holds, then for any $\varepsilon > 0$, for r large enough and n large enough we have $\mathbb{P}(|X_n| \geq r) \leq \varepsilon$.
- Use that for any $R > r > 0$, functions of the form $x \mapsto \sum_{k=-N}^{N} a_k e^{i\frac{k\pi x}{R}}$ are dense in the set of continuous function on $[-r, r]$, in combination with the previous step, to conclude.

□

Recall that the cumulant distribution function of X is defined by $F_X(x) = \mathbb{P}(X \leq x)$. This can also be used for convergence in law.

Theorem 7. *$X_n \to X$ in law if and only if for every $x \in \mathbb{R}$, if F_X is continuous at x then $F_{X_n}(x) \to F_X(x)$.*

Proof. Left in exercise (as in the previous proof: first deal with the tails of the X_n to reduce to a compact set). □

We also record the important compactness criterion for weak convergence:

Definition 15. *A family $(X_i)_{i \in I}$ of random variables is **tight** if, for any $\varepsilon > 0$, there exists a compact set K, such that for every $i \in I$ we have $\mathbb{P}(X_i \notin K) \leq \varepsilon$.*

Note that any finite family is tight. The main interest of this notion is that it characterizes (sequential) compactness for convergence in law.

Theorem 8. *(1) If $X_n \to X$ in law, then $(X_n)_{n \in \mathbb{N}}$ is tight.*

*(2) **(Prokhorov)** If $(X_n)_{n \in \mathbb{N}}$ is tight, then there exists a random variable X (possibly defined on a different probability space) and a subsequence $(X_{n_k})_{n \in \mathbb{N}}$ such that $X_{n_k} \to X$ in law.*

We prove the theorem with the help of the following lemma.

Lemma 2. *A function $F : \mathbb{R} \to \mathbb{R}$ is the cumulant distribution function of a random variable X if and only if:*

1. *F in non-decreasing;*
2. *F is cadlag (right-continuous and with left limits);*
3. *$\lim_{-\infty} F = 0$ and $\lim_{+\infty} F = 1$.*

Proof. Exercise. Hint: for the "if" direction, consider, $X = G(U)$, where $G(y) = \sup\{x : F(x) \le y\}$, and U is a uniform random variable on $[0, 1]$. Note: this may be in practice a useful way to simulate a random variable whose cumulant distribution function is known. □

Proof of Theorem 8. (1) is easier and left as exercise.

(2) (Sketch). By a diagonal procedure, we construct an extraction φ such that for each rational number q, $F_{X_{\varphi(n)}}(q)$ converges to some limit that we call $F_0(q)$. We then let for every $x \in \mathbb{R}$, $F(x) := \lim_{q\in\mathbb{Q}, q>x} F_0(q)$. We check that this function F satisfies the assumptions of Lemma 2 (tightness is only used to show that it satisfies point 3.). □

4 The Law of Large Numbers and the Central Limit Theorem

In this section, we consider a sequence $(X_n)_{n\ge1}$ of independent real random variables which all have the same law. This is called an **i.i.d.** sequence (for "independent and identically distributed").

4.1 The Law of Large Numbers

Theorem 9 (Weak Law of Large Numbers)**.** *Assume that $(X_n)_{n\ge1}$ is an i.i.d. sequence of real integrable variables, and let $m = \mathbb{E}[X_1]$. Then*

$$\frac{X_1 + \cdots + X_n}{n} \overset{n\to\infty}{\longrightarrow} m \qquad \textit{in probability.}$$

Proof. Since the limit is a constant, it suffices to check convergence in law, which can be done through the characteristic function. For $n \ge 1$, let us write $S_n = X_1 + \cdots + X_n$. Then, for every $t \in \mathbb{R}$ we have

$$\phi_{S_n/n}(t) \;=\; \mathbb{E}\Big[\exp\Big(it\frac{S_n}{n}\Big)\Big] \;=\; \phi_{X_1}(t/n)^n \;=\; \Big(1 + im\frac{t}{n} + o\Big(\frac{t}{n}\Big)\Big)^n \overset{n\to\infty}{\longrightarrow} e^{itm}\,.$$

□

In fact, the above can be strengthened to almost sure convergence.

Theorem 10 (Strong Law of Large Numbers)**.** *Under the same hypotheses than in Theorem 9, we have*

$$\frac{X_1 + \cdots + X_n}{n} \overset{n\to\infty}{\longrightarrow} m \qquad \textit{almost surely.}$$

Many different proofs of this result exist. One proof uses the following result, which can also be usefull in other situations.

Theorem 11 (Kolmogorov's 0-1 law)**.** *Let $(\mathcal{G}_n)_{n\in\mathbb{N}}$ be an independent sequence of sub σ-algebras of $\mathcal{F}$. We consider the terminal σ-algebra of this sequence, which is*

$$\mathcal{G} \;=\; \bigcap_{n\in\mathbb{N}} \sigma(\mathcal{F}_k,\, k \ge n)\,.$$

Then this σ-algebra is trivial, that is to say, for every event $A \in \mathcal{G}$, we have $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

Proof. The idea is to show that $\mathcal{G}$ is independent of itself, and thus $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, which implies that $\mathbb{P}(A)$ is 0 or 1. A detailed proof can be found for example in the notes of Jean-François Le Gall [3]. □

Proof of Theorem 10, the Strong Law of Large Numbers. To simplify, we assume that $m = 0$. For every $n \geq 1$, we write $S_n = X_1 + \cdots + X_n$. Let $a > 0$ and let $M = \sup_{n\in\mathbb{N}}(S_n - na)$. We aim to show that $M < \infty$ almost surely, which implies that, almost surely,

$$\limsup_{n\to\infty} \frac{S_n}{n} \;\leq\; \limsup_{n\to\infty} \left(a + \frac{M}{n}\right) \;=\; a\,.$$

If we show this for every $a > 0$, this would entail that $\limsup S_n/n \leq 0$, and the other bound $\liminf S_n/n \geq 0$ would follow by symmetry. Thus, $a > 0$ being fixed, there remains to show that $\mathbb{P}(M < \infty) = 1$.

For $n \in \mathbb{N}$, let us define

$$M_n \;=\; \max_{0\leq k\leq n}(S_n - na) \;=\; \max_{0\leq k\leq n}(X_1 + \cdots + X_n - na)$$

and

$$M'_n \;=\; \max_{0\leq k\leq n}(X_2 + \cdots + X_{n+1} - na) \;=\; \max_{0\leq k\leq n}(S_{k+1} - X_1 - na)\,.$$

First, note that M_n and M'_n have the same law. Besides, for every $n \in \mathbb{N}$ we can write

$$M_{n+1} \;=\; \max\big(0,\, \max_{1\leq k\leq n}(S_k - ka)\big) \;=\; \max(0,\, M'_n + X_1 - a) \;=\; M'_n - \min(M'_n,\, a - X_1)\,.$$

Taking the expectation, we obtain

$$\mathbb{E}\big[\min(M'_n,\, a - X_1)\big] \;=\; \mathbb{E}[M'_n] - \mathbb{E}[M_{n+1}] \;=\; \mathbb{E}[M_n] - \mathbb{E}[M_{n+1}] \;\leq\; 0\,,$$

because by definition we have $M_{n+1} \geq M_n$. Yet, the dominated convergence theorem, with domination by the integrable variable $|a - X_1|$, implies that

$$\mathbb{E}\big[\min(M'_n,\, a - X_1)\big] \;\overset{n\to\infty}{\longrightarrow}\; \mathbb{E}\big[\min(M',\, a - X_1)\big]\,,$$

where $M' = \lim M'_n$. Thus, we deduce that

$$\mathbb{E}\big[\min(M',\, a - X_1)\big] \;\leq\; 0\,.$$

If we had $M' = \infty$ almost surely, then the minimum above would be alsmot surely equal to $a - X_1$, and this would lead to a contradiction because we have $\mathbb{E}[a - X_1] = a > 0$. Therefore, we must have $\mathbb{P}(M' = \infty) < 1$. And since M' and M have the same law, the same holds for M. This is where the 0-1 law comes into play: the event $\{M = \infty\}$ belongs to the terminal σ-field

$$\bigcap_{n\geq 1} \sigma(X_k,\, k \geq n)$$

and the variables $(X_n)_{n\in\mathbb{N}}$ are independent. Hence, Kolmogorov's 0-1 law ensures that $\mathbb{P}(M = \infty)$ can only be 0 or 1, so we deduce that $\mathbb{P}(M = \infty) = 0$, that is to say, $M < \infty$ almost surely, which is what we wanted to show. □

4.2 The Central Limit Theorem

Definition 16. *The standard Gaussian measure, denoted by $\mathcal{N}(0,1)$, is the probability measure on $\mathbb{R}$ with density function given by $x \mapsto e^{-x^2/2}/\sqrt{2\pi}$.*

Exercise: check that the above is a well-defined probability measure (hint: compute $\int e^{-x^2-y^2}dxdy$ via polar coordinates). Further check that if Z has law $\mathcal{N}(0,1)$, then $\mathbb{E}[Z] = 0$, $\mathbb{E}[Z^2] = 1$, and the characteristic function is given by $\phi_Z(t) = e^{\frac{-t^2}{2}}$ (hint: use integration by parts to show that $\phi'_Z(t) = -t\phi_Z(t)$).

Theorem 12 (Central Limit Theorem)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of variables with $\mathbb{E}|X_1|^2 < \infty$, and let $m = \mathbb{E}[X_1]$ and $\sigma^2 = Var(X_1)$. For every $n \geq 1$, we write $S_n = X_1 + \cdots + X_n$. Then we have the convergence in law*

$$\frac{S_n - mn}{\sigma\sqrt{n}} \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)\,.$$

Proof. Let us assume that $m = 0$ and $\sigma = 1$ (the general case follows by considering $Y = (X - m)/\sigma$).

Since the variables are i.i.d., it holds that

$$\phi_{S_n/\sqrt{n}}(t) \;=\; \phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n \;=\; \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \overset{n\to\infty}{\longrightarrow} e^{-t^2/2}\,.$$

□

5 Conditional expectations

Let us start with some preliminary definitions. Let $(\Omega,\, \mathcal{F},\, \mathbb{P})$ be a probability space.

If $A,\, B \in \mathcal{F}$ are two events with $\mathbb{P}(B) > 0$, then we can define the conditional probability of A knowing B, which is

$$\mathbb{P}(A|B) \;=\; \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\,.$$

If $X : \Omega \to \mathbb{R}$ is an integrable variable and $B \in \mathcal{F}$ is an event with $\mathbb{P}(B) > 0$, we can also define the conditional expectation of X knowing B:

$$\mathbb{E}[X|B] \;=\; \frac{\mathbb{E}[X1_B]}{\mathbb{P}(B)}\,.$$

We can also define in a simple manner the conditional expectation with respect to a random variable Y, as long as Y takes finitely or countably many values $y_1,\, y_2,\, \ldots$ (each with positive probability). Then, for any integrable variable X, we let

$$\mathbb{E}[X|Y] \;=\; \sum_k \mathbb{E}[X|Y = Y_k]1_{\{Y=y_k\}} \;=\; \phi(Y)\,,$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is the function defined by $\phi(y) = \mathbb{E}[X|Y = y]$ if $\mathbb{P}(Y = y) > 0$, and $\phi(y) = 0$ otherwise.

Similarly, we can define the conditional expectation of an integrable variable X with respect to $\mathcal{G}$, a sub-σ-algebra of $\mathcal{F}$ which is generated by a finite or countable partition $(A_k)_k$ of Ω where $A_k \in \mathcal{F}$, by writing

$$\mathbb{E}[X|\mathcal{G}] \;=\; \sum_{k\,:\,\mathbb{P}(A_k)>0} \mathbb{E}[X|A_k]\,1_{A_k}\,.$$

If you think of a σ-algebra as a quantity of information (knowing $\mathcal{G}$ means knowing in which set of the partition you are), then the conditional expectation of X with respect to $\mathcal{G}$ is the best prediction of X you can make with the information at your disposal.

The above definitions can in fact be generalized to a more complete notion, which is the subject of this chapter.

Proposition 12. *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then for any $X \in L^1(\Omega,\, \mathcal{F},\, \mathbb{P})$, there exists a unique $Y \in L^1(\Omega,\, \mathcal{G},\, \mathbb{P})$ (which means that Y must be $\mathcal{G}$-measurable and integrable, and is unique up to equality almost everywhere) which satisfies one of the two following equivalent conditions:*

1. *For every event $A \in \mathcal{G}$, we have $\mathbb{E}[X1_A] = \mathbb{E}[Y1_A]$.*
2. *For every $Z \in L^\infty(\Omega,\, \mathcal{G},\, \mathbb{P})$ (i.e., for every $\mathcal{G}$-measurable and bounded variable), we have $\mathbb{E}[XZ] = E[YZ]$.*

Proof. **Existence:** Writing $X = X_+ - X_-$, it suffices to treat the case where X is nonnegative. We then check that the map

$$A \in \mathcal{G} \longmapsto \mathbb{E}\left[X 1_A\right]$$

is a measure on $(\Omega, \mathcal{G})$, which is absolutely continuous with respect to $\mathbb{P}$. By the Radon-Nikodym theorem, this measure admits a ($\mathcal{G}$-measurable) density Y, which by definition must satisfy

$$\forall A \in \mathcal{G}, \ \ \mathbb{E}\left[X 1_A\right] \ = \ \mathbb{E}\left[Y 1_A\right].$$

It follows from an approximation argument that this identity extends with 1_A replaced by arbitrary $\mathcal{G}$-measurable random variables.

Uniqueness: Let Y and Y' be two variables which satisfy the above. Since Y and Y' are $\mathcal{G}$-measurable, the event $A = \{Y > Y'\}$ is in $\mathcal{G}$, and it follows that

$$\mathbb{E}[Y 1_A] \ = \ \mathbb{E}[X 1_A] \ = \ \mathbb{E}[Y' 1_A],$$

which implies that $\mathbb{E}\left[(Y - Y') 1_{\{Y > Y'\}}\right] = 0$, so that almost surely $Y \leq Y'$. By symmetry the reverse inequality also holds almost surely. Therefore, $Y = Y'$ almost surely. □

The random variable Y obtained from the above proposition is denoted $\mathbb{E}[X|\mathcal{G}]$, and called the **conditional expectation of X with respect to $\mathcal{G}$** (or knowing $\mathcal{G}$). The conditional expectation can also be defined similarly in the case when the variable X is not integrable but $X \geq 0$.

Remark: the conditional expectation, when restricted to $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, coincides with the *orthogonal projection* on the closed subspace $L^2(\Omega, \mathcal{G}, \mathbb{P})$, with the scalar product $\langle X, Y \rangle = \mathbb{E}[XY]$. In particular, for $X \in L^2$, the conditional expectation $\mathbb{E}[X|\mathcal{G}]$ can be understood as the "best" approximation of X by a $\mathcal{G}$-measurable random variable.

We now detail some properties of the conditional expectation.

Proposition 13. *Assuming that the random variables X, Y are such that the conditional expectations below make sense, we have the following (in)equalities, understood in the almost sure sense.*

1. *(Linearity) If Y, Z are $\mathcal{G}$-measurable, then $\mathbb{E}[XY + Z|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]\ Y + Z$.*
2. *(Monotonicity) If $X \leq Y$ almost surely, then $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$.*
3. *$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.*
4. *(L^1-contraction) $\mathbb{E}\left[\ |\mathbb{E}\left[X|\mathcal{G}\right]|\right] \leq \mathbb{E}\left[|X|\right]$.*
5. *(Tower property) If $\mathcal{G}^1 \subset \mathcal{G}^2$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}^2]|\mathcal{G}^1]] = E[X|\mathcal{G}^1]$.*
6. *If X is independent from $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$.*
7. *If X is independent from $\mathcal{G}$, and Y is $\mathcal{G}$-measurable, then for any measurable f, it holds that $\mathbb{E}[f(X,Y)|\mathcal{G}] = g(Y)$ where $g(y) = \mathbb{E}[f(X,y)]$.*
8. *Conditional version of monotone convergence, Fatou's lemma, dominated convergence, Jensen's inequality,...*

Proof. Left as exercise. □

If Y is a random variable, we define the conditional expectation of X with respect to Y by

$$\mathbb{E}[X|Y] \ := \mathbb{E}[X|\sigma(Y)].$$

Remark: by Proposition 6, there exists a function h such that $\mathbb{E}[X|Y] \ = h(Y)$. To prove this identity, by definition it suffices to check that for any measurable (bounded) g, it holds that

$$\mathbb{E}\left[X g(Y)\right] = \mathbb{E}\left[h(Y) g(Y)\right].$$

Exercise 3. *Check that this coincides with the definition given in the beginning of the subsection if Y is discretely valued.*

When the considered random variables have densities, conditional expectations can be computed explicitly.

Proposition 14. *Assume that (X, Y) has a law which admits a density $p : \mathbb{R}^2 \to \mathbb{R}$ with respect to the Lebesgue measure on $\mathbb{R}$. Then, for any function $h : \mathbb{R} \to \mathbb{R}$ such that $h(X)$ is integrable, we have*

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

where $g : \mathbb{R} \to \mathbb{R}$ is the function given by

$$g(y) \;=\; \frac{\int_{\mathbb{R}} h(x)p(x,\, y)\, dx}{\int_{\mathbb{R}} p(x,\, y)\, dx}$$

if the denominator (which is the density of the variable Y) is positive, and $g(y) = 0$ otherwise.

6 Martingales in discrete time

Let $(\Omega,\, \mathcal{F},\, \mathbb{P})$ be a probability space.

Definition 17. *A* discrete real stochastic process *is a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ defined on a same probability space $(\Omega,\, \mathcal{F},\, \mathbb{P})$ and taking values in $(\mathbb{R},\, \mathcal{B}(\mathbb{R}))$.*

Definition 18. *A* filtration *of $(\Omega,\, \mathcal{F},\, \mathbb{P})$ is a non-decreasing sequence $(\mathcal{F}_n)_{n\in\mathbb{N}}$ of sub-σ-algebras of $\mathcal{F}$, that is to say, $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for every $n \in \mathbb{N}$. We say that $(\Omega,\, \mathcal{F},\, (\mathcal{F}_n)_{n\in\mathbb{N}},\, \mathbb{P})$ is a* filtered probability space.

In a filtered probability space, one can think of n as time and $\mathcal{F}_n$ represents the information known at time n.

Definition 19. *A process $(X_n)_{n\in\mathbb{N}}$ is said to be* adapted *to the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ if, for all $n \in \mathbb{N}$, the variable X_n is $\mathcal{F}_n$-measurable.*

Definition 20. *If $(X_n)_{n\in\mathbb{N}}$ is a process, we define the* natural filtration*, or* canonical filtration *of* filtration generated by $(X_n)_{n\in\mathbb{N}}$ *as $(\mathcal{F}_n)_{n\in\mathbb{N}}$ where, for every $n \in \mathbb{N}$, $\mathcal{F}_n = \sigma(X_0,\, \dots,\, X_n)$. It is the smallest filtration which makes the process $(X_n)_{n\in\mathbb{N}}$ adapted.*

Definition 21. *A process $(X_n)_{n\in\mathbb{N}}$ is called a* martingale *(respectively, a* submartingale*, a* supermartingale*) with respect to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ if:*

- *The process $(X_n)_{n\in\mathbb{N}}$ is adapted to $(\mathcal{F}_n)_{n\in\mathbb{N}}$, i.e., for every $n \in \mathbb{N}$, X_n is $\mathcal{F}_n$-measurable.*
- *For every $n \in \mathbb{N}$, the variable X_n is integrable.*
- *For every $n \in \mathbb{N}$, $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$, respectively $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$, $\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n$.*

Example 3. *Random walk: if $(X_n)_{n\geq 1}$ are i.i.d. integrable and centered variables, then the process $(S_n)_{n\in\mathbb{N}}$ defined by $S_0 = 0$ and $S_n = X_1 + \cdots + X_n$ for every $n \geq 1$ is a martingale.*

Closed martingale: if $Z \in L^1(\mathcal{F})$, and $(\mathcal{F}_n)_{n\in\mathbb{N}}$ is a filtration, then the process $(X_n)_{n\in\mathbb{N}}$ defined by $X_n = \mathbb{E}[Z|\mathcal{F}_n]$ for every $n \in \mathbb{N}$ is a martingale. It is called a closed martingale, or Doob's martingale.

Definition 22. *A* stopping time *with respect to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ is a random variable $T : \Omega \to \mathbb{N} \cup \{+\infty\}$ such that*

$$\forall n \in \mathbb{N}\,, \qquad \{T \leq n\} \,\in\, \mathcal{F}_n\,,$$

Exercise: check that this equivalent to the same definition with $\leq$ replaced by $=$.

Example 4. *If $(X_n)_{n\in\mathbb{N}}$ is an adapted process and A is a Borel set of $\mathbb{R}$ then the first hitting time of A, namely*

$$T_A \;=\; \inf\big\{n \in \mathbb{N} \,:\, X_n \in A\big\}\,,$$

with the usual convention that $\inf \varnothing = +\infty$*, is a stopping time.*

If T is a stopping time which is almost surely finite, then we can define the variable

$$X_T \;=\; \sum_{n\in\mathbb{N}} X_n 1_{\{T=n\}}\,.$$

Definition 23. *Given a stopping time T relative to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, we define the* ***$\sigma$-field of the past until*** *T as*

$$\mathcal{F}_T = \big\{A \in \mathcal{F} \,:\, \forall n \in \mathbb{N},\; A \cap \{T = n\} \in \mathcal{F}_n\big\}\,.$$

Exercise 4. *Check that $\mathcal{F}_T$ is a σ-algebra, that T is $\mathcal{F}_T$-measurable, that if $T < \infty$ almost surely, then X_T is $\mathcal{F}_T$-measurable., and that if S and T are two stopping times such that $S \le T$, then $\mathcal{F}_S \subset \mathcal{F}_T$.*

Theorem 13 (Doob's optional stopping (or sampling) theorem)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a martingale, respectively a submartingale, a supermartingale, with respect to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, and let T be a stopping time for this filtration.*

- *Then, the stopped process $(X_{n\wedge T})_{n\in\mathbb{N}}$ is a martingale (respectively, a submartingale, a supermartingale).*
- *Moreover, if $T < \infty$ almost surely and one of the following holds:*
  - *T is almost surely bounded, i.e., there exists a constant $C < \infty$ such that $T \le C$ almost surely;*
  - *$\mathbb{E}[T] < \infty$ and $(M_{t+1} - M_t)_{t\in\mathbb{N}}$ is bounded in L^∞;*
  - *$(M_{t\wedge T})_{t\in\mathbb{N}}$ is bounded in L^∞,*

  then $X_T \in L^1(\mathcal{F}_T)$ and $\mathbb{E}[X_T] = \mathbb{E}[X_0]$, respectively, $\mathbb{E}[X_T] \ge \mathbb{E}[X_0]$, $\mathbb{E}[X_T] \le \mathbb{E}[X_0]$.

Exercise 5. *Find a martingale M and a stopping time $T < \infty$ almost surely, and such that $\mathbb{E}[M_T] \neq \mathbb{E}[M_0]$.*

We then have the following generalization:

Proposition 15. *Let $S \le T$ be two bounded stopping times, and $(M_n)_{n\in\mathbb{N}}$ a martingale (all relative to the same filtration). Then*

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_S.$$

An important part of martingale theory is their convergence properties when $n \to \infty$. We state the below theorems without proofs.

Theorem 14. *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale which is bounded in L^1 (i.e., $\sup_n \mathbb{E}|M_n| < \infty$). Then M_n converges almost surely to a limit $M_\infty \in L^1$.*

Note that, although the limit is in L^1, the convergence does not hold in L^1 in general. Indeed, let $M_0 = 1$ and, for every $n \ge 1$, let $M_n = U_1 \dots U_n$, where the U_n are i.i.d. with $\mathbb{P}(U_1 = 0) = \mathbb{P}(U_1 = 2) = 1/2$. Then $M_n \to 0$ almost surely, but $\mathbb{E}[M_n] = 1$ for all n.

There exist the following variants of the above result:

Theorem 15. *Let $(M_n)_{n\in\mathbb{N}}$ be a submartingale such that $(M_n^+)_{n\in\mathbb{N}}$ is bounded in L^1, where $M_n^+ = \max(M_n, 0)$. Then M_n converges almost surely to a limit $M_\infty \in L^1$.*

Theorem 16. *Any non-negative supermartingale converges almost surely to a limiting integrable variable.*

In the case of a martingale bounded in L^p for $p > 1$, we can obtain a stronger convergence:

Theorem 17. *Let $(M_n)_{n\geq 0}$ be a martingale which is bounded in L^p, for a certain $1 < p < \infty$. Then M_n converges almost surely and in L^p to a limit M_∞.*

Proof of the case $p = 2$. In the case $p = 2$, recalling that conditional expectations are orthogonal projections, for every $0 \leq n \leq m$, we have

$$\| M_m\|_2^2 \;=\; \|M_m - M_n\|_2^2 + \|M_n\|_2^2 \,,$$

from which it follows that $\|M_n\|_2$ is an increasing sequence, which, since it is bounded, must converge to a finite limit. It then also follows that $(M_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in L^2, from which we can conclude. □

In order to state conditions under which the convergence holds in L^1, we need the following (important) notion.

Definition 24. *A family $(X_i)_{i\in I}$ of real random variables is **uniformly integrable** if*

$$\lim_{a\to\infty} \sup_{i\in I} \mathbb{E}\big[\,|X_i|\,1_{\{|X_i|\geq a\}}\big] \;=\; 0$$

or, equivalently, if $(X_i)_{i\in I}$ is bounded in L^1 and for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $A \in \mathcal{F}$, if $\mathbb{P}(A) \leq \delta$ then for every $i \in I$ we have $\mathbb{E}[|X_i|1_A] \leq \varepsilon$.

Exercise 6. *Check the equivalence between the two definitions. Note that uniform integrability implies L^1 boundedness. Find a counterexample showing that the converse is not true in general.*

Example 5.
1. *A finite collection of integrable variables is always uniformly integrable.*
2. *Domination by an integrable variable implies uniform integrability, that is to say, if $(X_i)_{i\in I}$ and $Z \in L^1(\mathcal{F})$ are such that, for all $i \in I$, we have $|X_i| \leq Z$, then $(X_i)_{i\in I}$ is uniformly integrable.*
3. *If a collection is bounded in L^p for a given $p > 1$, then it is uniformly integrable.*
4. *More generally, if there exists a function $\phi : \mathbb{R} \to \mathbb{R}$ with $\phi(x)/|x| \to +\infty$ when $|x| \to \infty$ and $\sup_i \mathbb{E}[\phi(X_i)] < \infty$, then the family $(X_i)_{i\in I}$ is uniformly integrable. This is called De La Vallée Poussin's criterion.*
5. *Let $X \in L^1(\mathcal{F})$, and let $(\mathcal{G}_i)_{i\in I}$ be a collection of sub-σ-fields of $\mathcal{F}$. Then, the family $(X_i)_{i\in I}$ defined by $X_i = \mathbb{E}[X|\mathcal{G}_i]$ is uniformly integrable.*

The following theorem motivates the introduction of the concept of uniform integrability. Indeed, it gives a necessary and sufficient condition to obtain equivalence between convergence in probability and convergence in L^1.

Theorem 18. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of integrable random variables. Then, the two following properties are equivalent:*

1. *The sequence $(X_n)_{n\in\mathbb{N}}$ converges in L^1 towards a variable $X \in L^1(\mathcal{F})$.*
2. *The sequence $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable and converges in probability towards a variable X.*

This theorem is a particular case of Lebesgue-Vitali's convergence theorem, which deals with the more general case of L^p, for $p \geq 1$.

This theorem can also be seen as a stronger version of the dominated convergence theorem, which only gives a sufficient condition for convergence in L^1: if $X_n \to X$ a.s. (and thus also in probability) and if there exists $Z \in L^1(\mathcal{F})$ such that, for all $n \in \mathbb{N}$, $|X_n| \leq Z$, then $(X_n)_{n\in\mathbb{N}}$ converges to X in L^1. The hypothesis of uniform integrability is weaker than domination by an integrable variable (i.e., the latter implies the former, as explained in point 1 of Example 5). Thus, the above theorem says that uniform integrability is the optimal hypothesis that one should require in order to obtain convergence in L^1.

To prove the direct implication, one only has to check that the definition of uniform integrability, using that a finite family is uniformly integrable. The reciprocal can be obtained by showing that the sequence $(X_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^1(\Omega, \mathcal{F}, \mathbb{P})$.

This important result has the following consequence for martingales:

Theorem 19 (Convergence in L^1)**.** *Let $(X_n)_{n\in\mathbb{N}}$ be a martingale. Then, the three following statements are equivalent:*

1. *The sequence $(X_n)_{n\in\mathbb{N}}$ converges almost surely and in L^1 to a random variable $X_\infty \in L^1$.*
2. *The family $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable.*
3. *The martingale $(X_n)_{n\in\mathbb{N}}$ is closed, that is to say, there exists a random variable $Z \in L^1$ such that, for all $n \in \mathbb{N}$, $X_n = \mathbb{E}[Z|\mathcal{F}_n]$.*

If condition 1 holds, then we can take $Z = X_\infty$ in 3.

This theorem leads to the following extension of Doob's stopping theorem:

Theorem 20 (Stopping theorem, second version)**.** *Let $(M_n)_{n\in\mathbb{N}}$ be a martingale with respect to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, which is uniformly integrable and thus converges almost surely and in L^1 towards a random variable $M_\infty \in L^1$. Let S and T be two stopping times relative to $(\mathcal{F}_n)_{n\in\mathbb{N}}$, taking values in $\mathbb{N}\cup\{\infty\}$, such that $S \le T$. Then we have*

$$\mathbb{E}\big[X_T \,\big|\, \mathcal{F}_S\big] \;=\; X_S\,,$$

where X_T is defined as

$$X_T \;=\; X_\infty\, 1_{\{T=\infty\}} + \sum_{n=0}^{\infty} X_n\, 1_{\{T=n\}}$$

and X_S is defined similarly.

7 Gaussian vectors

In this section, we deal with random vectors, i.e., random variables taking values in $\mathbb{R}^d$.

Note that many results stated above in the scalar case remain true in higher dimension. For instance, given a random vector $X = (X_1, \ldots, X_d)$, its characteristic function is defined by

$$\phi_X \,:\, \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d \;\longmapsto\; \mathbb{E}\big[\;\exp(i\langle \xi,\, X\rangle)\big]$$

where $\langle \cdot, \cdot\rangle$ is the usual scalar product ($\xi \cdot X =$
$xi_1 X_1 + \cdots + \xi_d X_d$).

Then it still holds that the law of a random vector is characterized by its characteristic function, and that a sequence $(X_n)_{n\in\mathbb{N}}$ converges in distribution to a variable X if and only if its characteristic functions ϕ_{X_n} converge pointwise to ϕ_X.

Definition 25. *A scalar random variable X is a **Gaussian** if there exists $m \in \mathbb{R}$ and $\sigma \ge 0$ such that X has the same law as $m + \sigma Z$, where $Z \sim \mathcal{N}(0,1)$ or, equivalently (if $\sigma \neq 0$), if the law of X has the density*

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

with respect to the Lebesgue measure on $\mathbb{R}$. We write $X \sim \mathcal{N}(m, \sigma^2)$.

Proposition 16. *If $X \sim \mathcal{N}(m,\, \sigma^2)$ then for every $t \in \mathbb{R}$ we have $\phi_X(t) = \exp(imt - t^2\sigma^2/2)$.*

Definition 26. *A random vector $X = (X^1, \ldots, X^d)$ is **a Gaussian vector** (also written: $(X^1, \ldots, X^d)$ are **jointly Gaussian**) if, for each $(\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d$, the scalar random variable $\lambda_1 X_1 + \cdots + \lambda_d X_d$ is Gaussian.*

Remark: to show that $(X_1, \ldots, X_d)$ is a Gaussian vector, it is not enough to show that each of its components is Gaussian. Exercise: find a counterexample such that each component is Gaussian but $(X_1, \ldots, X_d)$ is not a Gaussian vector.

Definition 27. *Given a Gaussian vector X, we define its mean $m = (m_1, \dots, m_d) \in \mathbb{R}^d$ and its covariance matrix $\Sigma = (\Sigma_{i,j})_{1 \le i,j \le d} \in \mathbb{R}^{d \times d}$ by*

$$m_i \;=\; \mathbb{E}[X_i] \qquad \textit{and} \qquad \Sigma_{i,j} \;=\; \mathrm{Cov}(X_i, X_j) = \mathbb{E}\big[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\big]\,.$$

Proposition 17. *The law of a Gaussian vector is characterized by its mean m and covariance matrix Σ. More precisely, the characteristic function of X is given for $\xi \in \mathbb{R}^d$ by*

$$\phi_X(\xi) \;=\; \exp\left(i\langle m, \xi\rangle - \frac{\langle \xi, \Sigma\xi\rangle}{2}\right).$$

We then write $X \sim \mathcal{N}(m, \Sigma)$.

Proof. By assumption, if $\xi \in \mathbb{R}^d$, the variable $\langle \xi, X\rangle$ is Gaussian, and we can compute

$$\mathbb{E}[\langle \xi, X\rangle] \;=\; \langle m, \xi\rangle \qquad \text{and} \qquad \mathrm{Var}(\langle \xi, X\rangle) \;=\; \sum_{i,j} \xi_i \xi_j \Sigma_{i,j} \;=\; \langle \xi, \Sigma\xi\rangle\,.$$

The formula then follows from that for the scalar Gaussian variables. Since the characteristic function can be expressed as a function of the mean and the covariance matrix, it shows that these two objects characterize the law of the Gaussian vector. □

This result has the following interesting consequences:

Corollary 1. *Let $(X^1, \dots, X^d)$ be jointly Gaussian. Then $X^1, \dots, X^d$ are independent if and only if they are pairwise uncorrelated.*

Corollary 2. *If $X \sim \mathcal{N}(m, \Sigma)$ and $A \in \mathbb{R}^{d\times d}$ then $AX \sim \mathcal{N}(Am, A\Sigma({}^tA))$.*

Proof. By the above proposition, it suffices to show that both members have the same law and covariance matrix. □

Remark: this gives a way to simulate any Gaussian vector with given mean m and covariance matrix Σ: it is enough to find a matrix A such that $\Sigma = A({}^tA)$, and to let $X = m + AY$, where $Y \sim \mathcal{N}(0, 1)$. There always exists such a matrix A, which can even be taken triangular: this is the so-called Cholesky decomposition of symmetric matrices.

Let us present one more corollary of Proposition 17:

Corollary 3. *If $(X_n)_{n\in\mathbb{N}}$ is a sequence of Gaussian vectors with mean m_n and covariance matrices Σ_n such that $m_n \to m$ and $\Sigma_n \to \Sigma$, then X_n converges in distribution fo $\mathcal{N}(m, \Sigma)$.*

Proof. It suffices to check pointwise convergence of the characteristic functions. □

We now show how, for Gaussian vectors, conditional expectations are easy to compute.

Proposition 18. *Let (X, Y) be a Gaussian vector in $\mathbb{R}^2$, with mean m and covariance matrix Σ given by*

$$m \;=\; \begin{pmatrix} m_X \\ m_Y \end{pmatrix} \qquad \textit{and} \qquad \Sigma \;=\; \begin{pmatrix} \sigma_{XX} & \sigma_{XY} \\ \sigma_{XY} & \sigma_{YY} \end{pmatrix}$$

and assume that $\sigma_{YY} > 0$. Then we have

$$\mathbb{E}\,[X|Y] \;=\; m_X + \frac{\sigma_{XY}}{\sigma_{YY}}\,(Y - m_Y)\,.$$

Proof. Let us define the variable

$$W \;=\; X - \frac{\sigma_{XY}}{\sigma_{YY}} Y\,.$$

Then a direct computation gives that $\mathrm{Cov}(W, Y) = 0$, and since (W, Y) is Gaussian, this implies that W is independent of Y. The result then follows from writing $X = W + (\sigma_{XY}/\sigma_{YY})Y$, with W independent of Y and the second term is $\sigma(Y)$-measurable. □

Note that the above computation extends to vectors of higher dimensions.

Finally, we remark that Gaussian vectors also arise naturally in fluctuation of i.i.d. random vectors.

Theorem 21 (Central Limit Theorem in $\mathbb{R}^d$)**.** *Let $(X^n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of random vectors, with square integrable entries. Let $m = \mathbb{E}[X^1]$, and $\Sigma = \big(\mathrm{Cov}(X^1_i, X^1_j)\big)_{1\le i,j\le d}$. Then we have the convergence in law*

$$\frac{X^1 + \cdots + X^n - nm}{\sqrt{n}} \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0, \Sigma)\,.$$

Proof. We treat the case $m = 0$. For every $n \ge 1$, we write $S^n = X^1 + \cdots + X^n$. For every fixed $\xi \in \mathbb{R}^d$, the scalar Central Limit Theorem applied to the sequence of real variables $(\langle \xi, X^n\rangle)_{n\in\mathbb{N}}$ shows that $\langle \xi, S^n\rangle/\sqrt{n}$ converges in distribution to $\mathcal{N}\big(0, \langle u, \Sigma u\rangle\big)$, which implies that $\phi_{S^n/\sqrt{n}}(\xi) \to \exp\big(-\langle u, \Sigma u\rangle/2\big)$. □

8 Brownian motion

8.1 Definition and first properties

Definition 28. *A* ***continuous-time stochastic process*** *is a family $X = (X_t)_{t\in\mathbb{R}_+}$ of random variables indexed by $\mathbb{R}_+$, defined on a same probability space.*

Definition 29. *A (standard)* ***Brownian motion*** *is a continuous-time stochastic process $(B_t)_{t\ge 0}$ such that*

1. *$B_0 = 0$ almost surely;*
2. *For each $k \ge 2$ and $0 < t_1 < \cdots < t_k$, the increments $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}}$ are independent;*
3. *For each $0 \le s \le t$, the increment $B_t - B_s$ has law $\mathcal{N}(0, t-s)$;*
4. *For every $\omega \in \Omega$, the function $t \mapsto B_t(\omega)$ is continuous on $\mathbb{R}_+$.*

Remark 1. *(Technical remark on continuity of sample paths)*

- *Condition 4 is sometimes replaced by the weaker condition:*

  (4') There exists a measurable $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$, such that for every $\omega \in \Omega_0$, the function $t \mapsto B_t(\omega)$ is continuous on $\mathbb{R}_+$.

  Note that if B satisfies 1, 2, 3 and 4', we can redefine B_t to be constant equal to 0 outside of Ω_0, to obtain a stochastic process satisfying 4, which is almost surely equal to the original process.

- *In fact, even though 1, 2 and 3 do not imply 4 or 4', they imply that we can find a modification of B (that is to say, a process B' such that for all $t \ge 0$, almost surely, $B_t = B'_t$) which is continuous (exercise: use for instance a similar construction to what is done below to find a sequence of continuous processes B^n, which almost surely converge uniformly on compacts, and such that for all $t \ge 0$, almost surely, $B^n_t \to B_t$.)*

Definition 30. *A continuous-time stochastic process $(X_t)_{t\ge 0}$ is a* ***Gaussian process*** *if for every $k \ge 1$ and any $t_1, \ldots, t_n$, $(X_{t_1}, \ldots, X_{t_k})$ is a Gaussian vector.*

Proposition 19. *Let $(B_t)_{t\geq 0}$ be a continuous-time stochastic process. Then $(B_t)_{t\geq 0}$ is a Brownian motion if and only if it is a Gaussian process with mean function given by $\mathbb{E}[B_t] = 0$ for all $t \geq 0$ and covariance function given by $\mathbb{E}[B_s B_t] = \min(s, t)$ for all $s, t \geq 0$ and such that for all $\omega \in \Omega$, the function $t \mapsto B_t(\omega)$ is continuous on $\mathbb{R}_+$.*

Proof. Assume first that $(B_t)_{t\geq 0}$ is a Brownian motion. Let $k \geq 1$ and let $0 \leq t_1 < \cdots < t_k$. Then the vector $(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}})$ has independent components each with a Gaussian distribution, so it is a Gaussian vector. Thus, $(B_{t_1}, \ldots, B_{t_k})$ is a Gaussian vector. For every $t \geq 0$, we have $\mathbb{E}[B_t] = \mathbb{E}[B_t - B_0] = 0$ because $B_t - B_0 \sim \mathcal{N}(0, t)$. And if $0 \leq s \leq t$, we have

$$\mathbb{E}[B_s B_t] \; = \; \mathbb{E}\big[B_s(B_s + B_t - B_s)\big] \; = \; \mathbb{E}\big[B_s^2\big] + \mathbb{E}\big[B_s(B_t - B_s)\big] \; = \; s + 0 = \min(s, t)\,.$$

using that $B_s \sim \mathcal{N}(0, s)$ and that it is independent of $B_t - B_s$.

For the reciprocal, assume that B is a Gaussian process with the above mean function and covariance function, and with continuous paths. First, $B_0 = 0$ almost surely because $\mathbb{E}[B_0^2] = 0$. Then, if $k \geq 2$ and $0 < t_1 < \cdots < t_k$ then $(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_k} - B_{t_{k-1}})$ is a Gaussian vector and we can check by computation that its components are uncorrelated. Thus, its components are independent. Lastly, if $0 \leq s \leq t$, we know that the increment $B_t - B_s$ is Gaussian and it suffices to check that it has mean 0 and variance $t - s$. □

Proposition 20. *Let $B = (B_t)_{t\geq 0}$ be a Brownian motion. Then:*

1. *(Symmetry) $(-B_t)_{t\geq 0}$ is a Brownian motion.*
2. *(Scaling) For each $\lambda > 0$, $B_t^\lambda = B_{\lambda t}/\sqrt{\lambda}$ is a Brownian motion.*
3. *(Weak Markov property) For each constant $T > 0$, $B_t^T := B_{T+t} - B_T$ is a Brownian motion (independent from $\sigma(B_s, s \leq T)$).*
4. *(Time inversion) $(tB_{1/t}1_{\{t>0\}})_{t\geq 0}$ is a Brownian motion.*

Proof. Left as an exercise. The most delicate point is to check that for the last point, the process is continuous at 0, which is equivalent to

$$\lim_{t\to\infty} \frac{B_t}{t} \; = \; 0\,. \tag{2}$$

Note that it follows from the Strong Law of Large Numbers that $B_n/n \to 0$, when n is an integer which tends to ∞. Exercise: prove that(2), holds almost surely, taking for granted that

$$\mathbb{E}\Big[\sup_{t\in[0,1]} |B_t|\Big] \; < \; \infty \tag{3}$$

and using independence of increments. We show (3) later. □

The fact that Brownian motions exist is not an obvious fact. We present later a rigorous construction of Brownian motion. Before this, let us introduce the law of the Brownian motion:

Definition 31. *Let $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ be the set of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}$, and let $\mathcal{E}$ be the Borel σ-algebra on this set associated with the topology of uniform convergence on compacts. Ket $B = (B_t)_{t\geq 0}$ be a Brownian motion, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We can see B as a map $B : \Omega \to \mathcal{C}(\mathbb{R}_+, \mathbb{R})$. The Wiener measure P_0 is the law of B on $(\mathcal{C}(\mathbb{R}_+, \mathbb{R}), \mathcal{E})$, i.e., the image measure of $\mathbb{P}$ by B: for all $A \in \mathcal{E}$, we let $P_0(A) = \mathbb{P}(B^{-1}(A)) = \mathbb{P}\big((t \mapsto B_t) \in A\big)$.*

Yet, to ensure that this definition makes sense, one needs to check the following fact:

Proposition 21. *A Brownian motion B defines a measurable map from $(\Omega, \mathcal{F})$ to $(\mathcal{C}(\mathbb{R}_+, \mathbb{R}), \mathcal{E})$.*

Proof. To show this, it is enough to check that for every open set $O \subset \mathcal{C}(\mathbb{R}_+, \mathbb{R})$, we have $B^{-1}(O) \in \mathcal{F}$. Yet, the topology of uniform convergence on compacts derives from the distance

$$d : (f, g) \longmapsto \sum_{n=1}^{\infty} \frac{1}{2^n} \min\Big(1, \sup_{[0, n]} |f - g|\Big) .$$

The space $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ with this distance is separable (i.e., there exists a countable dense family), which implies that each open set is a countable union of open balls. So it is enough to show that for every $f \in \mathcal{C}(\mathbb{R}_+, \mathbb{R})$ and every $\varepsilon > 0$, we have $\{\omega \in \Omega : d(B^\omega, f) < \varepsilon\} \in \mathcal{F}$. To show this, we may write

$$\big\{\omega \in \Omega : d(B^\omega, f) < \varepsilon\big\} \;=\; \Big\{\omega \in \Omega : \sum_{n=0}^{\infty} \frac{1}{2^n} \min\Big(1, \sup_{[0, n]\cap\mathbb{Q}} |B - f|\Big) < \varepsilon\Big\},$$

which belongs to $\mathcal{F}$ as the preimage of the measurable set $[0, \varepsilon)$ through a measurable function (the big sum). Note that here it is important that we reduced the supremum to the rational points, to that it is the supremum of countably many quantities, which makes the above function measurable. □

Remark: this Wiener measure P_0 does not depend on the choice of the Brownian motion (the Brownian motion in itself depends on the choice of the set Ω and on how it is constructed): all (standard) Brownian motion have the same law.

8.2 Construction of Brownian motion

We now present a rigorous construction of Brownian motion. Note that a Brownian motion can be viewed as a function of two variables:

$$B : \begin{cases} \Omega \times \mathbb{R}_+ \longrightarrow \mathbb{R} \\ \quad (\omega, t) \longmapsto B_t^\omega , \end{cases}$$

Hence, for every fixed $\omega \in \Omega$, we have a function $B^\omega : \mathbb{R}_+ \to \mathbb{R}$: this is why we can say that Brownian motion is a random function ("random" means that it depends on ω). Besides, for every fixed $t \in \mathbb{R}_+$, the map $B_t : \omega \in \Omega \mapsto B_t^\omega \in \mathbb{R}$ is a random variable, that is to say, a measurable map from Ω to $\mathbb{R}$. In the following proof, we often switch between these two points of view on Brownian motion, sometimes seen as a random function and sometimes as a family of random variables indexed by $t \geq 0$. Hence, in the notation B_t^ω, when we omit the ω it means that we consider the random variable B_t, and when we omit the t it means that we consider the function B^ω. Sometimes we write $B(t)$ instead of B_t, or $B^\omega(t)$ instead of B_t^ω.

We construct the process only on the interval $[0, 1]$, and then Brownian motion on $\mathbb{R}_+$ can be obtained by pasting together independent copies.

We consider $E = [0, 1]$, equipped with the Borel σ-algebra $\mathcal{E} = \mathcal{B}([0, 1])$ and the Lebesgue measure λ. Then, $L^2(E) = L^2(E, \mathcal{E}, \lambda)$, which is the set of square integrable functions from E to $\mathbb{R}$, is a Hilbert space, when equipped with the scalar product $\langle f, g\rangle_{L^2(E)} = \int_{[0,1]} fg \, d\lambda$.

The idea is that we will construct an isometry $\Phi : L^2(E) \to L^2(\Omega)$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is some probability space, that is to say, to each function it associates a scalar random variable, and it preserves the scalar product, i.e., for every $f, g \in L^2(E)$,

$$\langle \Phi(f), \Phi(g)\rangle_{L^2(\Omega)} \;=\; \langle f, g\rangle_{L^2(E)} ,$$

where the scalar product in $L^2(\Omega)$ is the usual scalar product $\langle X, Y\rangle_{L^2(\Omega)} = \mathbb{E}[XY]$. This map Φ will also be such that, for every $f \in L^2(E)$, the variable $\Phi(f)$ is a centered Gaussian variable with variance $\|f\|_{L^2}^2$.

Then, we will construct a Brownian motion $(B_t)_{t\in[0,1]}$ such that, for every $t \in [0, 1]$, $B_t = \Phi(1_{[0,t]})$. Note that this does not really define B because the right-hand side is only an element of $L^2(\Omega)$, namely, it is an equivalence class of random variables which are almost surely equal. Therefore, this only defines B_t up to almost sure equality, but not pointwise in ω. We will later define precisely B_t^ω for every $\omega \in \Omega$ and every $t \in [0, 1]$.

To get the intuition behind this construction, let us first check that if Φ and B are constructed in this way, then the process B is indeed a Gaussian process with the correct mean and covariance functions. For every $k \geq 1$, if $t_1, \ldots, t_k \in [0,1]$ and $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ then we have (note that Φ being an isometry implies that it is linear)

$$\sum_{j=1}^{k} \lambda_j B_{t_j} \;=\; \sum_{j=1}^{k} \lambda_j \Phi(1_{[0,t_j]}) \;=\; \Phi\Big(\sum_{j=1}^{k} \lambda_j 1_{[0,t_j]}\Big),$$

which is a Gaussian variable because as explained before the function Φ maps every function $f \in L^2(E)$ onto a Gaussian variable. Therefore, B is a Gaussian process on $[0,1]$.

Then, for every $t \in [0,1]$, we have $\mathbb{E}[B(t)] = \mathbb{E}\big[\Phi(1_{[0,t]})\big] = 0$ because $\Phi(1_{[0,t]})$ is a centered Gaussian variable. And if $0 \leq s \leq t \leq 1$, using that Φ is an isometry we can write

$$\begin{aligned}\mathbb{E}\big[B(s)B(t)\big] \;&=\; \mathbb{E}\big[\Phi(1_{[0,s]})\Phi(1_{[0,t]})\big] \;=\; \langle \Phi(1_{[0,s]}),\, \Phi(1_{[0,t]})\rangle_{L^2(\Omega)} \\ &=\; \langle 1_{[0,s]},\, 1_{[0,t]}\rangle_{L^2(E)} \;=\; \int_{[0,1]} 1_{[0,s]}(x) 1_{[0,t]}(x)\, dx \;=\; s \;=\; \min(s,\, t)\,.\end{aligned}$$

Therefore, B is a Gaussian process with the correct mean and covariance function, and there will remain to check later that its trajectories are continuous.

Let us now construct this map Φ. To define it, we consider a Hilbertian basis of the space $L^2(E)$. Namely, we consider the family of functions $(h_0,\, h_{n,k})_{n\in\mathbb{N},\, 0\leq k<2^n}$ defined by $h_0 = 1_{[0,1]}$ and, for $n \in \mathbb{N}$ and $0 \leq k \leq 2^n - 1$,

$$h_{n,k} \;=\; 2^{n/2}\Big(1_{\left[\frac{2k}{2^{n+1}},\, \frac{2k+1}{2^{n+1}}\right)} - 1_{\left[\frac{2k+1}{2^{n+1}},\, \frac{2k+2}{2^{n+1}}\right)}\Big)\,.$$

This family forms a Hilbertian orthonormal basis of $L^2(E)$ (the factor $2^{n/2}$ above is a normalization factor ensuring that $\|h_{n,k}\|_{L^2(E)} = 1$), so that for every $f \in L^2(E)$ we have the convergence, in the L^2 sense,

$$\langle f,\, h_0\rangle h_0 + \sum_{n=0}^{N} \sum_{k=0}^{2^n-1} \langle f,\, h_{n,k}\rangle h_{n,k} \;\overset{N\to\infty}{\longrightarrow}\; f\,. \tag{4}$$

Then, we consider a family $(Z_0,\, Z_{n,k})_{n\in\mathbb{N},\, 0\leq k<2^n}$ of i.i.d. standard Gaussian variables, defined on some probability space $(\Omega,\, \mathcal{F},\, \mathcal{P})$. This sequence is orthonormal in $L^2(\Omega,\, \mathcal{F},\, \mathbb{P})$, which is the set of square integrable scalar random variables defined on Ω, equipped with the usual scalar product $\langle X,\, Y\rangle_{L^2(\Omega)} = \mathbb{E}[XY]$, as explained above. Therefore, there exists an isometry $\Phi : L^2(E,\, \mathcal{E},\, \lambda) \to L^2(\Omega,\, \mathcal{F},\, \mathbb{P})$ which maps the orthonormal basis onto the orthonormal family, that is to say, such that $\Phi(h_0) = Z_0$ and for every $n \in \mathbb{N}$ and every $0 \leq k \leq 2^n - 1$, we have $\Phi(h_{n,k}) = Z_{n,k}$. Applying this isometry Φ on both sides of (4), we get that, for every $f \in L^2(E)$,

$$\Phi(f) \;=\; \lim_{N\to\infty} \langle f,\, h_0\rangle Z_0 + \sum_{n=0}^{N} \sum_{k=0}^{2^n-1} \langle f,\, h_{n,k}\rangle Z_{n,k}\,, \tag{5}$$

where the limit is in L^2. Note that this can also be taken as a definition of the map Φ.

At this point, we know that for every $f \in L^2(E)$, the variable $\Phi(f)$ is Gaussian, because it is a limit in L^2 of Gaussian variables, and we have the following:

Lemma 3. *If a sequence of Gaussian variables converges in L^2, then the limit is also Gaussian.*

Proof. Left as an exercise. Note that L^2 convergence implies convergence of the means and of the variances, and then use the characteristic function. □

Moreover, for every $f \in L^2(E)$, the variable $\Phi(f)$ is centered because it is a limit in L^2 of centered variables, and has variance $\mathrm{Var}(\Phi(f)) = \|\Phi(f)\|^2_{L^2(\Omega)} = \|f\|^2_{L^2(E)}$ because Φ is an isometry. Therefore, we checked that $\Phi(f)$ has distribution $\mathcal{N}\big(0,\, \|f\|^2_{L^2(E)}\big)$.

Let us now consider

$$g_0 \,:\, \begin{cases} [0,\,1] \to \mathbb{R} \\ \qquad t \mapsto \langle 1_{[0,\,t]},\, h_0 \rangle \;=\; t \end{cases}$$

and, for $n \in \mathbb{N}$ and $0 \leq k \leq 2^n - 1$,

$$g_{n,k} \,:\, \begin{cases} [0,\,1] \to \mathbb{R} \\ \qquad t \mapsto \langle 1_{[0,\,t]},\, h_{n,k} \rangle \;=\; \displaystyle\int_0^t h_{n,k}(s)\, ds\,. \end{cases}$$

Exercise: draw the graph of the first functions of the family. With this notation, for every $t \in [0,1]$, Equation (5) applied to the function $f = 1_{[0,\,t]}$ writes

$$\Phi(1_{[0,\,t]}) \;=\; \lim_{N\to\infty} g_0(t) Z_0 + \sum_{n=0}^{N} \sum_{k=0}^{2^n-1} g_{n,k}(t) Z_{n,k}\,, \tag{6}$$

where the convergence is still with respect to the L^2 norm on $L^2(\Omega)$. Then, for every $N \in \mathbb{N}$ we define a random function

$$B_N \,:\, \begin{cases} (\Omega,\,\mathcal{F},\,\mathbb{P}) \longrightarrow \mathcal{C}([0,1]) \\ \qquad \omega \longmapsto B_N^\omega \;=\; Z_0(\omega) g_0 + \displaystyle\sum_{n=0}^{N} F_n(\omega)\,, \qquad \text{with} \qquad F_n(\omega) \;=\; \displaystyle\sum_{k=0}^{2^n-1} Z_{n,k}(\omega) g_{n,k}\,, \end{cases}$$

where $\mathcal{C}([0,1])$ is the set of continuous functions on $[0,1]$, so that (6) boils down to

$$\forall t \in [0,1] \qquad \Phi(1_{[0,\,t]}) \;=\; \lim_{N\to\infty} B_N(t)\,, \tag{7}$$

where the convergence is once again in $L^2(\Omega)$.

We now want to show that, almost surely, the above convergence is uniform on the segment $[0,1]$, which we do by showing that the series of the functions F_n converges normally, that is to say, that the series of general term $\|F_n\|_\infty$ converges, except for ω in a negligible set.

For every $n \in \mathbb{N}$, since the functions $(g_{n,k})_{0\leq k<2^n}$ have disjoint support, we have

$$\|F_n\|_\infty \;=\; \max_{0\leq k\leq 2^n-1} |Z_{n,k}|\, \|g_{n,k}\|_\infty \;=\; \max_{0\leq k\leq 2^n-1} |Z_{n,k}|\, g_{n,k}\Big(\frac{2k+1}{2^{n+1}}\Big) \;=\; \frac{1}{2^{n/2+1}} \max_{0\leq k\leq 2^n-1} |Z_{n,k}|\,.$$

Then, for every $n \in \mathbb{N}$, we have

$$\mathbb{P}\bigg(\|F_n\|_\infty \geq \frac{1}{2^{n/4}}\bigg) \;=\; \mathbb{P}\Big(\exists k \in \{0,\,\ldots,\,2^n-1\} \,:\, |Z_{n,k}| \geq 2^{n/4+1}\Big) \;\leq\; 2^n\, \mathbb{P}\big(|Z_0| \geq 2^{n/4+1}\big)\,.$$

At this point, we use the following lemma:

Lemma 4. *If Z is a standard Gaussian variable, then for all $x \geq 0$, we have $\mathbb{P}(|Z| \geq x) \leq e^{-x^2/2}$.*

The proof of this lemma is left as an exercise. Hint: simply use that for $x,\, t \geq 0$, we have $(x+t)^2 \geq x^2 + t^2$.

With this lemma, we obtain

$$\sum_{n\in\mathbb{N}} \mathbb{P}\bigg(\|F_n\|_\infty \geq \frac{1}{2^{n/4}}\bigg) \;\leq\; \sum_{n\in\mathbb{N}} 2^n \exp\big(-2^{n/2+1}\big) \;<\; \infty\,.$$

Thus, Borel-Cantelli's lemma ensures that there exists $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ and for every $\omega \in \Omega_0$, for n large enough we have $\|F_n\|_\infty < 1/2^{n/4}$, which implies that

$$\sum_{n=0}^{\infty} \|F_n\|_\infty \;<\; \infty\,.$$

Let us fix $\omega \in \Omega_0$. Then, the sequence of functions $(B_N^\omega)_{N\in\mathbb{N}}$ converges uniformly on the segment $[0,1]$ to a function $B^\omega : [0,1] \to \mathbb{R}$. Since B_N^ω is continuous for every N and the convergence is uniform, the function B^ω is continous on $[0,1]$. For $\omega \in \Omega \setminus \Omega_0$, we define $B^\omega(t) = 0$ for every $t \in [0,1]$, so that for every ω, the function B^ω is continuous.

Now, note that for every $t \in [0,1]$ we have $B_N(t) \to B(t)$ almost surely, but from (7) we also have the convergence $B_N(t) \to \Phi(1_{[0,t]})$ in L^2. Thus, for every $t \in [0,1]$, almost surely, we have $B(t) = \Phi(1_{[0,t]})$. This concludes the construction of Brownian motion, because we already checked that if $B(t) = \Phi(1_{[0,t]})$ then B satisfies all the conditions to be a Brownian motion, apart from the continuity condition that we checked just above.

8.3 Other properties

We conclude this section with a few properties of the paths of Brownian motion.

Definition 32. *If B is a Brownian motion, we define its associated continuous-time filtration $(\mathcal{F}_t)_{t\geq 0}$ by writing $\mathcal{F}_t = \sigma(B_s,\, 0 \leq s \leq t)$ for every $t \geq 0$, and we also define $\mathcal{F}_\infty = \sigma(B_t,\, t \geq 0)$.*

Proposition 22 (Blumenthal's $0-1$ law)**.** *The σ-algebra $\mathcal{F}_{0+} = \cap_{st>0}\mathcal{F}_t$ is trivial, in the sense that every event $A \in \mathcal{F}_{0+}$ has probability 0 or 1.*

Proof. Let $A \in \mathcal{F}_{0+}$, and let us show that A is independent of $\mathcal{F}_\infty$. For every $k \geq 1$, for every choice of times $0 \leq t_1 < \cdots < t_k$ and for every bounded and continuous function $f : \mathbb{R}^k \to \mathbb{R}$, we have

$$\begin{aligned}
\mathbb{E}\big[1_A f(B_{t_1}, \ldots, B_{t_n})\big] &= \mathbb{E}\Big[\lim_{\varepsilon\to 0} 1_A f(B_{t_1+\varepsilon} - B_\varepsilon, \ldots, B_{t_n+\varepsilon} - B_\varepsilon)\Big] \\
&= \lim_{\varepsilon\to 0} \mathbb{E}\big[1_A f(B_{t_1+\varepsilon} - B_\varepsilon, \ldots, B_{t_n+\varepsilon} - B_\varepsilon)\big] \\
&= \lim_{\varepsilon\to 0} \mathbb{P}(A)\,\mathbb{E}\big[f(B_{t_1+\varepsilon} - B_\varepsilon, \ldots, B_{t_n+\varepsilon} - B_\varepsilon)\big] \\
&= \mathbb{E}\Big[\lim_{\varepsilon\to 0} 1_A f(B_{t_1}, \ldots, B_{t_n})\Big] \\
&= \mathbb{P}(A)\,\mathbb{E}\big[f(B_{t_1}, \ldots, B_{t_n})\big],
\end{aligned}$$

where the first and last equalities follow from the continuity of the Brownian motion and of f, the second and the fourth equalities follow from the dominated convergence theorem and for the third equality we used that $A \in \mathcal{F}_\varepsilon$ and independence of increments.

Hence, if $A \in \mathcal{F}_{0+}$, then A is independent of cF_∞, and therefore of itself, because $A \in \mathcal{F}_\infty$. Therefore, we have $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, which implies that $\mathbb{P}(A) \in \{0,1\}$. ☐

Corollary 4 (Kolmogorov's $0-1$ law)**.** *The tail σ-algebra $\cap_{t\geq 0}\sigma(B_s,\, s \geq t)$ is also trivial.*

Proof. It follows from Blumenthal's 0-1 by considering the time inversion of Brownian motion, i.e., the process $t \mapsto tB_{1/t}1_{t>0}$. ☐

Corollary 5. *It holds almost surely that for all $t > 0$,*

$$\sup_{[0,\,t]} B \;>\; 0 \qquad \textit{and} \qquad \inf_{[0,\,t]} B \;<\; 0\,.$$

Proof. Consider the event

$$A \;=\; \bigcap_{t>0} \Big\{ \sup_{[0,\,t]} B \;>\; 0 \Big\}\,.$$

For every $s > 0$, we have

$$A \;=\; \bigcap_{0<t\leq s} \Big\{ \sup_{[0,\,t]} B \;>\; 0 \Big\} \;\in\; \mathcal{F}_s\,,$$

so that we have $A \in \mathcal{F}_{0+}$. Hence, Blumenthal's 0-1 law entails that $\mathbb{P}(A) \in \{0, 1\}$.

Besides, we also have

$$A \;=\; \bigcap_{n\geq 1} A_n \qquad \text{with} \qquad A_n \;=\; \Big\{ \sup_{[0,\,1/n]} B \;>\; 0 \Big\}$$

and this intersection is non-increasing, so that $\mathbb{P}(A) = \lim \mathbb{P}(A_n)$. Yet, for every $n \geq 1$ we have

$$\mathbb{P}(A_n) \;\geq\; \mathbb{P}\big(B_{1/n} > 0\big) \;=\; \frac{1}{2}\,,$$

because $B_{1/n}$ is a centered Gaussian variable with positive variance. Thus, we conclude that $\mathbb{P}(A) \neq 0$ and finally that $\mathbb{P}(A) = 1$. The statement on the infimum also holds almost surely by symmetry. □

Corollary 6. *It holds almost surely that*

$$\limsup_{t\to+\infty} B_t \;=\; +\infty \qquad \textit{and} \qquad \liminf_{t\to+\infty} B_t \;=\; -\infty\,.$$

Proof. First note that the continuity of trajectories implies that

$$\Big\{ \limsup_{t\to+\infty} B_t \;=\; +\infty \Big\} \;=\; \Big\{ \sup_{t\geq 0} B_t \;=\; +\infty \Big\}\,.$$

Let $a > 0$. For $n \in \mathbb{N}$ we can write

$$\begin{aligned} \mathbb{P}\Big(\sup_{t\geq 0} B_t \;>\; a \Big) \;\geq\; \mathbb{P}\Big(\sup_{0\leq t\leq n} B_t \;>\; a \Big) \;&=\; \mathbb{P}\Big(\sup_{0\leq t\leq 1} B_{nt} \;>\; a \Big) \;=\; \mathbb{P}\Big(\sup_{0\leq t\leq 1} \frac{B_{nt}}{\sqrt{n}} \;>\; \frac{a}{\sqrt{n}} \Big) \\ &=\; \mathbb{P}\Big(\sup_{0\leq t\leq 1} B_t \;>\; \frac{a}{\sqrt{n}} \Big) \;\overset{n\to\infty}{\longrightarrow}\; \mathbb{P}\Big(\sup_{[0,1]} B \;>\; 0 \Big) \;=\; 1\,, \end{aligned}$$

where the first equality on the second line follows from the time-change property of Brownian motion (see Proposition 20), and the last equality is given by Corollary 5. Therefore, for every $a > 0$ we have

$$\mathbb{P}\Big(\sup_{t\geq 0} B_t \;>\; a \Big) \;=\; 1\,,$$

which implies that this supremum is almost surely infinite. □

We now wish to present the strong Markov property for Brownian motion. To do so, we need the following definitions:

Definition 33. *A random variable T taking values in $\mathbb{R}_+ \cup \{+\infty\}$ is a stopping time relative to a continuous-time filtration $(\mathcal{F}_t)_{t\geq 0}$ if for every $t \geq 0$, we have $\{T \leq t\} \in \mathcal{F}_t$.*

Exercise 7. *Is the above definition equivalent to the same definition where instead of the events $\{T \leq t\}$ we consider the events $\{T = t\}$? (give a proof or a counterexample).*

Definition 34. *If T is a stopping time relative to a continuous-time filtration $(\mathcal{F}_t)_{t\geq 0}$, we define the σ-field of the past until T as*

$$\mathcal{F}_T \;=\; \big\{ A \in \mathcal{F}_\infty \;:\; \forall t \geq 0\,,\; A \cap \{T \leq t\} \in \mathcal{F}_t \big\}\,.$$

Exercise 8. *Same question as above: can $T \leq t$ be equivalently replaced by $T = t$ in the above definition?*

Theorem 22 (Strong Markov property for Brownian Motion)**.** *Let B be a Brownian motion, let T be a stopping time (relative to the filtration of B) such that $\mathbb{P}(T < \infty) > 0$. Then, under the conditional probability measure $\mathbb{P}(\cdot | T < \infty)$, the process $B^{(T)}$ defined by*

$$B^{(t)} \;:\; t \geq 0 \;\longmapsto\; B_{T+t} - B_T$$

is a Brownian motion, which is independent of $\mathcal{F}_T$.

Proof. To simplify, we assume that $T < \infty$ almost surely. Let us show that for every $A \in \mathcal{F}_T$, every $k \geq 1$ and $0 \leq t_1 < \cdots < t_k$ and $f : \mathbb{R}^k \to \mathbb{R}$ a continuous and bounded function,

$$\mathbb{E}\big[1_A f(B^{(T)}_{t_1}, \ldots, B^{(T)}_{t_k})\big] \;=\; \mathbb{P}(A)\,\mathbb{E}\big[f(B_{t_1}, \ldots, B_{t_k})\big]\,. \tag{8}$$

This would show that $B^{(T)}$ is independent of $\mathcal{F}_T$ and the case $A = \Omega$ shows that $B^{(T)}$ has the same distribution as B, i.e., it is a Brownian motion. Let A, k, $t_1, \ldots, t_k$, f be as above. The idea to show (8) is to discretize T and to use the weak Markov property. More precisely, for every $n \geq 1$, we define

$$T_n \;=\; \frac{\lceil nT \rceil}{n} \;=\; \inf\left\{\frac{k}{n},\, k \in \mathbb{N},\, \frac{k}{n} \geq T\right\}.$$

Then, we have $T_n - 1/n < T \leq T_n$, whence $T_n \to T$ almost surely, and therefore we have the almost sure convergence

$$f\big(B^{(T_n)}_{t_1}, \ldots, B^{(T_n)}_{t_k}\big) \;\overset{n\to\infty}{\longrightarrow}\; f\big(B^{(T)}_{t_1}, \ldots, B^{(T)}_{t_k}\big)\,.$$

Thus, by the dominated convergence theorem, we have

$$\mathbb{E}\big[1_A f(B^{(T)}_{t_1}, \ldots, B^{(T)}_{t_k})\big] \;=\; \lim_{n\to\infty} \mathbb{E}\big[1_A f(B^{(T_n)}_{t_1}, \ldots, B^{(T_n)}_{t_k})\big] \;=\; \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbb{E}\big[1_{A \cap \{T_n = k/n\}}\, f(B^{(k/n)}_{t_1}, \ldots, B^{(k/n)}_{t_k})\big]\,.$$

Note now that for every $n \geq 1$ and $k \in \mathbb{N}$,

$$A \cap \left\{T_n = \frac{k}{n}\right\} \;=\; A \cap \left\{T \leq \frac{k}{n}\right\} \setminus \left\{T \leq \frac{k-1}{n}\right\} \;\in\; \mathcal{F}_{k/n}\,,$$

because $A \in \mathcal{F}_T$. Yet, the weak Markov property tells us that for every $n \geq 1$ and $k \in \mathbb{N}$, the process $B^{(k/n)}$ is a Brownian motion independent of $\mathcal{F}_{k/n}$, whence

$$\mathbb{E}\big[1_A f(B^{(T)}_{t_1}, \ldots, B^{(T)}_{t_k})\big] \;=\; \lim_{n\to\infty} \sum_{k=0}^{\infty} \mathbb{P}\left(A \cap \left\{T_n = \frac{k}{n}\right\}\right) \mathbb{E}\big[f(B_{t_1}, \ldots, B_{t_k})\big] \;=\; \mathbb{P}(A)\,\mathbb{E}\big[f(B_{t_1}, \ldots, B_{t_k})\big]\,,$$

applying again dominated convergence. □

This strong Markov property has the following nice property:

Theorem 23 (Reflection principle)**.** *For $t \geq 0$, let $S_t = \sup_{s \leq t} B_s$. Then, we have*

1. *For every $0 \leq a \leq b$, we have*
$$\mathbb{P}\big(S_t \geq a,\, B_t \leq b\big) \;=\; \mathbb{P}\big(B_t \geq 2a - b\big)\,.$$
2. *For every $t \geq 0$, the variable S_t has the same law as $|B_t|$.*

Proof. We start by proving the first point. We consider the stopping time

$$T \;=\; \inf\big\{t \geq 0 \,:\, B_t = a\big\}\,.$$

We already know that $T < \infty$ almost surely, and the idea is to reflect the part of the trajectory of the Brownian motion which is after this time, with respect to the horizontal line at height a. First, we write

$$\mathbb{P}\big(S_t \geq a,\, B_t \leq b\big) \;=\; \mathbb{P}\big(T \leq t,\, B_t \leq b\big) \;=\; \mathbb{P}\big(T \leq t,\, B^{(T)}_{t-T} \leq b - a\big)\,. \tag{9}$$

The last term above is the probability of an event which is expressed as a function of T and of $B^{(T)}$. Yet, by the strong Markov property, we have that $B^{(T)}$ is independent of T (note that T is $\mathcal{F}_T$-measurable), and we have

$$B^{(T)} \;\overset{\text{in law}}{=}\; B \;\overset{\text{in law}}{=}\; -B \;\overset{\text{in law}}{=}\; -B^{(T)}\,,$$

where the second equality is symmetry of Brownian motion (see Proposition 20). Therefore, Equation (9) becomes

$$\mathbb{P}\big(S_t \geq a,\, B_t \leq b\big) \;=\; \mathbb{P}\big(T \leq t,\, -B^{(T)}_{t-T} \leq b-a\big) \;=\; \mathbb{P}\big(T \leq t,\, B_t - B_T \geq a-b\big) \;=\; \mathbb{P}\big(S_t \geq a,\, B_t \geq 2a-b\big)\,.$$

Then, the second point of the statement follows by taking $a = b$ and writing

$$\mathbb{P}\big(S_t \geq a\big) \;=\; \mathbb{P}(S_t \geq a,\, B_t > a) + \mathbb{P}(S_t \geq a,\, B_t \leq a) \;=\; \mathbb{P}(B_t \geq a) + \mathbb{P}(B_t \geq 2a-a) \;=\; \mathbb{P}(|B_t| \geq a)\,,$$

which concludes the proof. □

Note that the second point in the above theorem implies that $S_1 \in L^p$ for every $p < \infty$, which in particular implies (3).

Students interested in learning more on properties of Brownian motion can consult the book "Brownian motion" by Mörters and Peres.

References

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

[2] Allan Gut. *Probability: a graduate course*. Springer Texts in Statistics. Springer, New York, 2005.

[3] Jean-Fran¸cois Le Gall. *Measure theory, probability, and stochastic processes*, volume 295 of *Graduate Texts in Mathematics*. Springer, Cham, 2022.