A REVIEW OF PROBABILITY FOUNDATIONS

The contents of this class are covered in numerous textbooks and lecture notes, here are just a few of them :
Le Gall "Measure Theory, Probability, and Stochastic Processes"
Durrett "Probability : Theory and examples"
Billingsley "Probability and Measure"
Gut "Probability : a graduate course"

1. Basics of measure theory and integration

Definition 1.1. *$\mathcal{E} \subset \mathcal{P}(E)$ is a **σ-algebra** on a set E if :*

- $\emptyset \in \mathcal{E}$
- *$A \in \mathcal{E} \Rightarrow A^c \in \mathcal{E}$ (where $A^c = E \setminus A$)*
- *If $(A_n)_n$ is a countable family of elements of $\mathcal{E}$, then $\cup_n A_n \in E$.*

The pair $(E, \mathcal{E})$ is then called a measurable space. Examples : $\{\emptyset, E\}$, $\mathcal{P}(E)$ are σ-algebras.

Given a family $\mathcal{F}$ of subsets of E, the σ-algebra generated by $\mathcal{F}$, denoted $\sigma(\mathcal{F})$, is the smallest σ-algebra on E containing $\mathcal{F}$. Since an intersection of σ-algebras is a σ-algebra, it can in fact be obtained as

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

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 1.2. *Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. A map $f : E \to F$ is called measurable if*

$$\forall A \in \mathcal{F},\ f^{-1}(A) := \{x \in E,\ f(x) \in A\} \in \mathcal{E}.$$

Definition 1.3. *A **measure** μ on a measurable space $(E, \mathcal{E})$ (where $\mathcal{E}$ is a σ-algebra on E), is a function $\mu : \mathcal{E} \to [0, +\infty]$ such that, if $A_n, n \in \mathbb{N}$ are disjoint elements of $\mathcal{E}$, it holds that*

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

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

A basic example is the counting measure, i.e. $\mu(A) = card(A)$. Another one is the Dirac mass δ_x at $x \in E$, defined by $\delta_x(A) = 1_A(x)$.

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

Remark : a useful technical tool to prove equality of two measures is the **monotone class theorem** (not defined in these notes, but easy to look up), which has the following consequence.

Proposition 1.1. *Let μ, ν be measures on $(E, \mathcal{E})$. Assume that $\mathcal{C} \subset \mathcal{E}$ is stable under finite intersections, satisfies $\sigma(\mathcal{C}) = \mathcal{E}$, and $\mu(A) = \nu(A)$ for all $A \in \mathcal{C}$.*

If in addition, there is an increasing sequence $(E_n)_n \subset \mathcal{C}$, $\cup_n E_n = E$ s.t. $\mu(E_n) = \nu(E_n) < \infty$, then it holds that $\mu = \nu$ (on $\mathcal{E}$).

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

The main use of measures is that they allow to define integrals $\int f d\mu$ of measurables functions $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. More precisely, there exists a unique way of defining, for any non-negative f, $\int f d\mu \in [0, +\infty]$ (also written $\int f(x) d\mu(x)$), s.t. the following is satisfied :

- If $A \in \mathcal{E}$, then $\int 1_A d\mu = \mu(A)$,
- For any measurable $f, g \geq 0$ and $a, b \in \mathbb{R}_+$, $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$,
- If $0 \leq f \leq g$, then $\int f d\mu \leq \int g d\mu$.

For $\mathbb{R}$-valued functions, we then say that f is integrable if $\int |f| d\mu < +\infty$, and we then define $\int f d\mu = \int f_+ d\mu - \inf f_- d\mu$. The set of integrable functions is denoted $L^1(E, \mathcal{E}, \mu)$.

The three following limit theorems are very useful.

Theorem 1.1. ***(Monotone convergence)*** *Let $f_n \geq 0$ be a non-decreasing sequence of measurable 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 $f_n \geq 0$ be measurable functions, then it holds that*

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

(Dominated convergence) *Let f_n be measurable and such that, it holds that*

$$\lim_{n\to\infty} f_n(x) = f(x) \text{ for } \mu\text{-a.e. } x,$$

and there exists an integrable function g such that

$$\forall n,\ |f_n(x)| \leq g(x) \text{ for } \mu\text{-a.e. } x.$$

Then it holds that

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

Derivation under integral sign :

Proposition 1.2. *Let μ be a measure on $(E, \mathcal{E})$, and let $f = f(\lambda, x)$ be a function on $I \times E$ (where I is a subinterval of $\mathbb{R}$) s.t. for some $k \in \{0, 1\}$:*

(i) For all λ, $f(\lambda, \cdot)$ is $\mathcal{E}$-measurable and μ-integrable,
(ii) For μ-a.e. x, $\lambda \mapsto f(\lambda, x)$ is in $C^k(I)$,
(iii) There exists a μ-integrable function g s.t. for μ-a.e. x, $\forall \lambda \in I$, $\left|\partial_\lambda^k f(\lambda, x)\right| \leq g(x)$.
Then

$$F : \lambda \mapsto \int f(\lambda, x) \mu(dx)$$

is in $C^k(I)$, and

$$F^{(k)}(\lambda) = \int (\partial_\lambda^k) f(\lambda, x) \mu(dx).$$

Definition 1.4. *Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two measured spaces. We can define the product σ-algebra $\mathcal{E} \otimes \mathcal{F} = \sigma(\mathcal{E} \times \mathcal{F})$. The* product measure *$\mu \otimes \nu$ is the unique measure on $(E \times F, \mathcal{E} \otimes \mathcal{F})$ such that*

$$\forall A \in \mathcal{E}, B \in \mathcal{F}, \ \ (\mu \otimes \nu)(A \times B) = \mu(A)\nu(B).$$

A measure is said **σ-finite** if there exists a sequence of measurable sets E_n with $\cup_{n\in\mathbb{N}} E_n = E$ and $\mu(E_n) < +\infty$.

Theorem 1.2 (Fubini)**.** *Assume that both μ and ν are σ-finite. Let $f : E \times F \to \mathbb{R}$ be a measurable function w.r.t. $\mathcal{E} \otimes \mathcal{F}$. Then the following are equivalent :*

(1) *f is integrable w.r.t. $\mu \otimes \nu$,*
(2) $\int \left(\int |f(x, y)| \, \mu(dx) \right) \nu(dy) < +\infty$,
(3) $\int \left(\int |f(x, y)| \, \nu(dy) \right) \mu(dx) < +\infty$,

and if this holds, one has

$$\int f(x, y)(\mu \otimes \nu)(dx, dy) \ = \ \int \left(\int f(x, y)\mu(dx) \right) \nu(dy) \ = \ \int \left(\int f(x, y)\nu(dy) \right) \mu(dx).$$

The formula above also holds for measurable non-negative f.

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

$$\nu(A) = \int 1_A(x) f(x) \mu(dx).$$

We say that f is the density of ν w.r.t. μ, also written $\frac{d\nu}{d\mu} = f$.

A measure ν is said to be **absolutely continuous** w.r.t. μ, (written $\nu \ll \mu$) if for all $A \in \mathcal{E}$, $\mu(A) = 0 \Rightarrow \nu(A) = 0$.

It is easy to check that if ν admits a density w.r.t. μ, then $\nu \ll \mu$. The converse turns out to be also true.

Theorem 1.3 (Radon-Nikodym)**.** *Let ν and μ be two σ-finite measures on a measured space $(E, \mathcal{E})$, s.t. $\nu \ll \mu$. Then ν admits a density f w.r.t. μ.*

2. Probability : random variables, independence,

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

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

Definition 2.1. *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** (abbreviated **r.v.**) X is a measurable map from $(\Omega, \mathcal{F})$ to a measurable space $(E, \mathcal{E})$. (When E is not specified, this will always mean $E = \mathbb{R}$).*

*The **expectation** of a non-negative or integrable r.v. X, is $\mathbb{E}[X] = \int X(\omega)\mathbb{P}(d\omega)$.*

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

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

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).

The cumulant distribution function (c.d.f.) of a scalar r.v. 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.

Proposition 2.1. *Let X be a scalar r.v., $h : \mathbb{R} \to \mathbb{R}$ measurable, s.t. $h(X)$ is integrable. Then*

$$\mathbb{E}\left[h(X)\right] = \int h(x)\mathcal{L}^X(dx).$$

For instance, if the law of X admits a density f w.r.t. Lebesgue measure, then

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

Given a r.v. X with values in $(E, \mathcal{E})$, the σ-algebra generated by X, denoted $\sigma(X)$, is the smallest σ-subalgebra of $\mathcal{F}$ for which X is measurable, explicitely it can be written as

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

Proposition 2.2. *Let X be a $(E, \mathcal{E})$- valued r.v. and Y be a $\sigma(X)$-measurable r.v., then there exists a measurable $\psi : E \to \mathbb{R}$ s.t. $Y = \psi(X)$.*

Definition 2.2 (L^p spaces)**.** *Fix $1 \le p < \infty$. Given a r.v. X, its L^p norm is defined by*

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

Then

$$L^p(\Omega) = \left\{X : \|X\|_{L^p} < \infty\right\} / \sim$$

equipped with $\|\cdot\|_{L^p}$, is a Banach space. ($X \sim Y$ iff $X = Y$ $\mathbb{P}-a.s.$).

We can also define L^∞, with norm

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

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

Proposition 2.3. ***(Jensen)*** *Let $\phi : \mathbb{R} \to \mathbb{R}$ be convex and X a r.v., 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 r.v. X,Y,*

$$\mathbb{E}\left[|XY|\right] \le \|X\|_{L^2}\|Y\|_{L^2}.$$

(Hölder) *Fix $1 \le p, q \le \infty$ s.t. $\frac{1}{p} + \frac{1}{q} = 1$. Then for any two r.v. X,Y,*

$$\mathbb{E}\left[|XY|\right] \le \|X\|_{L^p}\|Y\|_{L^q}.$$

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

Definition 2.3. *Let $X \in L^1(\Omega)$. The **mean** of X is simply $\mathbb{E}[X]$. X is **centered** if its mean is 0.*

*Let $X \in L^2$. The **variance** of X is defined by*

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

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

Proposition 2.4. *Let U be a non-decreasing positive function. Then for any r.v. X s.t. 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. Follows from taking $\mathbb{E}$ in the inequality $U(a)1_{\{X\geq a\}} \leq U(X)$. □

This implies the following classical special cases,

$$\forall a > 0 \quad \mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|]}{a} \text{ (Markov) }, \mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{Var(X)}{a^2} \text{ (Bienaymé-Tchebychev) },$$

$$\forall \lambda > 0, \ \forall a \in \mathbb{R}, \ \mathbb{P}(X \geq a) \leq \mathbb{E}[e^{\lambda X}]e^{-\lambda a} \text{ (Chernoff) }.$$

Definition 2.4. *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 iff, for any $i_1, \dots, i_n \in I$,

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

Similarly, two $(E, \mathcal{E})$-valued r.v. X and Y are ***independent*** *if*

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

and a family X_i, $i \in I$ of random variables is independent iff, for any $i_1, \dots, i_n \in I$, any measurable $A_1, \dots, A_n$,

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

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

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

Note that it is obvious from the definition that 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 family of random variables). In particular, in conjunction with Proposition 2.1 and Fubini's theorem, this implies that if X, Y are independent, f, g functions , 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

$$\limsup_n A_n = \cap_{n\geq 0} \left(\cup_{k\geq n} A_k\right),$$

this event can alternately be characterized as

$$\limsup_n A_n = \{\omega \ : \ \omega \in \text{ infinitely many } A_n \text{ 's}\},$$

Proposition 2.5 (Borel-Cantelli)**.** *(1) Let A_n, $n \in \mathbb{N}$ be a sequence of events s.t. $\sum_n \mathbb{P}(A_n) < \infty$. Then $\mathbb{P}(\limsup A_n) = 0$.*

(2) Let A_n, $n \in \mathbb{N}$ be an independent *sequence of events s.t. $\sum_n \mathbb{P}(A_n) = \infty$. Then $\mathbb{P}(\limsup A_n) = 1$.*

Proof. (1) By Fubini,

$$\mathbb{E}\left[\sum_n 1_{A_n}\right] = \sum_n \mathbb{P}(A_n) < \infty,$$

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

(2) Note that

$$(\limsup_n A_n)^c = \cup_{n\geq 0} \left(\cap_{k\geq n} A_k^c\right),$$

and since this is an increasing union, we have

$$\mathbb{P}((\limsup_n A_n)^c) = \lim_n \mathbb{P}\left(\cap_{k\geq n} A_k^c\right)$$

each of these can be bounded for $p \geq 0$ by

$$\mathbb{P}\left(\cap_{k\geq n} A_k^c\right) \leq \mathbb{P}\left(\cap_{k=n}^{n+p} A_k^c\right) = (1-\mathbb{P}(A_n))\dots(1-\mathbb{P}(A_{n+p})) \leq \exp\left(-\mathbb{P}(A_n) - \dots - \mathbb{P}(A_{n+p})\right)$$

where we used independence, and this goes to 0 as $p \to \infty$ by assumption. □

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

We first record the important fact :

Lemma 2.1. *Let $X_1, \dots, X_n$ be pairwise independent elements of $L^2(\Omega)$. Then $Var(X_1 + \dots + X_n) = \sum_{i=1}^n Var(X_i)$.*

Exercise 1. *Let A_n be pairwise independent, s.t. $\sum_n \mathbb{P}(A_n) = +\infty$. 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 $Var(S_n) \leq m_n$. Deduce from Chebychev's inequality that $\mathbb{P}(S \leq m_n/2) \leq \frac{4}{m_n}$, and conclude that $\mathbb{P}(S = \infty) = 1$.*

3. Convergence of random variables

Let X, and X_n, $n \in \mathbb{N}$ be some r.v.'s defined on the same probability space.

Definition 3.1. *We say that X_n converges to X :*
in probability *if $\forall \varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to_{n\to\infty} 0$.*
in L^p *(for a given $p \geq 1$) if $\mathbb{E}[|X_n - X|^p] \to_{n\to\infty} = 0$.*
almost surely *if $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.*
in law *if for any bounded continuous ϕ, $\mathbb{E}[\phi(X_n)] \to_{n\to\infty} \mathbb{E}[\phi(X)]$.*

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

Proposition 3.1. *(1) $X_n \to X$ almost surely or in $L^p \Rightarrow X_n \to X$ in probability.*
(2) $X_n \to X$ in probability $\Rightarrow$ There is a subsequence (n_k) s.t. $X_{n_k} \to X$ almost surely.
(3) $X_n \to X$ in probability $\Rightarrow X_n \to X$ in law
(4) If $X \equiv c \in \mathbb{R}$ is constant, then $X_n \to X$ in law $\Rightarrow X_n \to X$ in probability.

Proof. (1) Dominated convergence for 'a.s.', Markov's inequality for L^p.

(2) Since $X_n \to X$ in probability, by a diagonal procedure we can find a subsequence n_k s.t.

$$\mathbb{P}\left(|X_{n_k} - X| \geq \frac{1}{k}\right) \leq \frac{1}{k^2}.$$

By the Borel-Cantelli lemma, this implies that a.s., for k large enough, $|X_{n_k} - X| < \frac{1}{k}$ which of course implies that $\lim_k X_{n_k} = X$.

(3) Assume that $X_n \to X$ in probability, but not in law. This implies the existence of ϕ, continuous and bounded, and of a subsequence n_k s.t.

$$\liminf_k |\mathbb{E}[\phi(X_{n_k})] - \mathbb{E}[\phi(X)]| > 0.$$

On the other hand, by (2), up to taking another subsequence, we can assume that $X_{n_k} \to X$ almost surely. By dominated convergence, this implies $\lim_k \mathbb{E}[\phi(X_{n_k})] = \mathbb{E}[\phi(X)]$, a contradiction with the above.

(4) For $\varepsilon > 0$, let ϕ be a continuous bounded function s.t. $\phi(c) = 0$ and $\phi(x) = 1$ if $|x - c| \geq \varepsilon$. Then $\mathbb{P}(|X_n - c| \geq \varepsilon) \leq \mathbb{E}[\phi(X_n)] \to \phi(c) = 0$. □

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 a (law of) r.v. X is defined by

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

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

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

Proposition 3.2. *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 3.1. *The following are equivalent :*
(1) $X_n \to_{n\to\infty} X$ *in law,*
(2) $\forall t \in \mathbb{R}$, $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$.

Proof. 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 $|\sin(x)| \leq \frac{|x|}{2}$ for $|x| \geq 2$.)
- Deduce that if (2) holds, then for any $\varepsilon > 0$, for r large enough, $\limsup_n \mathbb{P}(|X_n| \geq r) \leq \varepsilon$.
- Use that for any $R > r > 0$, functions of the form $\sum_{k=-N}^{N} a_k e^{i\frac{k\pi x}{R}}$ are dense in $C([-r, r])$, in combination with the previous step, to conclude.

□

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

Theorem 3.2. $X_n \to_{n\to\infty} X$ *in law if and only if* $\forall\ x \in \mathbb{R}$, *if* F_X *is continuous at* x, *then* $\lim_{n\to\infty}\ F_{X_n}(x) = F_X(x)$.

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 3.2. *A family* $\{X_i, i \in I\}$ *of random variables is* ***tight*** *if, for any* $\varepsilon > 0$, *there exists a compact* K, *s.t.* $\sup_i \mathbb{P}(X_i \notin K) \leq \varepsilon$.

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

Theorem 3.3. *(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 r.v.* X *(possibly on a different probability space) and a subsequence* X_{n_k} *s.t.* $X_{n_k} \to X$ *in law.*

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

Lemma 3.1. *A function $F : \mathbb{R} \to \mathbb{R}$ is the c.d.f. 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) *$\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$*

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 r.v. on $[0, 1]$ (if F is injective, then $G = F^{-1}$). (Note : this may be in practice a useful way to simulate a r.v. whose c.d.f. is known). □

Proof of Theorem 3.3. (1) is easier and left as exercise.

(2) (Sketch). By a diagonal procedure, we construct a subsequence n_k s.t. for each rational q, $F_{X_{n_k}}(q)$ converges to some limit $\tilde{F}(q)$. We then let for $x \in \mathbb{R}$, $F(x) := \lim_{q\in\mathbb{Q},q\downarrow x} \tilde{F}(q)$. We then verify that F satisfies the assumptions of Lemma 3.1 (tightness is only used in point 3.). □

4. LLN and CLT

In this section, we consider a sequence X_n, $n \ge 1$ of independent random variables, s.t. for each n, X_n and X_1 have the same law. We say that this is an **i.i.d.** sequence (for "independent and identically distributed").

4.1. Law of large numbers (LLN).

Theorem 4.1 (Weak LLN). *Assume that X_n is an i.i.d. sequence, with $\mathbb{E}|X_1| < \infty$, and let $m = \mathbb{E}[X_1]$. Then*

$$\frac{\sum_{k=1}^n X_k}{n} \to_{n\to\infty} m \text{ in probability.}$$

Proof. It suffices to check convergence in law, which can be done by the characteristic function :

$$\mathbb{E}\left[\exp(itn^{-1}\sum_{k=1}^n X_k)\right] = \Phi_{X_1}(t/n)^n = \left(1 + m\frac{t}{n} + o(\frac{t}{n})\right)^n \to e^{tm}.$$

□

In fact, the above can be strenghtened to a.s. convergence.

Theorem 4.2 (Strong LLN). *Assume that X_n is an i.i.d. sequence, with $\mathbb{E}|X_1| < \infty$, and let $m = \mathbb{E}[X_1]$. Then*

$$\frac{\sum_{k=1}^n X_k}{n} \to_{n\to\infty} m \text{ almost surely.}$$

Proof. We sketch a proof due to Etemadi who in fact only uses pairwise independence.

First, considering separately X_n^+ and X_n^-, we may assume that $X_1 \ge 0$ a.s., and we aim to show that letting $S_n = \sum_{k=1}^n X_k$, $S_n/n \to \mathbb{E}[X_1]$ a.s.

We let $Y_n = X_n 1_{\{X_n \le n\}}$. Show that

$$\sum_{n\ge1} \mathbb{P}(X_n \ne Y_n) \le \mathbb{E}[X_1]$$

(this involves rearranging double sums). By Borel-Cantelli, this implies that a.s., for n large enough, $Y_n = X_n$. It therefore suffices to show

$$\text{almost surely, } \lim_{n\to\infty} \frac{\tilde{S}_n}{n} = \mathbb{E}[X_1], \text{ where } \tilde{S}_n = \sum_{k=1}^n Y_k.$$

We now fix $\alpha > 1$, and let $k_n = \lfloor \alpha^k \rfloor$. Then show that for any fixed $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left| \frac{\tilde{S}_{k_n} - \mathbb{E}[\tilde{S}_{k_n}]}{k_n} \right| > \varepsilon \right) \leq C \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{i=1}^{k_n} Var(Y_i) \leq \ldots \leq C' \mathbb{E}[X_1] < \infty.$$

(Again the computation abbreviated as $(\ldots)$ involves re-arranging a double sum). Deduce that a.s., $\frac{\tilde{S}_{k_n}}{k_n} \to \mathbb{E}[X_1]$.

Now for any arbitrary n, there exists $k_{n'}$ with $k_{n'} \leq n \leq k_{n'+1}$ and $n/k_{n'} \leq \alpha$, $n/k_{n'+1} \geq \alpha^{-1}$. Since $\tilde{S}_n$ is increasing, this implies

$$\text{a.s., } \alpha^{-1}\mathbb{E}[X_1] \leq \liminf_n \frac{\tilde{S}_n}{n} \leq \limsup_n \frac{\tilde{S}_n}{n} \leq \alpha \mathbb{E}[X_1].$$

Taking a sequence $\alpha_n \downarrow 1$, we conclude that a.s., $\lim_n \frac{\tilde{S}_n}{n} = \mathbb{E}[X_1]$. □

4.2. **Central Limit Theorem (CLT).**

Definition 4.1. *The standard Gaussian measure (denoted $\mathcal{N}(0,1)$), is the measure on $\mathbb{R}$ with probability density function given by $f(x) = \sqrt{2\pi}^{-1} e^{-\frac{x^2}{2}}$.*

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 4.3 (CLT)**.** *Let X_n be an i.i.d. sequence, with $\mathbb{E}|X_1|^2 < \infty$, and let $m = \mathbb{E}[X_1]$, $\sigma^2 = Var(X_1)$. Then*

$$\frac{\sum_{k=1}^n (X_k - m)}{\sigma\sqrt{n}} \to_{n\to\infty} \mathcal{N}(0,1) \textit{ in law }.$$

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

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

$$\phi_{\frac{\sum_{k=1}^n X_k}{\sqrt{n}}}(t) = \phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 - \frac{1}{2}\frac{t^2}{n} + o\left(\frac{t^2}{n}\right)\right)^n \to_{n\to\infty} e^{-\frac{t^2}{2}}.$$

□

5. Conditional expectations

Preliminary definitions : let A be an event with $\mathbb{P}(A) > 0$. Then, given another event B, the probability of B conditionally on A is

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

and given an integrable r.v. X, we can define the expectation of X, conditionally on A, by

$$\mathbb{E}[X|A] = \frac{\mathbb{E}[X 1_A]}{\mathbb{P}(A)}.$$

We can also define in a simple manner conditional expectation w.r.t. a r.v. Y, as long as Y takes countably many values $y_1, y_2, \dots$ (each with positive probability). Then we let

$$\mathbb{E}[X|Y] = \sum_k \mathbb{E}[X|Y = Y_k] 1_{\{Y=y_k\}}.$$

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

Proposition 5.1. *Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. Then for any integrable (resp. nonnegative) r.v. X on $(\Omega, \mathcal{F}, \mathbb{P})$, there exists a unique (up to a.s. equality) r.v. Y, s.t.*
(1) Y is $\mathcal{G}$-measurable,
(2) For any $\mathcal{G}$-measurable Z s.t. XZ is integrable (resp. $Z \geq 0$), it holds that $\mathbb{E}[XZ] = \mathbb{E}[YZ]$.

Proof. We will treat the case where $X \in L^1$.

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

$$\mathbb{E}[X 1_{\{Y>Y'\}}] = \mathbb{E}[X 1_{\{Y>Y'\}}] = \mathbb{E}[X 1_{\{Y>Y'\}}]$$

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

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

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

is a measure on $(\Omega, \mathcal{G})$, which is absolutely continuous w.r.t. $\mathbb{P}$. 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. ☐

The random variable Y obtained from the above proposition is denoted $\mathbb{E}[X|\mathcal{G}]$, and called **conditional expectation of X w.r.t. $\mathcal{G}$.**

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 $L^2(\Omega, \mathcal{G}, \mathbb{P})$.

We now detail some properties of the conditional expectation.

Proposition 5.2. *Assuming that the r.v.'s X, Y are such that the conditional expectations below make sense, we have the following (in)equalities, understood in the a.s. 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 $X \leq Y$ a.s., 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 $g(y) = \mathbb{E}[f(X, y)]$, which is often written equivalently as*

$$\mathbb{E}[f(X, Y)|\mathcal{G}] = \mathbb{E}[f(X, y)]_{y=Y}.$$

(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 w.r.t. Y by

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

Remark : by Proposition 2.2, there exists a function h s.t. $\mathbb{E}[X|Y] \ = h(Y)$. To prove this identity, by definition it suffices to check that for any measurable (bounded) g, it holds that

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

Exercise 2. *Check that this coincides with the definition given in the beginning of the subsection if Y is discretely valued.*

When the considered r.v.'s have densities, conditional expectations can be computed explicitely.

Proposition 5.3. *Assume that (X,Y) has a law which admits a density $f = f(x,y)$. (for simplicity wrt $dxdy$ but also holds w.r.t. any measure $\mu(dx) \otimes \nu(dy)$). Then it holds, for any h s.t. $\mathbb{E}[h(X)]$ makes sense :*

$$\mathbb{E}[h(X)|Y] = \frac{\int h(x)f(x,Y)dx}{\int f(x,Y)dx} \ a.s. \tag{5.1}$$

Proof. Under the assumption on (X,Y), Y admits a density given by

$$f_y(y) = \int f(x,y)dx$$

Indeed, for any function ψ, by Fubini,

$$\mathbb{E}[\psi(Y)] = \int \psi(y)f(x,y)dxdy = \int \psi(y)\left(\int f(x,y)dx\right)dy.$$

Denote $h^Y(Y)$ the r.h.s. of (5.1). For any bounded function ϕ, we compute (again using Fubini)

$$\begin{aligned}
\mathbb{E}[h(X)\phi(Y)] &= \int h(x)\phi(y)f(x,y)dxdy \\
&= \int \phi(y)\left(\int h(x)f(x,y)dx\right)dy \\
&= \int \phi(y)h^Y(y)f_y(y)dy \\
&= \mathbb{E}[\phi(Y)h^Y(Y)],
\end{aligned}$$

which concludes the proof. □

Let us now discuss the notion of **conditional law**.

Definition 5.1. *A* ***probability kernel*** *on is a map $\nu : \mathbb{R} \times \mathcal{B}(\mathbb{R})$ such that : (i) for all $y \in \mathbb{R}$, $\nu(y,\cdot)$ is a probability measure on $\mathbb{R}$, (ii) for all $A \in \mathcal{B}(\mathbb{R})$, $y \mapsto \nu(y,A)$ is measurable.*

Theorem 5.1. *Let X,Y be two random variables. There exists a probability kernel $\mathcal{L}^{X|Y}$, which is called the* ***law of X, conditionally on*** *Y, and which satisfies, for any bounded measurable f,*

$$\mathbb{E}\left[f(X)|Y\right] = \int f(x)\mathcal{L}^{X|Y}(Y,dx).$$

Examples :
(1) if X independent from Y, then $\mathcal{L}^{X|Y}(y,\cdot) = \mathcal{L}^X$ for any value of y.
(2) If $X = g(Y)$, then $\mathcal{L}^{X|Y}(y,\cdot) = \delta_{g(y)}$.
(3) If X and Y have a joint density $f = f(x,y)$, the conditional distribution is given by Proposition 5.3.
(4) Exercise : if $X = Z$, $Y = |Z|$ where $Z \sim \mathcal{N}(0,1)$, check that $\mathcal{L}^{X|Y}(y,\cdot) = \frac{1}{2}\delta_y + \frac{1}{2}\delta_{-y}$.

6. Martingales in discrete time

Definition 6.1. *A **filtration** on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a sequence $(\mathcal{F}_n)_{n\geq 0}$ of sub-σ-algebras of $\mathcal{F}$, which is non-decreasing i.e. $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for each $n \geq 0$.*

$(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_n)_{n\geq 0})$ is then called a filtered probability space. We fix one in the below.
(example : if (X_n) is a sequence of r.v., we can take $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$.)

Definition 6.2. *A stochastic process $(X_n)_{n\geq 0}$ (i.e. each X_n is a measurable r.v.) is **adapted** if, for each n, X_n is $\mathcal{F}_n$-measurable.*

Definition 6.3. *A **martingale** (resp. submartingale, supermartingale) is an adapted process $(M_n)_{n\geq 0}$ s.t.*
(1) $\forall n \geq 0$, $M_n \in L^1(\mathbb{P})$,
(2) $\forall n \geq 0$, $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ (resp. $\geq$, $\leq$).

Note that this implies $\mathbb{E}[M_n|\mathcal{F}_m]$ for each $n \geq m$, and $\mathbb{E}[M_n] = \mathbb{E}[M_0]$.
Example : simple random walk. $M_n = \mathbb{E}[Z|\mathcal{F}_n]$.

Definition 6.4. *A stopping time τ is a random variable taking values in $\mathbb{N} \cup \{\infty\}$ such that*

$$\forall n \geq 0, \quad \{\tau \leq n\} \in \mathcal{F}_n.$$

(exercise : check that this equivalent to the same definition with $\leq$ replaced by $=$).
Example : if (Y_n) is an adapted process, and A is a Borel subset of $\mathbb{R}$, then

$$\tau_A := \inf\{n \geq 0, \quad Y_n \in A\}$$

is a stopping time.

Given a stopping time τ and a process M, we let M^τ (the **stopped process**) be defined by $M^\tau_n = M_{n\wedge\tau}$.

Proposition 6.1. *If M is a martingale (resp. super, sub), then so is M^τ for any stopping time τ.*

Proof. First, note that

$$M^\tau_n = M_n 1_{\{n\leq\tau\}} + M_\tau 1_{\{\tau<n\}} = M_n 1_{\{n\leq\tau\}} + \sum_{k<n} M_k 1_{\{\tau=k\}}$$

is an adapted process, and integrable since $|M^\tau_n \ \leq \sum_{k=0}^n |M_k|$.

We then write

$$M^\tau_{n+1} = M_{n+1} 1_{\{n+1\leq\tau\}} + M_\tau 1_{\{\tau<n+1\}} = M_{n+1} 1_{\{n<\tau\}} + M_\tau 1_{\{\tau\leq n\}}$$

and taking conditional expectation,

$$\mathbb{E}[M^\tau_{n+1}|\mathcal{F}_n] = \mathbb{E}[M_{n+1}|\mathcal{F}_n] 1_{\{n<\tau\}} + M_\tau 1_{\{\tau\leq n\}} = M_n 1_{\{n<\tau\}} + M_\tau 1_{\{\tau\leq n\}} = M^\tau_n.$$

□

Corollary 6.1. *If τ is a bounded stopping time, and M is a martingale, then $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$.*

Exercise 3. *(1) Show that the above also holds if we replace the boundedness assumption on τ by*

$$\mathbb{E}[\tau] < +\infty, \quad \text{and } \exists K \in \mathbb{R}, \forall n \in \mathbb{N}, |M_{n+1} - M_n| \leq K \, a.s.$$

(Hint : apply dominated convergence to $M_{n\wedge\tau}$).

(2) Find a martingale M and a stopping time $\tau < \infty$ a.s., and such that $\mathbb{E}[M_\tau] \neq \mathbb{E}[M_0]$. (Hint : consider the first time when a simple random walk hits 1).

Definition 6.5. *Given a stopping time τ, we define*

$$\mathcal{F}_\tau = \{A \in \mathcal{F} \, : \, \forall n \geq 0, \; A \cap \{\tau = n\} \in \mathcal{F}_n\}.$$

It is easy to check that this defines a σ-algebra. Also note that this notation is consistent with $\mathcal{F}_n$ (i.e. if $\forall \omega$, $\tau(\omega) = n$ then $\mathcal{F}_\tau = \mathcal{F}_n$).

Check that if (X_n) is an adapted process, then $X_\tau 1_{\{\tau<\infty\}}$ is $\mathcal{F}_\tau$-measurable. This follows from

$$\left\{X_\tau 1_{\{\tau<\infty\}} \in A, \, \tau = n\right\} = \{X_n \in A, \tau = n\}.$$

If $\rho \leq \tau$ are two stopping times, then $\mathcal{F}_\rho \subset \mathcal{F}_\tau$, since

$$A \cap \{\, \tau \leq n\} = A \cap \{\, \rho \leq n\} \cap \{\tau \leq n\}.$$

We then have the following generalization of Corollary 6.1.

Proposition 6.2. *Let $\rho \leq \tau$ be two bounded stopping times, and M a martingale. Then*

$$\mathbb{E}[M_\tau | \mathcal{F}_\rho] = M_\rho.$$

Proof. First $\mathbb{E}[M_\tau|\mathcal{F}_n] = M_{\tau\wedge n}$ by Proposition 6.1, and then, for any $A \in \mathcal{F}_\rho$,

$$\begin{aligned}
\mathbb{E}[M_\tau 1_A] &= \sum_n \mathbb{E}[M_\tau 1_A 1_{\rho=n}] \\
&= \sum_n \mathbb{E}[\mathbb{E}[M_\tau|\mathcal{F}_n] 1_A 1_{\rho=n}] \\
&= \sum_n \mathbb{E}[M_n 1_A 1_{\rho=n}] \\
&= \mathbb{E}[M_\rho 1_A].
\end{aligned}$$

(We have used that $1_A 1_{\rho=n}$ is $\mathcal{F}_n$-measurable by definition of $\mathcal{F}_\rho$) □

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

Theorem 6.1. *Let $(M_n)_{n \geq 0}$ 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_∞.*

Note the convergence does not hold in L^1 in general. Indeed, let $M_n = U_1 \dots U_n$, where the U_n are i.i.d. with $\mathbb{P}(U_1 = 0) = \mathbb{P}(U_1 = 2) = \frac{1}{2}$. Then $M_n \to 0$ a.s., but $\mathbb{E}[M_n] = 1$ for all n. Note in particular that in that case $M_n \neq \mathbb{E}[M_\infty \, |\mathcal{F}_n]$.

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

Definition 6.6. *A family X_i, $i \in I$ of r.v.'s is **uniformly integrable (u.i.)** if*

$$\lim_{K\to\infty} \sup_{i\in I} \mathbb{E}\left[|X_i| \, 1_{\{|X_i|\geq K\}}\right] = 0.$$

Note that any u.i. family is bounded in L^1.

For instance, if $\forall i$, $|X_i| \leq |Y|$ with Y integrable, then the X_i are uniformly integrable (exercise).

A convenient way to check the uniform integrability is via De La Vallée Poussin's criterion, which states that a family is u.i. if and only if there exists Φ with $\lim_{|x|\to\infty} \frac{\Phi(x)}{|x|} = +\infty$ and $\sup_i \mathbb{E}[\Phi(X_i)] < \infty$. (Exercise : prove the "if" part, which is both the simplest and the most useful direction.) For instance, families which are bounded in $L^p, p > 1$ are uniformly integrable.

The importance of the u.i. condition, is the following result, which is an extension of the dominated convergence theorem.

Proposition 6.3. *Let $X_n \to X$ a.s., and assume that the $X_n, n \in \mathbb{N}$ are uniformly integrable. Then $\lim_n \mathbb{E}[X_n] = \mathbb{E}[X]$.*

Proof. Fix $\varepsilon > 0$, by definition there exists K s.t., letting $X_n^K = X_n 1_{\{|X_n| \leq K\}}$, it holds that

$$\sup_n \left|\mathbb{E}\left[X_n\right] - \mathbb{E}\left[X_n^K\right]\right| \leq \varepsilon.$$

On the other hand, by dominated convergence, $\lim_n \mathbb{E}[X_n^K] = \mathbb{E}[X^K]$, and we deduce that

$$\limsup_n |\mathbb{E}[X_n] - \mathbb{E}[X]| \leq 2\varepsilon.$$

We conclude by letting $\varepsilon \to 0$. □

Theorem 6.2. *Let $(M_n)_{n \geq 0}$ be a martingale which is uniformly integrable. Then M_n converges a.s. and in L^1 to a limit M_∞. In addition, it holds that $M_n = \mathbb{E}[M_\infty \mid \mathcal{F}_n]$.*

Proof. Let M_∞ be the a.s. limit given by Theorem 6.1. Then L^1 convergence follows from Proposition 6.3 applied to $X_n = |M_n - M_\infty|$. To check that $M_n = \mathbb{E}[M_\infty \mid \mathcal{F}_n]$, it suffices to check that for any $A \in \mathcal{F}_n$,

$$\mathbb{E}\left[M_n 1_A\right] = \mathbb{E}\left[\lim_{m\to\infty} M_m 1_A\right],$$

which again follows from Proposition 6.3. □

Theorem 6.3. *Let $(M_n)_{n \geq 0}$ be a martingale which is bounded in L^p, $1 < p < \infty$. Then M_n converges a.s. and in L^p to a limit M_∞.*

Let us prove the L^2-convergence when $p = 2$. In that case, recalling that conditional expectations are orthogonal projections, it holds that

$$\forall m \geq n, \ \| M_m\|_2^2 = \|M_m - M_n\|_2^2 + \|M_n\|_2^2,$$

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) is a Cauchy sequence in L^2, from which we can conclude.

7. Gaussian vectors

In this section, we will work with random vectors, i.e. r.v.'s Y with 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 $Y = (Y_1, \ldots, Y_d)$, its characteristic function is defined by

$$\phi_Y : \xi = (\xi_1, \ldots, \xi_d) \in \mathbb{R}^d \mapsto \mathbb{E}\left[\ e^{i\xi\cdot Y}\right]$$

where $\cdot$ is the scalar product ($\xi \cdot X = \sum_{i=1}^d \xi_i X_i$).

Then it still holds that a sequence Y_n converges in distribution to Y if and only if its characteristic function converges pointwise. (In particular, if two r.v.'s share the same characteristic function, they have the same law).

Definition 7.1. *X , scalar r.v., is a Gaussian, if there exists $m \in \mathbb{R}$, $\sigma > 0$ s.t. X has the same law as $m + \sigma Z$, where $Z \sim \mathcal{N}(0,1)$. This is equivalent to the characteristic function satisfying $\Phi_X(t) = e^{mt - \frac{t^2\sigma^2}{2}}$. We write $X \sim \mathcal{N}(m, \sigma^2)$.*

Definition 7.2. *A random vector $Y = (Y^1, \ldots, Y^d)$ is **a Gaussian vector** (also written : $(Y^1, \ldots, Y^d)$ are **jointly Gaussian**), if, for each $(\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^d$, the scalar r.v. $\sum_{i=1}^d \lambda_i Y_i$ is Gaussian.*

Definition 7.3. *Given a Gaussian vector Y, we define its mean $m \in \mathbb{R}^d$ and its covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$ by*

$$m_i = \mathbb{E}[Y_i],$$
$$\Sigma_{ij} = cov(Y_i Y_j) = \mathbb{E}[(Y_i - \mathbb{E}[Y_i])(Y_j - \mathbb{E}[Y_j])].$$

Proposition 7.1. *The law of a Gaussian vector is characterized by its mean m and covariance Σ. More precisely, the characteristic function of Y is then given for $\xi \in \mathbb{R}^d$ by*

$$\mathbb{E}[\exp(i\xi \cdot Y)] = \exp\left(m \cdot \xi - \frac{1}{2}\xi \cdot \Sigma\xi\right).$$

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

In addition, if $\Sigma = CC^T$ for a matrix C, then

$$Y =^{law} m + \Sigma X,$$

where $X = (X_1, \ldots, X_d)$ with the X_i i.i.d. $\mathcal{N}(0,1)$.

Proof. For the first part : by assumption, if ξ is in $\mathbb{R}^d$, $\xi \cdot Y$ is Gaussian, and it is immediate to compute

$$\mathbb{E}[\xi \cdot Y] = \xi \cdot m, \quad Var(\xi \cdot Y) = \sum_{i,j} \xi_i \xi_j \Sigma_{ij} = \xi \cdot \Sigma\xi.$$

The formula then follows from that for the scalar Gaussians.

For the second part, it suffices to check that the r.h.s. is also a Gaussian vector, with same mean and covariance matrix. ☐

Remark : this gives a way to simulate any Gaussian vector. (and there always exists such a C, which can be taken triangular : this is the so-called Cholesky decomposition of symmetric matrices).

Corollary 7.1. *Let $(Y^1, \ldots, Y^d)$ be jointly Gaussian. Then $Y^1, \ldots, Y^d$ are independent if and only if they are pairwise uncorrelated.*

More generally : if $X \sim \mathcal{N}(m, \Sigma)$, then for any matrix $A \in \mathbb{R}^{d' \times d}$, $AX \sim \mathcal{N}(Am, A\Sigma A^T)$.

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

Proposition 7.2. *Let $Z = (X, Y)$ be a Gaussian vector in $\mathbb{R}^2$, with mean vector $\begin{pmatrix} m_X \\ m_Y \end{pmatrix}$ and covariance matrix $\begin{pmatrix} \sigma_{XX} & \sigma_{XY} \\ \sigma_{XY} & \sigma_{YY} \end{pmatrix}$, and assume that $\sigma_{YY} > 0$.*

Then it holds that

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

More generally, conditionally on Y, X is Gaussian with mean $\mathbb{E}[X|Y]$ and variance $\sigma_{XX} - \frac{\sigma_{XY}}{\sigma_{YY}}$.

Proof. Let
$$W = X - \frac{\sigma_{XY}}{\sigma_{YY}} Y.$$
Then a direct computation gives that $cov(W, Y) = 0$, and since (W, Y) is Gaussian, W is independent of Y. The result then follows from writing $X = W + \frac{\sigma_{XY}}{\sigma_{YY}} Y$, with the first term of the r.h.s. is independent of Y and the second 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 7.1 (CLT in $\mathbb{R}^d$.)**.** *Let X^n be an i.i.d. sequence of random vectors, with square integrable entries. Let $M = \mathbb{E}[X^1]$, and $\Sigma = \left(cov(X_i^1, X_j^1)\right)_{1 \le i,j \le d}$. Then*
$$\frac{\sum_{k=1}^n (X_k - M)}{\sqrt{n}} \to_{n\to\infty} \mathcal{N}(0, \Sigma) \textit{ in law }.$$

Proof. Using the characteristic function, this reduces to the scalar CLT. □

8. Brownian motion : definition, existence

A continuous time stochastic process on a probability space is a family $X = (X_t)_{t \ge 0}$ of random variables indexed by $\mathbb{R}_+$.

Definition 8.1. *A (standard)* ***Brownian motion*** *is a stochastic process $(B_t)_{t \ge 0}$ s.t.*

(1a) $B_0 = 0$ a.s.

(1b) For each $0 = t_0 < t_1 < \ldots < t_m$, then $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}}$ are independent,

(1c) For each $0 \le s \le t$, $B_t - B_s$ has law $\mathcal{N}(0, t - s)$.

(2) $\forall \omega \in \Omega$, $t \mapsto B_t$ is continuous.

Remark 8.1. *(Technical remark on continuity of sample paths)*

- *Condition (2) is sometimes replaced by the weaker condition :*

  (2') There exists a measurable $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$, such that $t \mapsto B_t$ is continuous on Ω_0.

  Note that if B satisfies (1) and (2'), we can redefine the B_t to be 0 outside of Ω_0 to obtain a stochastic process satisfying (1)-(2).
- *In fact, even though (1) does not imply (2) or (2'), it implies that we can find a modification of B (i.e. a process $\tilde{B}$ such that for all $t \ge 0$, a.s., $B_t = \tilde{B}_t$) which is continuous.*

  (exercise : use for instance a similar construction to what is done below to find a sequence of continuous B^n's, which almost surely converge uniformly on compacts, and such that for all $t \ge 0$, a.s., $B_t = \lim_n B_t^n$.)

A stochastic process $(X_t)_{t \ge 0}$ is a **Gaussian process** if for any $t_1, \ldots, t_n$, $(X_{t_1}, \ldots, X_{t_n})$ is a Gaussian vector.

Proposition 8.1. *In the definition of Brownian motion, we can alternatively replace (1) by*

(1') B is a Gaussian process, with mean function $\mathbb{E}[B_t] = 0$ and covariance $\mathbb{E}[B_s B_t] = s \wedge t$

Proof. (1) ⇒ (1') : $(B_{t_1}, \ldots, B_{t_n})$ is a linear function of $(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}})$. Since the latter has independent Gaussian components, it is a Gaussian vector, and so is the former. The mean and covariance computation is immediate

(1') ⇒ (1) : (1a) is immediate. For (1b), note that $(B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_m} - B_{t_{m-1}})$ is a Gaussian vector, and it suffices to check that its entries are uncorrelated, which follows from the covariance function. For (1c), again $B_t - B_s$ is Gaussian, and it suffices to check that it has mean 0 and variance $t - s$. □

Proposition 8.2. *Let $B = (B_t)_{t \geq 0}$ be a Brownian motion. Then :*

(1) $(-B_t)_{t \geq 0}$ is a Brownian motion,
(2) For each $\lambda > 0$, $B^\lambda_t := \lambda^{-1/2} B_{\lambda t}$ is a Brownian motion.
(3) For each $T > 0$, $B^T_t := B_{T+t} - B_T$ is a Brownian motion (independent from $\sigma(B_s, s \leq T)$).
(4) $(tB_{\frac{1}{t}} 1_{\{t>0\}})_{t \geq 0}$ is a Brownian motion.

Proof. Straightforward. The only delicate point is to check that in case (4), $\tilde{B}(t) = tB_{\frac{1}{t}}$ is continuous at 0, which is equivalent to

$$\lim_{t \to \infty} \frac{B_t}{t} = 0. \tag{8.1}$$

Note that it follows from the strong law of large numbers that $\lim_n \frac{B_n}{n} = 0$. Exercise : prove (8.1), taking for granted that

$$\mathbb{E}\left[\sup_{t \in [0,1]} |B_t|\right] < \infty \tag{8.2}$$

(and using independence of increments). □

The fact that Brownian motions exist is not an obvious fact. We will now present a construction of Brownian motion on $[0, 1]$ due to Paul Lévy. Note that this suffices to define a Brownian motion on $\mathbb{R}_+$ by "pasting together" successive independent copies.

The construction proceeds by successive approximations, defining the value of B at dyadic points. Let

$$\mathcal{D} = \cup_n \mathcal{D}_n, \quad \mathcal{D}_n = \left\{k2^{-n}, \quad 0 \leq k \leq 2^n\right\}.$$

We then fix $(Z_t)_{t \in \mathcal{D}}$, an i.i.d. family of $\mathcal{N}(0, 1)$, and define for any $n \geq 0$, the functions F_n on $\mathcal{D}$ by $F_0(0) = 0, F_0(1) = Z_1$, and for $n \geq 1$,

$$F_n(t) = \begin{cases} 2^{-\frac{n+1}{2}} Z_t, & t \in \mathcal{D}_n \setminus \mathcal{D}_{n-1}, \\ 0, & t \in \mathcal{D}_{n-1}, \end{cases}$$

which are then extended to the whole of $[0, 1]$ by linear interpolation. We also let

$$B_n = \sum_{k=0}^{n} F_k,$$

and aim to show that the B_n (which are continuous functions by definition), almost surely converge uniformly on $[0, 1]$ as $n \to \infty$.

We then use the following property of Gaussians : if $Z \sim \mathcal{N}(0, 1)$, then for all $x \geq 1$, $\mathbb{P}(|Z| \geq x) \leq e^{-x^2}$.

It follows that for any $c > 0$,

$$\begin{aligned}\mathbb{P}\left(\|F_n\|_\infty \geq c\right) &= \mathbb{P}\left(\exists t \in \mathcal{D}_n - \mathcal{D}_{n-1},\ \ 2^{-\frac{n+1}{2}}|Z_t| \geq c\right) \\ &\leq \sum_{t \in \mathcal{D}_n - \mathcal{D}_{n-1}} \mathbb{P}(|Z_t| \geq c) \quad = \quad 2^{n-1}\mathbb{P}(|Z| \geq c2^{\frac{n+1}{2}}) \\ &\leq 2^{n-1}\exp\left(-c^2 2^{-n-1}\right).\end{aligned}$$

Hence, taking $c_n = 2^{-\alpha n}$ (for fixed $\alpha < 1/2$)), it holds that $\sum_{n\geq 0}\mathbb{P}\left(\|F_n\|_\infty \geq c_n\right) < \infty$. By the Borel-Cantelli lemma, this implies that a.s., for n large enough, $\|F_n\|_\infty \leq c_n$ and in particular, a.s. it holds that $\sum_{n\geq 0}\|F_n\|_\infty < \infty$, so that B_n converges uniformly to a limit.

It remains to show that B is a Brownian motion. This is done by first checking that $(B(t))_{t\in\mathcal{D}}$ is a Gaussian process with the correct mean and covariance (this is easily checked on $\mathcal{D}_n$ by an induction on n). The extension to the whole interval is then a consequence of the following lemma.

Lemma 8.1. *Let X_n be a sequence of Gaussian vectors with mean M_n and covariance Σ_n, and assume that M_n and Σ_n converge to limits M and Σ as $n \to \infty$. Then X_n converges in distribution to $\mathcal{N}(M, \Sigma)$ as $n \to \infty$.*

Proof. This is obvious from the characteristic function. □

Exercise 4. *Show that there exists constant $C > 0$ s.t. $\mathbb{P}\left(\sup_{t\in[0,1}|B_t| \geq x\right) \leq Ce^{-Cx^2}$. (Hint : use the previous construction and write for any $\lambda > 0$, $\mathbb{E}\left[e^{\lambda\|B\|_\infty}\right] \leq \Pi_{n\geq 0}\mathbb{E}\left[e^{\lambda\|F_n\|_\infty}\right]$.) Note that this implies* (8.2)*.*

Remark 8.2. *Wiener measure (the law of Brownian motion) Let $E = C(\mathbb{R}_+, \mathbb{R})$ (path-space), and $\mathcal{E}$ the Borel σ-algebra for the topology of uniform convergence on compacts.*

Then, given a Brownian motion B on $(\Omega, \mathcal{F}, \mathbb{P})$, it can be shown that letting

$$\mathbb{W} : A \in \mathcal{E} \mapsto \mathbb{P}((t \mapsto B_t) \in A)$$

defines a unique measure on $(E, \mathcal{E})$. This measure is called the ***Wiener measure****.*

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

Proposition 8.3 (Blumenthal's $0-1$ law)**.** *Let $\mathcal{F}_t = \sigma\left(B_s; 0 \leq s \leq t\right)$, and let $\mathcal{F}_{0+} = \cap_{s\geq 0}\mathcal{F}_s$. Then $\mathcal{F}_{0+}$ is trivial in the sense that*

$$A \in \mathcal{F}_{0+} \ \Rightarrow\ \mathbb{P}(A) = 0 \text{ or } 1.$$

Proof. We show that $\mathcal{F}_{0+}$ is independent from $\mathcal{F}_t$, for any $t > 0$. Indeed, let f be a continuous function and $A \in \mathcal{F}_{0+}$, then for any $t_1, \ldots, t_n$

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

where in the second equality we have used that $A \in \mathcal{F}_\varepsilon$ and independence of increments.

Hence, if $A \in \mathcal{F}_{0+}$, then A is independent of itself, and $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$, which implies $\mathbb{P}(A) \in \{0, 1\}$. □

Remark 8.3 (Kolmogorov's $0-1$ law)**.** *By time inversion, this also implies that the tail σ-algebra $\cap_{t\geq 0}\sigma\left(B_s; s \geq t\right)$ is trivial.*

Corollary 8.1. *It holds almost surely, that*

$$\forall t_0 > 0, \quad \sup_{t\in[0,t_0]} B_t > 0 \text{ and } \inf_{t\in[0,t_0]} B_t < 0.$$

Proof. Let $t_n \downarrow 0$, $A_n = \{\sup_{[0,t_n]} B > 0\}$ and $A = \cap_n A_n$. Then, since A_n is decreasing (for set inclusion), $A \in \mathcal{F}_{0+}$ and $\mathbb{P}(A) = \lim_n \mathbb{P}(A_n)$. On the other hand, for each n, $\mathbb{P}(A_n) \geq \mathbb{P}(B_{t_n} > 0) = \frac{1}{2}$ since B_{t_n} is a centered Gaussian. It follows that $\mathbb{P}(A) \geq \frac{1}{2}$, and by the $0-1$ law, this implies $\mathbb{P}(A) = 1$.

The statement on the infimum also holds a.s. by symmetry. □

Corollary 8.2. *It holds almost surely that*

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

Proof. Fix $a > 0$. By scaling (point (2) in Proposition 8.2), it holds that, for any $T > 0$,

$$\mathbb{P}\left(\sup_{t\in[0,+\infty)} B_t > a\right) \geq \mathbb{P}\left(\sup_{t\in[0,T]} B_t > a\right) = \mathbb{P}\left(\sup_{t\in[0,1]} B_t > \frac{a}{\sqrt{T}}\right).$$

Letting $T \to \infty$, and using that $\sup_{t\in[0,1]} B_t > 0$ a.s. by the previous corollary, we obtain $\limsup_{t\to+\infty} B_t > a$ a.s., for arbitrary a. □

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