MDI115 – Part II:
Convergence of probability measures

Lecture Notes

P. Bianchi

February 8, 2024

Contents

Foreword

Chapter 1

Construction of probability measures

1.1 Definitions

Definition 1.1. Let $\mathcal{X}$ be an arbitrary set. A class $\mathcal{F}$ of subsets of $\mathcal{X}$ is called a *field* if the following conditions are satisfied:

i) $\mathcal{X} \in \mathcal{F}$

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

iii) $A, B \in \mathcal{F}$ implies $A \cup B \in \mathcal{F}$.

It is called a *σ-field* if, in addition, the following condition holds:

iv) $A_1, A_2, \cdots \in \mathcal{F}$ implies $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$.

Example 1.2. On $\mathcal{X} = (0, 1]$, let $\mathscr{B}_0$ the class formed by finite unions of intervals of the form $(a, b]$, augmented by the empty set. That is, each element A of $\mathscr{B}_0$ has the form:

$$A = (a_1, b_1] \cup \cdots \cup (a_n, b_n] \tag{1.1}$$

for some integer n. Then, $\mathscr{B}_0$ is a field on $\mathcal{X}$. It is not a σ-field, though.

If $\mathcal{F}_0$ is a class of subsets of $\mathcal{X}$, we denote by $\sigma(\mathcal{F}_0)$ the σ-field generated by $\mathcal{F}_0$, that is, the smallest σ-field which contains $\mathcal{F}_0$.

Definition 1.3. If $\mathcal{X}$ is a topological space, we denote by $\mathcal{B}(\mathcal{X})$ the *Borel* σ-field on $\mathcal{X}$, that is, the σ-field generated by the class of all open sets.

Example 1.4. The Borel σ-field on $\mathcal{X} = (0, 1]$ is generated by open subsets of $(0, 1]$, that is, sets of the form $U \cap (0, 1]$, where U is an open subset of $\mathbb{R}$. Alternatively, it is also the σ-field generated by the class of intervals of the form $(a, b]$. Also, it is the σ-field generated by the field $\mathscr{B}_0$ defined in Example 1.2.

A *set function* is a real-valued function defined on some class of subsets of Ω.

Definition 1.5. A set function P on a field $\mathcal{F}_0$ is a *probability measure* if it satisfies the following conditions:

i) $0 \le P(A) \le 1$ for all $A \in \mathcal{F}_0$

ii) $P(\emptyset) = 0$, $P(\mathcal{X}) = 1$

iii) If $A_1, A_2, \dots \in \mathcal{F}_0$ is a disjoint sequence of $\mathcal{F}_0$-sets, and if $\cup_n A_n \in \mathcal{F}_0$, then

$$P\Big(\bigcup_n A_n\Big) = \sum_n P(A_n).$$

The last condition is called *countable additivity.* In this last condition, note that the assumption that $\cup_n A_n \in \mathcal{F}_0$ is needed. Indeed, $\mathcal{F}_0$ is supposed to be a field and not necessarily a σ-field. Thus, it is not guaranteed in general that $\cup_n A_n \in \mathcal{F}_0$.

Theorem 1.6. *Let P be a probability measure on a field $\mathcal{F}_0$.*

i) $A \subset B$ *implies* $P(A) \le P(B)$

ii) $P(A^c) = 1 - P(A)$

iii) $A_n \uparrow A$ *implies* $P(A_n) \uparrow P(A)$

iv) $A_n \downarrow A$ *implies* $P(A_n) \downarrow P(A)$

v) (countable subadditivity) $P\big(\bigcup_n A_n\big) \le \sum_n P(A_n)$.

Proof. Standard, see MDI104 textbook. □

Example 1.7. Define the map $\lambda : \mathscr{B}_0 \to [0,1]$ given, for every A of the form (1.1),

$$\lambda(A) = \sum_{i=1}^{n} b_i - a_i \,. \tag{1.2}$$

Of course, for a given $A \in \mathscr{B}_0$, the decomposition (1.1) of A as a finite union of intervals may not be unique, but it is a simple exercise to establish that the sum in the righthand side of Eq. (1.2) does not depend on the specific decomposition. Hence, λ is well defined. You can check that:

$$\lambda \text{ is a probability measure on the field } \mathscr{B}_0.$$

The main goal is now to extend λ to a probability measure on the generated σ-field $\sigma(\mathscr{B}_0) = \mathcal{B}((0,1])$.

1.2 Existence and extension

The main theorem to be proved is the following.

Theorem 1.8. *A probability measure on a field has a unique extension to the generated σ-field.*

The uniqueness is a consequence of the π-λ-theorem, as seen in MDI104, which asserts that when two probability measures coincide on a π-system generating the σ-field, they are equal: the result applies here because a field is a π-system. The main point here is to establish the existence.
Before proving the existence result, note that Theorem 1.8 along with Example 1.7 shows the existence of the Lebesgue measure on $\mathcal{X} = (0, 1]$.

Corollary 1.9. *There exists a unique probability measure λ on $\mathcal{B}((0,1])$ such that $\lambda((a,b]) = b - a$ for all $0 < a < b < 1$.*

Proof. The set function λ given by (1.2) is a probability measure on the field $\mathscr{B}_0$. Therefore, it admits a unique extension on $\sigma(\mathscr{B}_0) = \mathcal{B}((0,1])$. □

Let us now pass to the proof of Theorem 1.8.

Definition 1.10. An *outer measure* is a set function $P^* : 2^{\mathcal{X}} \to [0, \infty]$ such that:

i) $P^*(\emptyset) = 0$

ii) $A \subset B$ implies $P^*(A) \le P^*(B)$

iii) $P^*\big(\bigcup_n A_n\big) \le \sum_n P^*(A_n)$.

Proposition 1.11. *Let P be a probability measure on a field $\mathcal{F}_0$. The function P^* defined for all $A \subset \mathcal{X}$ by:*

$$P^*(A) = \inf\{\sum_n P(A_n) \,:\, A_1, A_2, \dots \in \mathcal{F}_0,\ A \subset \bigcup_n A_n\} \tag{1.3}$$

is an outer measure.

Proof. The other being obvious, only iii) needs proof. Set $\epsilon > 0$. For each n, choose a $\mathcal{F}_0$-cover $\bigcup_k B_{nk}$ of A_n such that $\sum_k P(B_{nk}) \le P^*(A_n) + \epsilon 2^{-n}$. Then, using that $\bigcup_{nk} B_{nk}$ is a $\mathcal{F}_0$-cover of $\bigcup_n A_n$,

$$P^*\big(\bigcup_n A_n\big) \le \sum_n \sum_k P(B_{nk}) \le \sum_n (P^*(A_n) + \epsilon 2^{-n}) = \sum_n P^*(A_n) + \epsilon/2\,.$$

Letting $\epsilon \to 0$ completes the proof. □

Intuitively, if the A_n form an efficient covering of A, in the sense that they do not overlap one another very much, or extend much beyond A, then $\sum_n P(A_n)$ should be a good approximation to the measure of A, whenever A is indeed to have a measure assigned. For such A, we also expect that $P^*(A) = 1 - P^*(A^c)$, and we may, by definition, call *P^*-measurable set*, any set A satisfying $P^*(A) + P^*(A^c) = 1$. For technical reasons, it is easier to use the following definition.

Definition 1.12. Let P^* be an outer measure. A set $A \subset \mathcal{X}$ is called P^*-measurable if for every set $E \subset \mathcal{X}$,

$$P^*(A \cap E) + P^*(A^c \cap E) = P^*(E)\,. \tag{1.4}$$

Proposition 1.13. *Let P^* be an outer measure. The class $\mathscr{M}$ of P^*-measurable sets is a σ-field, and the restriction of P^* to $\mathscr{M}$ is countably additive.*

Proof. First we establish that $\mathscr{M}$ is a field. Let $A, B \in \mathscr{M}$. For any set $E \subset \mathcal{X}$,

$$\begin{aligned} P^*(E) &= P^*(B \cap E) + P^*(B^c \cap E) \\ &\overset{(a)}{=} P^*(A \cap B \cap E) + P^*(A^c \cap B \cap E) + P^*(A \cap B^c \cap E) + P^*(A^c \cap B^c \cap E) \\ &\overset{(b)}{\geq} P^*(A \cap B \cap E) + P^*\left((A^c \cap B \cap E) \cup (A \cap B^c \cap E) \cup (A^c \cap B^c \cap E)\right) \\ &= P^*(A \cap B \cap E) + P^*\left((A \cap B)^c \cap E\right)\,, \end{aligned}$$

where in (a), we used the definition of measurability in Eq. (1.4), with $B \cap E$ instead of E, and where in (b), we used the subadditivity condition iii) of Definition 1.10. Moreover, by subadditivity again, $P^*(E) \leq P^*(A \cap B \cap E) + P^*\left((A \cap B)^c \cap E\right)$, hence an equality. This proves that $\mathscr{M}$ is a field.
Second, we establish that P^* is additive on $\mathscr{M}$. Consider $A, B, \cdots \in \mathscr{M}$ being disjoint. Note that $A \cup B \in \mathscr{M}$, because $\mathscr{M}$ is a field. Therefore,

$$\begin{aligned} P^*(A \cup B) &= P^*((A \cup B) \cap A) + P^*((A \cup B) \cap A^c) \\ &= P^*(A) + P^*(B \backslash A) \\ &= P^*(A) + P^*(B)\,, \end{aligned}$$

because A, B are disjoint. By induction, for any disjoint $A_1, A_2, \cdots \in \mathscr{M}$,

$$\forall n,\ P^*\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P^*(A_k)\,.$$

By condition ii) in Definition 1.10, $P^*(\bigcup_{k=1}^{\infty} A_k) \geq \sum_{k=1}^{n} P^*(A_k)$. Letting n goes to infinity,

$$P^*(\bigcup_{k=1}^{\infty} A_k) \geq \sum_{k=1}^{\infty} P^*(A_k)\,.$$

The reverse inequality hold by countable subadditivity of P^*, that is, condition iii) in Definition 1.10. Therefore,

$$P^*(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P^*(A_k)\,.$$

This proves that P^* is countably additive on $\mathscr{M}$. By a marginal adaptation of the proof, one can slightly generalize the above equality by:

$$P^*(E \cap \bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P^*(E \cap A_k)\,, \tag{1.5}$$

for any set E. This will be revealed useful to establish the last point, which is that $\mathscr{M}$ is a σ-field. Let us prove this. Consider $A_1, A_2, \dots \in \mathscr{M}$, and at first, assume that these sets are disjoints. For every n, $\bigcup_{k \le n} A_k \in \mathscr{M}$, because $\mathscr{M}$ is a field. Thus, for every E,

$$\begin{aligned} P^*(E) &= P^*(E \cap \bigcup_{k \le n} A_k) + P^*(E \cap (\bigcup_{k \le n} A_k)^c) \\ &\ge P^*(E \cap \bigcup_{k \le n} A_k) + P^*(E \cap (\bigcup_{k=1}^{\infty} A_k)^c) \\ &\overset{(c)}{=} \sum_{k=1}^{n} P^*(E \cap A_k) + P^*(E \cap (\bigcup_{k=1}^{\infty} A_k)^c)\,, \end{aligned}$$

where we used Eq. (1.5) to obtain (c). Letting $n \to \infty$,

$$\begin{aligned} P^*(E) &\ge \sum_{k=1}^{\infty} P^*(E \cap A_k) + P^*(E \cap (\bigcup_{k=1}^{\infty} A_k)^c) \\ &\overset{(d)}{=} P^*(E \cap \bigcup_{k=1}^{\infty} A_k) + P^*(E \cap (\bigcup_{k=1}^{\infty} A_k)^c) \end{aligned}$$

where we again used Eq. (1.5) to obtain (d). Therefore, $\bigcup_{k=1}^{\infty} A_k \in \mathscr{M}$. It remains to establish the same, but for sets $B_1, B_2, \dots \in \mathscr{M}$ that are not disjoint. One can define disjoint sets $A_1, A_2, \dots$ by $A_1 = B_1$, $A_2 = B_2 \backslash B_1$, etc. As $\mathscr{M}$ is a field, all A_n are in $\mathscr{M}$. Moreover, $\bigcup_n B_n = \bigcup_n A_n \in \mathscr{M}$. Thus, $\mathscr{M}$ is a σ-field. □

Proof of Theorem 1.8. Let P be a probability measure on a field $\mathcal{F}_0$. By Proposition 1.11, the set function P^* given by (1.3) is an outer measure. Let $\mathscr{M}$ denote the class of P^*-measurable sets. We establish the following.

1. $\mathcal{F}_0 \subset \mathscr{M}$. Indeed, consider $A \in \mathcal{F}_0$. Let $E \subset \mathcal{X}$ be any set. Set $\epsilon > 0$. Choose $\mathcal{F}_0$-sets A_n such that $\sum_n P(A_n) \le P^*(E) + \epsilon$. Then,

$$P^*(E \cap A) \le P(\bigcup_n A_n \cap A) \le \sum_n P(A_n \cap A)$$

where we used the fact that $A_n \cap A \in \mathcal{F}_0$, along with the countable subadditivity of the measure P. Similarly, $P^*(E \cap A^c) \le \sum_n P(A_n \cap A^c)$. Adding these inequalities,

$$\begin{aligned} P^*(E \cap A) + P^*(E \cap A^c) &\le \sum_n P(A_n \cap A) + \sum_n P(A_n \cap A^c) \\ &\le \sum_n P(A_n) \le P^*(E) + \epsilon\,. \end{aligned}$$

Letting ϵ go to zero, we observe that $P^*(E \cap A) + P^*(E \cap A^c) \le P^*(E)$, and the reverse inequality holds by subadditivity of P^*. Thus, $A \in \mathscr{M}$. We have shown that $\mathcal{F}_0 \subset \mathscr{M}$. As $\mathscr{M}$ is a σ-field, this in turn implies that $\sigma(\mathcal{F}_0) \subset \mathscr{M}$.

2. For every $A \in \mathcal{F}_0$, $P(A) = P^*(A)$. To establish this, consider a cover $A \subset \bigcup_n A_n$ of $\mathcal{F}_0$-sets. One has $P(A) \leq P(\bigcup_n A_n) \leq \sum_n P(A_n)$. Taking the infemum w.r.t. all covers, we obtain that $P(A) \leq P^*(A)$. On the otherhand, it is obvious from Definition 1.10 that $P^*(A) \leq P(A)$ (because A is trivially a cover of A). Hence, P and P^* coincide on $\mathcal{F}_0$. In particular $P^*(\mathcal{X}) = P(\mathcal{X}) = 1$

The restriction of P^* to $\sigma(\mathcal{F}_0)$ is a probability measure. Indeed, it is countably additive by Proposition 1.13, and is such that $P^*(\mathcal{X}) = 1$ by the second point above. Moreover, it coincides with P on $\mathcal{F}_0$. The proof is complete.

Chapter 2

Convergence of probability measures

A complete separable metric space is called a *Polish space*. In this chapter, $\mathcal{X}$ is assumed Polish, equipped with a distance d.
We define the distance of a point x to a set $A \subset \mathcal{X}$ as

$$d(x, A) := \inf\{d(x, y) : y \in A\}\,,$$

with the convention that $\inf \emptyset_{\mathbb{R}} = +\infty$. We denote by $B(x, r)$ the ball of center x and radius r, in $\mathcal{X}$.

2.1 The space of probability measures

We denote by $\mathcal{P}(\mathcal{X})$ the set of Borel probability measures on $\mathcal{X}$, that is, the set of probability measures defined on the Borel σ-field $\mathcal{B}(\mathcal{X})$.
We denote by $C(\mathcal{X})$, the set of continuous functions on $\mathcal{X} \to \mathbb{R}$. We denote by $C_b(\mathcal{X})$ the set of bounded continuous functions on $\mathcal{X} \to \mathbb{R}$. The set $C_b(\mathcal{X})$ is a normed vector space, equipped with the supremum norm: $\|f\| = \sup_{x\in\mathcal{X}} |f(x)|$.
The topological dual of $C_b(\mathcal{X})$ is denoted as $C_b(\mathcal{X})^*$.

2.1.1 Probability measures as linear operators

A probability measure $P \in \mathcal{P}(\mathcal{X})$ is a set function on $\mathcal{B}(\mathcal{X})$, but can as well be interpreted as an element of $C_b(\mathcal{X})^*$, namely, the map $f \mapsto \int f dP$. This is justified by Proposition 2.1 below, which states that a probability P is characterized by the values of $\int f dP$ where $f \in C_b(\mathcal{X})$.

Proposition 2.1. *The map $\psi : \mathcal{P}(\mathcal{X}) \to C_b(\mathcal{X})^*$ given by $\psi(P) : f \mapsto \int f dP$, is injective.*

Proof. The function $\psi(P)$ is well defined in the sense that $\psi(P)$ is an element of $C_b(\mathcal{X})^*$. Indeed, $\psi(P)$ is linear and continuous, by Lebesgue's dominated convergence theorem.
Consider two measure P, Q such that $\psi(P) = \psi(Q)$. Let F be a non empty closed subset of $\mathcal{X}$. Set $\epsilon > 0$. Define $f(x) = \max(0, 1 - d(x,F)/\epsilon)$. As $d(\,.\,,F)$ is continuous, f is continuous, and obviously bounded. Thus,

$$P(F) \leq \int f dP = \int f dQ \leq Q(F^\epsilon)\,,$$

where $F^\epsilon = \bigcup_{x\in F} \overline{B(x,\epsilon)}$. As $F^\epsilon \downarrow F$, $Q(F^\epsilon) \downarrow Q(F)$. Thus $P(F) \leq Q(F)$. Exchanging the roles of P and Q, $P(F) = Q(F)$. The two probabilities coincide on the closed sets, hence on the open sets. The class of open sets is a π-system which generates $\mathcal{B}(\mathcal{X})$. This proves that $P = Q$. □

We can now identify $\mathcal{P}(\mathcal{X})$ and $\psi(\mathcal{P}(\mathcal{X}))$, and write $\mathcal{P}(\mathcal{X})$ as a subset of $C_b(\mathcal{X})^*$, up to the injection ψ. That is, with slight notation abuse:

$$\mathcal{P}(\mathcal{X}) \subset C_b(\mathcal{X})^*\,.$$

In the same way, we shall abusively use the same notation P to designate the probability measure P acting on Borel sets, and the linear operator $\psi(P)$ acting on bounded continuous functions. For instance, the following notations represent the same value, and will be used interchangeably:

$$P(f),\ \langle P, f\rangle,\ \int f dP\,.$$

2.1.2 The weak⋆ topology

As a subset of $C_b(\mathcal{X})^*$, the space $\mathcal{P}(\mathcal{X})$ can be equipped with any of the induced topologies existing on $C_b(\mathcal{X})^*$: strong, weak or weak⋆. Our main interest in this course is the weak⋆ topology, which is the coarsest topology on $C_b(\mathcal{X})^*$ such that, for each $f \in C_b(\mathcal{X})^*$, the map $L \mapsto \langle L, f\rangle$ is continuous.
From now on, we endow $\mathcal{P}(\mathcal{X})$ with the topology induced by the weak⋆ topology on $C_b(\mathcal{X})^*$. This means that the open sets of $\mathcal{P}(\mathcal{X})$ are the sets of the form $U \cap \mathcal{P}(\mathcal{X})$ for U any weak⋆-open set of $C_b(\mathcal{X})^*$.
As a consequence, a neighborhood basis for a point $P \in \mathcal{P}(\mathcal{X})$ is formed by the sets

$$\{Q \in \mathcal{P}(\mathcal{X})\ :\ Q(f_i) \leq P(f_i) + \epsilon_i,\ \forall i = 1\ldots n\}\,,$$

where n, $f_1, \ldots, f_n \in C_b(\mathcal{X})$, $\epsilon_1, \ldots, \epsilon_n > 0$ are arbitrary. A sequence (P_n) on $\mathcal{P}(\mathcal{X})$ converges weak⋆ to $P \in \mathcal{P}(\mathcal{X})$ if:

$$\forall f \in C_b(\mathcal{X}),\ \langle P_n, f\rangle \to \langle P, f\rangle\,. \tag{2.1}$$

The above condition (2.1) is a characterization of converging *sequences*. As a matter of fact, this condition fully characterizes the weak⋆ topology because,

as we shall see below in Section A.5, $\mathcal{P}(\mathcal{X})$ equipped with the weak$\star$ topology is metrizable (recall that, in metrizable spaces, the set of converging sequences characterizes the topology).
The fact that $\mathcal{P}(\mathcal{X})$ is metrizable is evident in the case where $\mathcal{X}$ is compact. In this case, $C_b(\mathcal{X}) = C(\mathcal{X})$ is separable, and therefore, the strong unit ball of its dual is weak$\star$ metrizable. Note that $\mathcal{P}(\mathcal{X})$ is a subset of the strong unit ball in $C_b(\mathcal{X})^*$, because, for any $P \in \mathcal{P}(\mathcal{X})$, the operator norm satisfies

$$\|P\|_* := \sup\{P(f) : f \in C_b(\mathcal{X}), \|f\| \leq 1\} = 1\,.$$

Therefore, $\mathcal{P}(\mathcal{X})$ endowed with the weak$\star$ topology is metrizable as a subset of the unit ball.
A less evident result is the fact that $\mathcal{P}(\mathcal{X})$ endowed with the weak$\star$ topology is metrizable, eventhough $\mathcal{X}$ is not compact. This point will be addressed in Section A.5. As a consequence, the characterization of weak$\star$ converging sequences in (2.1) can be taken as an alternative *definition* of the topology on $\mathcal{P}(\mathcal{X})$.

In the sequel, the notation $P_n \to P$ will always refer to a weak$\star$ limit. Similarly, when we use words such as “limit, compact, closure” for sequences or sets of probability measures, we mean weak$\star$ limit, weak$\star$ compact, weak$\star$ closure in $\mathcal{P}(\mathcal{X})$, unless we state it otherwise.

Example 2.2. Consider a sequence x_n on $\mathcal{X}$, which converges to x. For any $f \in C_b(\mathcal{X})$, $\delta_{x_n}(f) = f(x_n)$ converges to $f(x) = \delta_x(f)$ because f is continuous. This shows that $\delta_{x_n} \to \delta_x$. Yet, for some sets A, $\delta_{x_n}(A)$ may not converge to $\delta_x(A)$. For instance, unless x_n is identically equal to x for large n, $\delta_{x_n}(\{x\})$ does not converge to $\delta_x(\{x\}) = 1$. This raises the question: for which sets A does $P_n \to P$ implies $P_n(A) \to P(A)$? The answer is provided below.

Remark 2.1. In most books, the term *weak convergence* is generally used instead of weak$\star$ convergence. This terminology is bit confusing, because it has nothing to do with a convergence in a weak topology. The term “weak” here refers to the fact that $P_n(A) \to P(A)$ is not required to holds for *all* Borel sets A, but only to P-continuity sets, as will be clear from Theorem 2.3 below. To avoid the confusion, other other use the terminology of *narrow convergence*.

2.1.3 Portmanteau theorem

Given a Borel probability P, a set A is called a P-continuity set if $P(\bar{A} \backslash \mathring{A}) = 0$. Given a real function f on $\mathcal{X}$, A is called a f-continuity set if f is continuous at each point of A.

Theorem 2.3. *The following conditions are equivalent.*

i) $P_n \to P$

ii) $P_n(f) \to P(f)$ *for all bounded Lipschitz continuous* f

iii) $\limsup_n P_n(F) \leq P(F)$ *for all closed* F

iv) $\liminf_n P_n(U) \geq P(U)$ *for all open* U

v) $P_n(A) \to P(A)$ *for all* P*-continuity set* A.

Proof. The implication $i \Rightarrow ii$ is trivial.
We prove $ii \Rightarrow iii$. Assume that F is non empty, otherwise there is nothing to prove. The map $f(x) = \max(0, 1 - d(x, F)/\epsilon)$ is bounded and 1-Lipschitz continuous, because the function $\max(0, 1 - \cdot\,)$ is 1-Lispchitz continuous on $\mathbb{R}$, and the distance function $d(\,\cdot\,, F)$ is Lipschitz continuous as a consequence of the triangular inequality. Thus, $P_n(f) \to P(f)$. Therefore, considering the ϵ-enlargement $F^\epsilon = \bigcup_{x\in F} B(x, \epsilon)$, we obtain:

$$P(F^\epsilon) \geq P(f) = \lim_n P_n(f) \geq \limsup_n P_n(F)\,.$$

Letting $\epsilon \downarrow 0$ completes the proofs, because $F^\epsilon \downarrow F$, thus $P(F^\epsilon) \downarrow P(F)$.
The proof that $iii \Leftrightarrow iv$ follows easily by complementation.
We prove that $iii\&iv \Rightarrow v$. As $P(\bar{A}\backslash\mathring{A}) = 0$,

$$\begin{aligned} P(A) = P(A^{\mathrm{o}}) &\leq \liminf_n P_n(A^{\mathrm{o}}) \\ &\leq \liminf_n P_n(A) \leq \limsup_n P_n(A) \\ &\leq \limsup_n P_n(\overline{A}) \leq P(\overline{A}) = P(A)\,. \end{aligned}$$

Therefore, the inequalities are in fact equalities. In particular, $\liminf_n P_n(A) = \limsup_n P_n(A) = P(A)$. Hence, $P_n(A) \to P(A)$.
We prove that $v \Rightarrow i$. Consider $f \in C_b(\mathcal{X})$, and assume without restriction that $0 \leq f \leq 1$. By Fubini's theorem, $P(f) = \int_0^1 P(\{f > t\})dt$. As f is continuous, $\{f > t\}$ is open, and its closure is $\{f \geq t\}$. Its boundary is therefore $\{f = t\}$. The set of t such that $P(\{f = t\}) > 0$ are precisely the discontinuity points of the distribution function $t \mapsto P(\{f \leq t\})$. There are at most countably many such points. Therefore, $P_n(\{f > t\}) \to P(\{f > t\})$ for every t outside a countable set of points. By the dominated convergence theorem,

$$P_n(f) = \int_0^1 P_n(\{f > t\})dt \to \int_0^1 P_n(\{f > t\})dt = P(f)\,.$$

□

Remark 2.2. The weak⋆ convergence of probability measure on $\mathcal{X} = \mathbb{R}$ can be characterized by means of the *distribution function* $F_P(x) = P((-\infty, x])$ of a measure $P \in \mathcal{P}(\mathbb{R})$. In this case, $P_n \to P$ is also equivalent to: $F_{P_n}(x) \to F_P(x)$ for every F_P-continuity point x. The implication $\Rightarrow$ is a consequence of condition v) in Theorem 2.3, used with $A = (-\infty, x]$. The proof of the converse is similar to the proof of v)$\Rightarrow$i) in Theorem 2.3.

2.1.4 Mapping theorem

Let $h : \mathcal{X} \to \mathcal{Y}$ be a measurable function between two Polish spaces $\mathcal{X}$ and $\mathcal{Y}$. Let D_h be the set of discontinuities of h (*n.b.* it is a Borel set). The pushforward measure of P w.r.t. h is the measure $h_\# P$ defined by $h_\# P(A) = P(h^{-1}(A))$ for any $A \in \mathcal{B}(\mathcal{Y})$. Equivalently, by the transfer theorem, it is the measure given by:

$$h_\# P(f) = \int f(h(x)) dP(x)$$

for any $f \in C_b(\mathcal{Y})$.

Theorem 2.4. *If $P_n \to P$ and $P(D_h) = 0$, then $h_\# P_n \to h_\# P$.*

Proof. Let F be a closed subset of $\mathcal{Y}$. One has:

$$\limsup_n h_\# P_n(F) \leq \limsup_n P_n(\overline{h^{-1}(F)})) \leq P(\overline{h^{-1}(F)})) = P(\overline{h^{-1}(F)}) \cap D_h^c) \,,$$

where we used that $P(D_h^c) = 1$. Let $x \in \overline{h^{-1}(F)}) \cap D_h^c$. By definition, x is a continuity point of h, and $x = \lim_n x_n$ for some sequence x_n such that $h(x_n) \in F$. Passing to the limit, $h(x_n) \to h(x)$ because h is continuous at x. As F is closed, $h(x) \in F$. We have shown that $\overline{h^{-1}(F)}) \cap D_h^c \subset h^{-1}(F)$. Thus, $\limsup_n h_\# P_n(F) \leq h_\# P(F)$. This shows that $P_n \to P$. □

2.2 Tightness and the Prokhorov theorem

Unless $\mathcal{X}$ is compact, the space $\mathcal{P}(\mathcal{X})$ is in general to compact. For instance, if $\mathcal{X} = \mathbb{R}$, the sequence δ_n admits no converging subsequence as n tends to infinity. The aim of this section is to characterize the *relatively compact* subsets of $\mathcal{P}(\mathcal{X})$, that is, the subsets which have a compact closure in $\mathcal{P}(\mathcal{X})$ equipped with the weak⋆ topology.

2.2.1 Tightness

Definition 2.5. A subset $A \subset \mathcal{P}(\mathcal{X})$ is called *tight* if, for every $\epsilon > 0$, there exists a compact subset $K \subset \mathcal{X}$ such that $P(K) > 1 - \epsilon$ for all $P \in A$.

Proposition 2.6. *Any singleton in $\mathcal{P}(\mathcal{X})$ is tight.*

Proof. As $\mathcal{X}$ is separable, it admits a dense denumerable subset given by the sequence $x_1, x_2, \ldots$. Thus, for every $n \geq 1$, $\mathcal{X}$ can be covered by open balls B_{nk} of radius $1/n$, centered at x_k: $\mathcal{X} = \bigcup_k B_{nk}$. Equivalently, $\bigcup_{k \leq K} B_{nk} \uparrow_K \mathcal{X}$. Consider $P \in \mathcal{P}(\mathcal{X})$ and $\epsilon > 0$. As $P(\bigcup_{k \leq \kappa} B_{nk}) \uparrow_\kappa 1$, there exists k_n such that $P(\bigcup_{k \leq k_n} B_{nk}) > 1 - \epsilon 2^{-n}$. By countable subadditivity of P, $P(\bigcap_n \bigcup_{k \leq k_n} B_{nk}) > 1 - \epsilon$. The set $A := \bigcap_n \bigcup_{k \leq k_n} B_{nk}$ is totally bounded, which means that, for any $r > 0$, it can be covered by finitely many balls of radius r. As $\mathcal{X}$ is complete, A has a compact closure. The compact set $\overline{A}$ satisfies $P(\overline{A}) < 1 - \epsilon$. □

2.2.2 Prokhorov theorem

Theorem 2.7. *A subset of $\mathcal{P}(\mathcal{X})$ is relatively compact if and only if it is tight.*

Proof. We only establish the converse of Theorem 2.7, which is the most important part, and refer to [1, Theorem 5.2] for the direct half.
Let $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ be a tight subset. We establish that it is relatively compact. As $\mathcal{P}(\mathcal{X})$ is metrizable, it is sufficient to establish sequential compactness. In other words, let (P_n) be a $\mathcal{A}$-valued sequence. The converse of the theorem will be established if one can extract a converging subsequence from (P_n).

For every integer $l \geq 1$, choose a compact set $K_l \subset \mathcal{X}$ such that $P_n(K_l) > 1-1/l$ for all n. One can choose (K_l) in such a way that $K_1 \subset K_2 \subset \dots$. The set $\mathcal{X}$ being separable, it can be covered by a countable family of open balls B_{kn}, where B_{nk} has radius $1/n$, and is centered at x_k, where $(x_1, x_2, \dots)$ is a dense sequence on $\mathcal{X}$. Consider a class of set $\mathcal{H}$, formed by the emptyset $\emptyset_{\mathcal{X}}$ and all finite unions of sets of the form $B_{nk} \cap K_l$ for $n, k, l \geq 1$. As $\mathcal{H}$ is denumerable, one can index its elements as $\mathcal{H} = \{H_1, H_2, \dots\}$. We prove that one can extract a subsequence (P_{ψ_n}) such that $\forall H \in \mathcal{H}$, the limit

$$\alpha(H) := \lim_{n\to\infty} P_{\psi_n}(H) \tag{2.2}$$

exists. To construct this extracted sequence, we use the so called "diagonal process". The sequence $(P_n(H_1))$ being bounded (hence relatively compact), there exists a value $\alpha(H_1) \in [0,1]$ and a subsequence $P_{\varphi_1(n)}(H_1)$ such that $P_{\varphi_1(n)}(H_1) \to \alpha(H_1)$. Then, the sequence $(P_{\varphi_1(n)}(H_2))$ being bounded, there exists $\alpha(H_2)$ and an other extractor φ_2 such that $P_{\varphi_1(\varphi_2(n))}(H_2) \to \alpha(H_2)$. Iterating the process, for every $k \geq 1$, there exists $\alpha(H_k)$ such that $P_{\varphi_1\circ\cdots\circ\varphi_k(n)}(H_k) \to \alpha(H_k)$. Define $\psi_n = \varphi_1 \circ \cdots \circ \varphi_n(n)$. It holds that: $\forall k,\ P_{\psi_n}(H_k) \to \alpha(H_k)$. Thus, Eq. (2.2) holds.
The main part of the proof consists in proving that $\alpha(\,\cdot\,)$, which is only defined on the class $\mathcal{H}$, generalizes to a probability measure P, defined on all Borel sets of $\mathcal{X}$. Eventually, this probability P will be shown to be the limit of P_{ψ_n}.
We proceed as in Chapter 1. Define for all subsets $A \subset \mathcal{X}$, the set function:

$$P^*(A) = \inf_{U \text{ open}: A\subset U} \ \sup_{H\in\mathcal{H}: H\subset U} \alpha(H)\,.$$

Suppose that one is able to establish the following points:

a) P^* is an outer measure,

b) Each closed subset of $\mathcal{X}$ is P^*-measurable.

Then, as the class of P^*-measurable sets is a σ-field, it contains all Borel sets. By Proposition 1.13, the restriction P of P^* to $\mathcal{B}(\mathcal{X})$ is countably additive, hence a probability measure, because:

$$1 \geq P(\mathcal{X}) = \sup_{H\in\mathcal{H}} \alpha(H) \geq \sup_l \alpha(K_l) \geq \sup_l (1 - \frac{1}{l}) = 1\,.$$

Moreover, for any open set U,

$$P(U) = \sup_{H \in U} \alpha(H) = \sup_{H \in U} \lim_n P_{\psi_n}(H) \leq \liminf_n P_{\psi_n}(U)$$

By the Portmanteau theorem, $P_n \to P$. The proof is concluded. It remains to show the two points a-b) above. The proof of this is technical. We refer to [1, pp.62]. □

Chapter 3

Convergences of random variables

In this chapter, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and $\mathcal{X}$ is a Polish (separable complete metric) space, equipped with its Borel σ-field. The letters X, X_n designate random variables on $\Omega \to \mathcal{X}$.

3.1 Almost sure, in probability

3.1.1 Definitions and properties

Definition 3.1. The sequence (X_n) converges *in probability* to X, noted $X_n \xrightarrow{P} X$, if

$$\forall \varepsilon > 0, \ \lim_n \mathbb{P}(d(X_n, X) > \varepsilon) = 0 \, .$$

The sequence (X_n) converges *almost surely* (a.s.) to X, noted $X_n \xrightarrow{a.s.} X$, if $X_n \to X$ $\mathbb{P}$-almost everywhere (a.e.). Otherwise stated, if there exists $A \in \mathcal{F}$ such that (s.t.) $\mathbb{P}(A) = 1$, and s.t. for every $\omega \in A$, $X_n(\omega) \to X(\omega)$.

Proposition 3.2. *If $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{P} X$.*

Proof. This follows from Lebesgue's dominated convergence theorem. □

Proposition 3.3. *If it holds that*

$$\forall \varepsilon > 0, \ \sum_n \mathbb{P}(d(X_n, X) > \varepsilon) < \infty \, ,$$

then $X_n \xrightarrow{a.s.} X$.

Proof. By the Borel-Cantelli lemma A.3, $\mathbb{P}(\limsup_n \{d(X_n, X) > \varepsilon\}) = 0$. Choosing $q \in \mathbb{N}^*$ and setting $\varepsilon = 1/q$, this yields

$$\mathbb{P}\left(\exists n, \, \forall k \geq n, \, d(X_n, X) \leq 1/q\right) = 1 \, ,$$

and by taking the intersection of these events for all $q \in \mathbb{N}^*$, it follows that

$$\mathbb{P}\left(\forall q \in \mathbb{N}^*,\, \exists n,\, \forall k \geq n,\, d(X_n, X) \leq 1/q\right) = 1\,,$$

which reads $\mathbb{P}(\lim_n d(X_n, X) = 0) = 1$. □

Proposition 3.4. *The following statements are equivalent:*

i. $X_n \xrightarrow{P} X$ *;*

ii. From any subsequence of (X_n)*, one can extract a further subsequence that converges to* X *almost surely.*

Proof. i⇒ ii. For every $\varepsilon > 0$, $\mathbb{P}(d(X_n, X) > \varepsilon) \to 0$. Thus, for every n, there exists φ_n such that $\mathbb{P}(d(X_{\varphi_n}, X) > \varepsilon) \leq 2^{-n}$. This implies that $\sum_n \mathbb{P}(d(X_{\varphi_n}, X) > \varepsilon) < \infty$. Therefore, $X_{\varphi_n} \xrightarrow{a.s.} X$ by Prop. 3.3. This proves the point *ii.*
ii ⇒ i. By contradiction, assume that X_n does not converge in probability to X. Thus, there exists some $\varepsilon > 0$ and some subsequence (X_{φ_n}) such that for every n, $\mathbb{P}(d(X_{\varphi_n}, X) > \varepsilon) > \varepsilon$. By the standing assumption, one can extract a further subsequence $(X_{(\phi\circ\psi)_n})$ which converges a.s. to X. The latter satisfies as well $\mathbb{P}(d(X_{(\phi\circ\psi)_n}, X) > \varepsilon) > \varepsilon$. But the dominated convergence theorem implies that $\mathbb{P}(d(X_{(\phi\circ\psi)_n}, X) > \varepsilon) \to 0$, which leads to a contradiction. □

Define $\mathcal{Y}$ be an other Polish space. Let $h : \mathcal{X} \to \mathcal{Y}$ be a Borel map. Denote by D_h the (measurable) set of discontinuities of h.

Theorem 3.5 (Mapping theorem)**.** *Assume that* $\mathbb{P}(X \in D_h) = 0$.

i. If $X_n \xrightarrow{a.s.} X$*, then* $h \circ X_n \xrightarrow{a.s.} h \circ X$ *;*

ii. If $X_n \xrightarrow{P} X$*, then* $h \circ X_n \xrightarrow{P} h \circ X$.

Proof. i. Choose $A \in \mathcal{F}$ s.t. $\mathbb{P}(A) = 1$ and $X_n(\omega) \to X(\omega)$ for every $\omega \in A$. For every $\omega \in A \cap X^{-1}(D_h)^c$, h is continuous at point $X(\omega)$, hence $h(X_n(\omega)) \to h(X(\omega))$. As $\mathbb{P}(A \cap X^{-1}(D_h)^c) = 1$, the statement is proven.
ii. Consider an arbitrary subsequence $(h \circ X_{\varphi_n})$. As $X_n \xrightarrow{P} X$, Prop. 3.4 implies that there exists that one can extract a subsequence from (X_{φ_n}) which converges a.s. to zero. Denote by $(X_{(\varphi\circ\psi)_n})$ this subsequence: $X_{(\varphi\circ\psi)_n} \xrightarrow{a.s.} X$. Using point *i.*, $h \circ X_{(\varphi\circ\psi)_n} \xrightarrow{a.s.} h \circ X$. Hence, we have shown that from any subsequence $(h \circ X_{\varphi_n})$, one can extract a further subsequence which converges a.s. to zero. By Prop. 3.4, the conclusion follows. □

We shall always equipp the product space $\mathcal{X} \times \mathcal{Y}$ with its Borel σ-field $\mathscr{B}(\mathcal{X} \times \mathcal{Y}) = \mathscr{B}(\mathcal{X}) \otimes \mathscr{B}(\mathcal{Y})$.

Proposition 3.6. *Let (X_n), X be r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{B}(\mathcal{X}))$. Let (Y_n), Y be r.v. on $(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{Y}, \mathscr{B}(\mathcal{Y}))$. If $X_n \xrightarrow{a.s.} X$ and $Y_n \xrightarrow{a.s.} Y$, then $(X_n, Y_n) \xrightarrow{a.s.} (X, Y)$. The same holds when the almost sure convergence is replaced by the convergence in probability.*

Proof. The proof is immediate for the almost sure convergence, thus we only prove the statement for the convergence in probability. Assume that $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Consider a subsequence $(X_{\varphi_n}, Y_{\varphi_n})$. By Prop. 3.4, one can extract a subsequence, say $\varphi' := \varphi \circ \psi$ s.t. $X_{\varphi'_n} \xrightarrow{a.s.} X$, and a further subsequence say $\varphi'' := \varphi' \circ \kappa$ s.t. $Y_{\varphi''_n} \xrightarrow{a.s.} Y$. Along this subsequence, $(X_{\varphi''_n}, Y_{\varphi''_n}) \xrightarrow{a.s.} (X, Y)$. This shows $(X_n, Y_n) \xrightarrow{P} (X, Y)$ again by Prop. 3.4. □

Corollary 3.7. *Assume that $\mathcal{X} = \mathbb{R}^d$, and denote by $(X_n^{(1)}, \cdots, X_n^{(d)})$ the real coordinates of X_n. The sequence X_n converges a.s. to X if and only if for every $i = 1 \dots d$, $X_n^{(i)}$ converges a.s. to $X^{(i)}$. The same holds when the almost sure convergence is replaced by the convergence in probability.*

Proof. The proof is immediate for the almost sure convergence, thus we only prove the statement for the convergence in probability. Assume that $X_n \xrightarrow{P} X$. For every $i = 1 \dots d$, the projection map $(x^{(1)}, \dots, x^{(n)}) \mapsto x^{(i)}$ being continuous, it holds that $X_n^{(i)} \xrightarrow{P} X^{(i)}$ by Th. 3.5. Conversely assume that $X_n^{(i)} \xrightarrow{P} X^{(i)}$ for all i. By induction from Prop. 3.6, $(X_n^{(1)}, \dots, X_n^{(k)}) \xrightarrow{P} (X^{(1)}, \dots, X^{(k)})$, and the result follows. □

3.1.2 Strong Law of Large Numbers (LLN)

The acronym iid stands for independent and identically distributed. If X is a r.v. on $\mathcal{X}$, the notation $\mathbb{E}(X)$ stands for the vector $(\mathbb{E}(X^{(1)}), \dots, \mathbb{E}(X^{(d)}))$, whenever it is well defined.

Theorem 3.8. *Let $(X_n : n \in \mathbb{N}^*)$ be an iid sequence on $\mathbb{R}^d$. Assume that $\mathbb{E}(\|X_1\|) < \infty$. Then, $n^{-1} \sum_{i=1}^n X_i \xrightarrow{a.s.} \mathbb{E}(X_1)$.*

Proof. By Prop. 3.7, it is sufficient to prove the result on $\mathbb{R}$. It is even sufficient to prove the result in the case where $X_n \geq 0$ for every n. Indeed, if X_n is not necessarily positive, just use the decomposition $X_n = X_n^+ - X_n^-$ where $X_n^\pm = \max(\pm X_n, 0)$, apply the LLN to X_n^+ and X_n^- separately, and conclude using Th. 3.5. We thus assume from now on that $X_n \geq 0$. We set $S_n := \sum_{i=1}^n X_i$. We first provide the proof in the simpler case where $\mathrm{Var}(X_1) < \infty$, and then explain the extension to the general case.

Case 1: Var$(X_1) < \infty$. For all $\varepsilon > 0$, Markov's inequality implies that

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}S_n}{n}\right| > \varepsilon\right) \leq \frac{\mathrm{Var}(X_1)}{n\varepsilon^2}.$$

Choose $\alpha > 1$ and define for each n, $k_n := \lceil \alpha^n \rceil$, where $\lceil x \rceil$ represents the smallest integer greater than or equal to x. There exists a constant $c_\varepsilon < \infty$ s.t.

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\left| \frac{S_{k_n} - \mathbb{E}S_{k_n}}{k_n} \right| > \varepsilon \right) \leq c_\varepsilon \sum_{n=1}^{\infty} k_n^{-1} < \infty \,.$$

By Prop. 3.3, $\frac{S_{k_n}}{k_n} \xrightarrow{a.s.} \mathbb{E}(X_1)$. Choose any $i \in \mathbb{N}^*$. Denote by $n(i)$ an integer s.t.

$$k_{n(i)} \leq i \leq k_{n(i)+1} \,.$$

Such an integer $n(i)$ exists because (k_n) is an increasing sequence. As $X_n \geq 0$ for all n, we deduce that $S_{k_{n(i)}} \leq S_i \leq S_{k_{n(i)}+1}$. Dividing by i,

$$\frac{S_{k_{n(i)}}}{k_{n(i)}} \frac{k_{n(i)}}{i} \leq \frac{S_i}{i} \leq \frac{S_{k_{n(i)+1}}}{k_{n(i)+1}} \frac{k_{n(i)+1}}{i} \,.$$

It is not difficult to show that

$$\frac{k_{n(i)}}{i} \geq \frac{1}{\alpha} \quad \text{and} \quad \frac{k_{n(i)+1}}{i} \leq \alpha + \frac{1}{i} \,.$$

Denote by A_α an event s.t. $\mathbb{P}(A_\alpha) = 1$ and for all $\omega \in A_\alpha$, $\frac{S_{k_n}(\omega)}{k_n} \xrightarrow{a.s.} \mathbb{E}(X_1)$. We obtain that for all $\omega \in A_\alpha$,

$$\frac{1}{\alpha}\mathbb{E}(X_1) \leq \liminf_{i \to \infty} \frac{S_i(\omega)}{i} \leq \limsup_{i \to \infty} \frac{S_i}{i} \leq \alpha \mathbb{E}(X_1) \,.$$

We now set α of the form $\alpha = 1 + \frac{1}{q}$ for $q \in \mathbb{N}^*$. For any such q, there exists a probability one event $A_{1+1/q}$ s.t. the above inequality hold on that event. The event $A := \bigcap_{q \in \mathbb{N}^*} A_{1+1/q}$ is s.t. $\mathbb{P}(A) = 1$, and for every $\omega \in A$,

$$\forall q \in \mathbb{N}^*, \quad \frac{1}{1+q^{-1}}\mathbb{E}(X_1) \leq \liminf_{i \to \infty} \frac{S_i(\omega)}{i} \leq \limsup_{i \to \infty} \frac{S_i(\omega)}{i} \leq (1+q^{-1})\mathbb{E}(X_1) \,.$$

Letting $q \to \infty$, we obtain that for all $\omega \in A$, $\lim_i \frac{S_i}{i} = \mathbb{E}(X_1)$.

Case 2: Var$(X_1) = +\infty$. Define $Y_n := X_n \mathbb{1}_{X_n < n}$ and $S_n^\star := \sum_{i \leq n} Y_i$. For any $\alpha > 1$, define k_n as above. For all $\varepsilon > 0$,

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}\left(\left| \frac{S_{k_n}^\star - \mathbb{E}S_{k_n}^\star}{k_n} \right| > \varepsilon \right) &\leq \frac{1}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{k_n^2} \sum_{i=1}^{k_n} \mathrm{Var}(Y_i) \\ &\leq \frac{1}{\varepsilon^2} \sum_{i=1}^{\infty} \mathrm{Var}(Y_i) \sum_{n : k_n \geq i} \frac{1}{\alpha^{2n}} \\ &\leq \frac{1}{\varepsilon^2 (1-\alpha^{-1})} \sum_{i=1}^{\infty} \frac{\mathbb{E}(Y_i^2)}{i^2} \,. \end{aligned}$$

Using that $\mathbb{E}(Y_i^2) = \sum_{k=0}^{i-1} \mathbb{E}(X_1^2 \mathbb{1}_{k \le X_1 < k+1})$ and permuting the indices k and i, and setting $c := \frac{1}{\varepsilon^2(1-\alpha^{-1})}$, we obtain

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}\left(\left|\frac{S^\star_{k_n} - \mathbb{E}S^\star_{k_n}}{k_n}\right| > \varepsilon\right) &\le c \sum_{k=0}^{\infty} \mathbb{E}(X_1^2 \mathbb{1}_{k \le X_1 < k+1}) \sum_{i>k} \frac{1}{i^2} \\ &\le c \sum_{k=0}^{\infty} \mathbb{E}(X_1^2 \mathbb{1}_{k \le X_1 < k+1}) \frac{1}{k+1} \\ &\le c \sum_{k=0}^{\infty} \mathbb{E}(X_1 \mathbb{1}_{k \le X_1 < k+1}) \ \le c\mathbb{E}(X_1)\,.\end{aligned}$$

By Prop. 3.3, $\frac{S^\star_{k_n}}{k_n} \xrightarrow{a.s.} \mathbb{E}(X_1)$. Following the exact same proof as in Case 1, we deduce that $\frac{S^\star_n}{n} \xrightarrow{a.s.} \mathbb{E}(X_1)$. It remains to prove that $\frac{S_n}{n}$ has the same a.s. limit as $\frac{S^\star_n}{n}$, and the proof will be complete. To that end, remark that $\mathbb{P}(X_n \neq Y_n) = \mathbb{P}(X_n > n) = \mathbb{P}(X_1 > n)$. By the same kind of derivations,

$$\begin{aligned}\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbb{P}(k+1 \ge X_1 > k) \\ &= \sum_{k=1}^{\infty} k\, \mathbb{P}(k+1 \ge X_1 > k) \\ &\le \sum_{k=1}^{\infty} \mathbb{E}(X_1 \mathbb{1}_{k+1 \ge X_1 > k}) \ \le \mathbb{E}(X_1) \ < \infty\,.\end{aligned}$$

By the Borel-Cantelli lemma A.3, we obtain that $\mathbb{P}(\limsup_n \{X_n \neq Y_n\}) = 0$, which means that, $\mathbb{P}$-a.e., $X_n = Y_n$ for all n outside a finite set. Thus, $\mathbb{P}$-a.e., $\lim_n n^{-1} S_n = \lim_n n^{-1} S^\star_n$. This concludes the proof. □

3.2 Convergence in law

3.2.1 Definition and properties

We recall that $X_\# \mathbb{P}$ is the law of X. If $f : \mathcal{X} \to \mathbb{R}$ is a measurable mapping, we recall that the notation $f(X)$ means $f \circ X$. The expectation of $f(X)$ is the quantity

$$\mathbb{E}(f(X)) := \mathbb{P}(f \circ X) \ = X_\# \mathbb{P}(f)\,,$$

whenever it is well defined.

Definition 3.9. The sequence (X_n) converges *in law* to X, noted $X_n \xrightarrow{\mathcal{L}} X$, if $(X_n)_\# \mathbb{P} \to X_\# \mathbb{P}$. Equivalently, if $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ for all $f \in C_b(\mathcal{X})$.

Remark 3.1. The convergence in law is the weak$\star$ convergence of the laws of the random variables. Therefore, only their distributions matter. Actually, all variables (X_n) and X could have been defined on different probability spaces. The introduction of a unique probability space is mainly for notational convenience.

Remark 3.2. Because only the distribution of X matters in the definition (and not X itself), we will also use the following notation. If (X_n) is a sequence of r.v. on $\mathcal{X}$ and if $P \in \mathcal{P}(\mathcal{X})$, we will write $X_n \to P$ as an equivalent to $(X_n)_\#\mathbb{P} \to P$. Thus, if X_n, X have the distribution P_n, P respectively, the notations $X_n \xrightarrow{\mathcal{L}} X$, $P_n \to P$ and $X_n \xrightarrow{\mathcal{L}} P$ mean the same fact. Convergence in law is also called convergence in distribution. Notations $X_n \xrightarrow{\mathcal{L}} X$ and $X_n \xrightarrow{\mathcal{D}} X$ are often used with the same meaning as $X_n \xrightarrow{\mathcal{L}} X$.

Proposition 3.10. *If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{\mathcal{L}} X$.*

Proof. Let $f \in C_b(\mathcal{X})$. By Th. 3.5, $f(X_n) \xrightarrow{P} f(X)$. The sequence $(\mathbb{E}(f(X_n)))_n$ is bounded. Let ℓ be an accumulation point, that is, $\mathbb{E}(f(X_n)) \to \ell$ along some subsequence. By Prop. 3.4, one can extract a further subsequence such that $f(X_n) \xrightarrow{a.s.} f(X)$ along that subsequence. By the dominated convergence theorem, $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$ along that subsequence. Therefore, $\ell = \mathbb{E}(f(X))$. This proves that $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$. □

3.2.2 Characteristic Functions, Lévy's Theorem

In this parapgraph, we assume that $\mathcal{X} = \mathbb{R}^d$ is finite dimensional.

Theorem 3.11 (Lévy)**.** *Let $(P_n), P \in \mathcal{P}(\mathbb{R}^d)$. The following properties are equivalent:*

i. $P_n \to P$ *;*

ii. *For every $t \in \mathbb{R}^d$, $\phi_{P_n}(t) \to \phi_P(t)$.*

Proof. *i* ⇒ *ii.* The direct half is evident, as the mapping $x \mapsto e^{i\langle t,x\rangle}$ is bounded and continuous.
ii ⇒ *i.* Assume that (P_n) is tight (this point is proven below). By Th. 2.7, there exists a subsequence (P_{φ_n}) and a probability measure ν s.t. $P_{\varphi_n} \to \nu$. By the first point just proven, $\phi_{\varphi_n}(t) \to \phi_\nu(t)$ for every t. By identification of the limit, $\phi_\nu(t) = \phi_P(t)$, hence $P = \nu$ by Th. A.4. As a conclusion, (P_n) is tight and every subsequence converging weakly⋆ converges to P. This shows that $P_n \to P$.
Thus, the crux of the proof consists in showing that (P_n) is tight. We provide the proof in the case $d = 1$. One has for every $a > 0$, by Fubini's theorem

$$\begin{aligned}\frac{1}{a}\int_{-a}^{a}(1-\phi_{P_n}(t))dt &= \int\left(\frac{1}{a}\int_{-a}^{a}(1-e^{itx})dt\right)dP_n(x)\\ &= 2\int\left(1-\frac{\sin ax}{ax}\right)dP_n(x)\,,\end{aligned}$$

and since the integrand is positive, we have for every measurable set $C \subset \mathbb{R}$,

$$\begin{aligned}\frac{1}{a}\int_{-a}^{a}(1-\phi_{P_n}(t))dt &\geq 2\int_C\left(1-\frac{\sin ax}{ax}\right)dP_n(x)\\ &\geq \int_C 2\left(1-\frac{1}{a|x|}\right)dP_n(x)\,.\end{aligned}$$

Now setting in particular $C := \mathbb{R}\backslash[-\frac{2}{a}, \frac{2}{a}]$, we obtain:

$$\frac{1}{a}\int_{-a}^{a}(1-\phi_{P_n}(t))dt \geq P_n(\mathbb{R}\backslash[-\tfrac{2}{a}, \tfrac{2}{a}])\,.$$

Consider the lefthand side. By the dominated convergence theorem, it converges to $\frac{1}{a}\int_{-a}^{a}(1-\phi_P(t))dt$. Now consider the latter limit. As ϕ_P is continuous at zero, it is an easy exercise to show that $\frac{1}{a}\int_{-a}^{a}(1-\phi_P(t))dt$ converges to zero as $a \to \infty$. For any $\varepsilon > 0$, choose $a > 0$ s.t. $\frac{1}{a}\int_{-a}^{a}(1-\phi_P(t))dt < \varepsilon$. Since $\frac{1}{a}\int_{-a}^{a}(1-\phi_{P_n}(t))dt \to \frac{1}{a}\int_{-a}^{a}(1-\phi_P(t))dt$, we obtain: $\limsup_n P_n(\mathbb{R}\backslash[-\frac{2}{a}, \frac{2}{a}]) < \varepsilon$, which concludes the proof by Prop. 2.6-ii. □

Remark 3.3. Examining the proof of Th. 3.11, the key argument is to show that condition *ii* implies the tightness of (P_n). To prove this, we did not explicitely used the fact that the limit ϕ_P is a characteristic function: we just needed the continuity of ϕ_P at zero. Thus, one can state a more general formulation of Lévy's theorem, which you should be able to prove by your own: *If there exists a function $\psi : \mathbb{R}^d \to \mathbb{C}$ continuous at zero s.t. $\phi_{P_n}(t) \to \psi(t)$ for all t, then there exists $P \in \mathcal{P}(\mathbb{R}^d)$ s.t. $\phi_P = \psi$ and $P_n \to P$.*

The following lemma shows that the convergence in law of a sequence of random *vectors* can, in a certain sense, be reduced to convergence in law of *real-valued* r.v.: this can be convenient in some proofs.

Lemma 3.12 (Cramer-Wold). *The following points are equivalent:*

i. $X_n \xrightarrow{\mathcal{L}} X$;

ii. $\forall t \in \mathbb{R}^d,\ \langle t, X_n\rangle \xrightarrow{\mathcal{L}} \langle t, X\rangle$.

Proof. $X_n \xrightarrow{\mathcal{L}} X$ if and only if $\mathbb{E}(e^{\imath\langle t,X_n\rangle}) \to \mathbb{E}(e^{\imath\langle t,X\rangle})$ for all $t \in \mathbb{R}^d$. This is equivalent to $\mathbb{E}(e^{\imath u\langle t,X_n\rangle}) \to \mathbb{E}(e^{\imath u\langle t,X\rangle})$ for all $t \in \mathbb{R}^d$ and all $u \in \mathbb{R}$, which reads $\phi_{\langle t,X_n\rangle}(u) \to \phi_{\langle t,X\rangle}(u)$ for all t, u. □

3.2.3 Central Limit Theorem

We start with a short reminder.
We refer to the *covariance matrix* of a random vector $X = (X^{(1)}, \dots, X^{(d)})$ on $\mathbb{R}^d$ as the positive semidefinite linear operator on $\mathbb{R}^d \to \mathbb{R}^d$, noted $\mathrm{Cov}(X)$, whose coefficient (i,j) is the covariance $\mathrm{Cov}(X^{(i)}, X^{(j)})$ of $X^{(i)}$ and $X^{(j)}$. If $\mathbb{E}(\|X\|^2) < \infty$, then $\mathrm{Cov}(X)$ is well-defined. By definition, the r.v. X is a Gaussian vector if $\langle t, X\rangle$ is a Gaussian variable for every $t \in \mathbb{R}^d$. If X is a Gaussian vector, then $\mathbb{E}(\|X\|^2) < \infty$ and for every $t \in \mathbb{R}^d$,

$$\phi_X(t) = e^{\imath\langle t,\mathbb{E}(X)\rangle}e^{-\frac{1}{2}\langle t,\mathrm{Cov}(X)t\rangle}\,.$$

In particular, the law of the Gaussian vector X depends only on its expectation $\mathbb{E}(X)$ and its covariance matrix $\mathrm{Cov}(X)$. The notation $\mathcal{N}(c, \Sigma)$ represents the law of a Gaussian vector having expectation c and covariance matrix Σ. By Lemma A.7, if X has the distribution $\mathcal{N}(c, \Sigma)$, then the r.v. $AX + b$ has the distribution $\mathcal{N}(Ac + b, A\Sigma A^*)$ for every linear operator A and every vector b.

Let $(X_n : n \in \mathbb{N}^*)$ be a sequence of r.v. defined on $(\Omega, \mathcal{F}, \mathbb{P})$ into $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$.

Theorem 3.13. *Suppose that (X_n) is an iid sequence of random vectors s.t. $\mathbb{E}(\|X_1\|^2) < \infty$. Then,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mathbb{E}(X_1)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{Cov}(X_1))\,.$$

Th. 3.13 is a consequence of the more general result stated below in Th. 3.14. A collection of r.v. $(X_{i,n} : 1 \le i \le n,\, n \in \mathbb{N}^*)$ on $(\Omega, \mathcal{F}, \mathbb{P})$ into $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ will be refer to as a (triangular) *array*. Such an array is said to satisfy the *Lindebergh condition* if

$$\forall \varepsilon > 0,\ \lim_n \sum_{i=1}^{n} \mathbb{E}\left(\|X_{i,n}\|^2 \mathbb{1}_{\|X_{i,n}\| > \varepsilon}\right) = 0\,. \tag{3.1}$$

Theorem 3.14. *Suppose that $(X_{i,n} : 1 \le i \le n,\, n \in \mathbb{N})$ is an array of random vectors on $\mathbb{R}^d$. Assume the following.*

i. For every $n \in \mathbb{N}^*$, the r.v. $X_{1,n}, \ldots, X_{n,n}$ are independent;

ii. For every $n \in \mathbb{N}^*$ and $i = 1, \ldots, n$, $\mathbb{E}(X_{i,n}) = 0$;

iii. The Lindebergh condition (3.1) is satisfied;

iv. There exists a matrix Σ s.t.

$$\lim_n \sum_{i=1}^{n} \mathrm{Cov}(X_{i,n}) = \Sigma\,,$$

where the limit is taken componentwise.

Then,

$$\sum_{i=1}^{n} X_{i,n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma)\,.$$

Verify by your own that Th. 3.13 is indeed a consequence of Th. 3.14. Before proving Th. 3.14, we need the following technical lemmas.

Lemma 3.15. *For every $z \in \mathbb{C}$, $|e^z - 1 - z| \le |z|^2 e^{|z|}$.*

Proof. Just expand $|e^z - 1 - z| \le \sum_{k \ge 2} \frac{|z|^k}{k!}$ and the conclusion follows. □

Lemma 3.16. *For every $x \in \mathbb{R}$, $|1 + \imath x - \frac{x^2}{2} - e^{\imath x}| \le \min(x^2, |x|^3)$.*

Proof. Write the Taylor-Lagrange expansion of $e^{\imath x}$ at the second and third order. □

Lemma 3.17. *Define* $\mathcal{U} := \{z \in \mathbb{C} : |z| \leq 1\}$*. Consider* $n \in \mathbb{N}^*$ *and two sequences* $(z_i : i = 1 \dots n)$, $(w_i : i = 1 \dots n)$ *on* $\mathcal{U}$*. Then,*

$$\left| \prod_{i=1}^n z_i - \prod_{i=1}^n w_i \right| \leq \sum_{i=1}^n |z_i - w_i| \,.$$

Proof. By induction. □

Proof of Th. 3.14: the scalar case. We first consider the case where $d = 1$ (random variables are real-valued). The vector case will be handled at the end of this section. Then, $\Sigma \geq 0$ is a scalar. Moreover, the function $x \mapsto \sqrt{\Sigma}\, x$ being continuous, The mapping theorem implies that it is sufficient to prove the result for $\Sigma = 1$.
Define $s_{i,n} := \mathrm{Var}(X_{i,n})$, thus

$$\lim_n \sum_{i=1}^n s_{i,n} = 1 \,. \tag{3.2}$$

We define $S_n := \sum_{i=1}^n X_{i,n}$. Consider a fixed $t \in \mathbb{R}$. By independence, we obtain

$$\phi_{S_n}(t) = \prod_{i=1}^n \phi_{X_{i,n}}(t) \,.$$

By the triangular inequality,

$$\left| \phi_{S_n}(t) - e^{-t^2/2} \right| \leq A_n + B_n + C_n \,,$$

where

$$\begin{aligned}
A_n &:= \left| \prod_{i=1}^n \phi_{X_{i,n}}(t) - \prod_{i=1}^n (1 - \frac{s_{i,n} t^2}{2}) \right| \\
B_n &:= \left| \prod_{i=1}^n (1 - \frac{s_{i,n} t^2}{2}) - \prod_{i=1}^n e^{-s_{i,n} t^2/2} \right| \\
C_n &:= \left| \prod_{i=1}^n e^{-s_{i,n} t^2/2} - e^{-t^2/2} \right| \,.
\end{aligned}$$

Now, we must prove that each of the terms A_n, B_n, C_n converges to zero. By Prop. A.6, this will imply that, for every $t \in \mathbb{R}$, $\phi_{S_n}(t) \to \phi_{\mathcal{N}(0,1)}(t)$. By Th. 3.11, we will conclude that $S_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$.

By Eq. (3.2), the term C_n converges to zero as $n \to \infty$. Note that for all $\varepsilon > 0$,

$$\max_{1\le i\le n} s_{i,n} \le \varepsilon^2 + \sum_{i=1}^n \mathbb{E}\left((X_{i,n})^2 \mathbb{1}_{|X_{i,n}|>\varepsilon}\right) ,$$

which implies by the Lindebergh condition that

$$\lim_n \max_{1\le i\le n} s_{i,n} = 0 . \tag{3.3}$$

As a consequence, the values $1 - \frac{s_{i,n}t^2}{2}$ are all in the interval $[-1,1]$ for n large enough. By lemma 3.17,

$$A_n \le \sum_{i=1}^n |\phi_{X_{i,n}}(t) - (1 - \frac{s_{i,n}t^2}{2})|$$
$$B_n \le \sum_{i=1}^n |1 - \frac{s_{i,n}t^2}{2} - e^{-s_{i,n}t^2/2}| .$$

By Lemma 3.15, $B_n \le \sum_{i=1}^n (\frac{s_{i,n}^2 t^4}{4} e^{s_{i,n}t^2/2})$, and thus $B_n \to 0$ by Eq. (3.2) and (3.3). It remains to prove that $A_n \to 0$. Denote $r(x) := e^{\imath x} - (1 + \imath x - \frac{x^2}{2})$. Then, using that $\mathbb{E}(X_{i,n}) = 0$,

$$\phi_{X_{i,n}}(t) = 1 - \frac{s_{i,n}t^2}{2} + \mathbb{E}(r(tX_{i,n})) .$$

Hence, using Lemma 3.16

$$A_n \le \sum_{i=1}^n \mathbb{E}(|r(tX_{i,n})|) \ \le \ \sum_{i=1}^n \mathbb{E}(\min((tX_{i,n})^2, |tX_{i,n}|^3)) .$$

Note that for every $\varepsilon > 0$ small enough, we have for every $x \in \mathbb{R}$,

$$\min((tx)^2, |tx|^3) \le (tx)^2 \mathbb{1}_{|x|>\varepsilon} + |tx|^3 \mathbb{1}_{|x|\le\varepsilon} .$$

Thus,

$$A_n \le \sum_{i=1}^n \mathbb{E}((tX_{i,n})^2 \mathbb{1}_{|X_{i,n}|>\varepsilon}) + \sum_{i=1}^n \mathbb{E}(|tX_{i,n}|^3 \mathbb{1}_{|X_{i,n}|\le\varepsilon}) .$$

The first term in the righthand side converges to zero as $n \to \infty$ by the Lindebergh condition. Therefore, $\limsup_n A_n \le \varepsilon$. As this holds for every $\varepsilon > 0$ small enough, $A_n \to 0$. The proof that $S_n \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$ is complete.

Proof of Th. 3.14: the vector case. Recall that if Z has the distribution $\mathcal{N}(0,\Sigma)$ and if $t \in \mathbb{R}^d$, then $\langle t, Z\rangle$ has the distribution $\mathcal{N}(0, \langle t, \Sigma t\rangle)$. By Lemma 3.12, it is therefore sufficient that for every $t \in \mathbb{R}^d$,

$$\sum_{i=1}^n \langle t, X_{i,n}\rangle \xrightarrow{\mathcal{L}} \mathcal{N}(0, \langle t, \Sigma t\rangle) .$$

The above convergence holds by applying Th. 3.14 to the real array $(\langle t, X_{i,n}\rangle)$.

3.3 Exercises

Exercice 1 (Borel-Cantelli)**.** 1. Let $(A_n)_n$ be a sequence of independent events. Establish the "second Borel-Cantelli lemma":

$$\sum_n \mathbb{P}(A_n) = +\infty \implies \mathbb{P}(\limsup_n A_n) = 1\,.$$

2. Let X_n be iid variables over $\{-1, 1\}$ such that $\mathbb{P}(X_1 = 1) = p$ where $0 < p < 1$. Show that, with probability 1, the sequence X_n contains an infinity of 1 and an infinity of -1. Define $S_n = \sum_{i=1}^n X_i$. Compute the probabilities of the events $S_{2n} = 0$ and $S_{2n+1} = 0$. Suppose $p \neq \frac{1}{2}$. Show that, with probability 1, the event $S_n = 0$ only occurs a finite number of times.

Exercice 2 (Borel-Cantelli)**.** Let $(X_n)_n$ be an iid sequence of exponential random variables with parameter 1.

1. Compute, for ϵ such that $|\epsilon| < 1$, the probability of the event $X_n/\ln n \geq 1+\epsilon$. Conclude that, with probability 1, the event $X_n/\ln n \geq 1+\epsilon$ occurs infinitely often when $-1 < \epsilon < 0$. Show that, always with probability 1, it occurs only finitely many times when $0 < \epsilon < 1$.

2. Conclude that
$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{X_n}{\ln n} = 1\right) = 1\,.$$

Exercice 3 (Recurrence events)**.** Disasters occur at times $T_1, T_2, \ldots$. Let X_i be the time interval between the $(i-1)$th and the ith disaster, so that $T_i = X_1 + \cdots + X_i$. Assume the sequence X_i is iid and that $\mathbb{E}X_1 < \infty$. Let

$$N(t) = \max\{n : T_n \leq t\}$$

be the number of disasters that occur up to time t.

1. Show that $N(t) < n$ if and only if $T_n > t$. Conclude that for all n, $\mathbb{P}(N(t) < n)$ tends to zero as $t \to \infty$. Conclude that $N(t) \xrightarrow{p.s.} +\infty$ as $t \to \infty$.

2. Show that for all t,
$$\frac{T_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{T_{N(t)+1}}{N(t)+1}\,(1 + \frac{1}{N(t)})\,.$$

   Conclude that $N(t)/t \xrightarrow{p.s.} \mathbb{E}(X_1)$ as $t \to \infty$.

Exercice 4 (Binomial convergence)**.** Let X_n be a binomial distribution with parameters $(n, \frac{\lambda}{n})$. Study the convergence in law of the sequence (X_n).

Exercice 5 (Convergence in law of an estimator)**.** We observe $X_1, \cdots, X_n$ independent identically distributed random variables with uniform distribution on $[0, \theta]$ where θ is an unknown value that we want to estimate. We are interested in the estimator:

$$\hat{\theta}_n = \max(X_1, \cdots, X_n) \,.$$

1. Compute the cumulative distribution function of $\hat{\theta}_n$.
2. Deduce the cumulative distribution function of $\Delta_n = n(\theta - \hat{\theta}_n)$.
3. Study the convergence in law of Δ_n.
4. Compare with the estimator $\tilde{X}_n = \frac{2}{n} \sum_{k=1}^n X_k$.

Exercice 6 (A tightness criterion)**.** Let $(P_n)_n$ be a sequence of probability measures on $\mathbb{R}^d$.

1. Suppose that $\sup_n \int \|x\| \, dP_n(x) < \infty$. Show that the sequence P_n is tight (use the Markov inequality).
2. More generally, suppose that $\sup_n \int V(x) dP_n(x) < \infty$ where $V : \mathbb{R}^d \to \mathbb{R}_+$ is a function with compact level sets (for every $M > 0$, the set $\{x : V(x) \leq M\}$ is compact). Show that the sequence P_n is tight.

Chapter 4

The space $C([0,1])$

To do.

Appendix A

Appendix

A.1 Topology refresher

Let $\mathcal{X}$ be a topological set. A neighborhood of a point $x \in \mathcal{X}$ is a subset of $\mathcal{X}$ that contains an open set containing x. A neighborhood basis at x is a collection B of neighborhoods of x such that every open set containing x contains at least one element from B.

The *closure* $\overline{A}$ of A is the smalled closed set which contains A. The *interior* $\mathring{A}$ (also noted A^{o}) of A is the largest open set containing A. A set A is said *compact* if for any cover of A by open sets, one can extract a finite subcover. It is said *relatively compact* if it has a compact closure.

A sequence $(v_n : n \in \mathbb{N})$ is said to be a *subsequence* of the sequence (u_n) if there exists a strictly increasing function $\varphi : \mathbb{N} \to \mathbb{N}$ s.t. for every $n \in \mathbb{N}$, $v_n = u_{\varphi_n}$. A subsequence of the subsequence (u_{φ_n}) is a sequence of the form $(u_{\varphi_{\psi_n}})$ where $\psi : \mathbb{N} \to \mathbb{N}$ is strictly increasing. This also writes $(u_{(\varphi\circ\psi)_n})$ where $\circ$ stands for the composition.

Let $\mathcal{X}$ be a metric space. A subset A of a metric space $\mathcal{X}$ is called *totally bounded* (or sometimes *precompact*) if, for every $\epsilon > 0$, it can be covered by finitely many balls of radius ϵ.

Proposition A.1. *A totally bounded subset of a complete metric space $\mathcal{X}$ has a compact closure.*

Proof. Let $A \subset \mathcal{X}$ be totally bounded. For every integer $n \geq 1$, let $\bigcup_{k=1}^{k_n} B_{kn}$ be a cover of A, formed by balls of radius $1/n$. Let (x_i) be a sequence on A. As $A \subset \bigcup_{k=1}^{k_1} B_{k1}$, there exists $\kappa_1 \leq k_1$ such that $B_{\kappa_1 1}$ contains an infinite number of points in the sequence (x_i). Otherwise stated, one can extract a subsequence, say $(x_{\varphi_1(i)} : i = 1, 2, \dots)$, contained in $B_{\kappa_1 1}$. We iterate the process. As the extracted sequence is covered by finitely many balls B_{k2}, on of these balls, say $B_{\kappa_2 2}$ contains infinitely many points $x_{\varphi_1(i)}$. One can therefore extract an other subsequence, say $x_{\varphi_1(\varphi_2(i))}$, which belongs to $B_{\kappa_1 1} \cap B_{\kappa_2 2}$. Iterating again the process, for any iteration n, we obtain a subsequence of the form $(x_{\varphi_1\circ\dots\circ\varphi_n(i)}$:

$i = 1, 2, \dots)$ which belongs to $B_{\kappa_1 1} \cap \cdots \cap B_{\kappa_n n}$. Now consider the sequence $y_n := x_{\varphi_1 \circ \cdots \circ \varphi_n(n)}$. For every $m \geq n$, $y_m \in B_{\kappa_n n}$, therefore, $d(y_n, y_m) \leq 2/n$. This proves that (y_n) is a Cauchy sequence on A. As $\mathcal{X}$ is complete, the Cauchy sequence (y_n) converges. We have shown that every sequence on A admits a converging subsequence, which makes A relatively compact. □

A.2 Borel-Cantelli lemma

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. An *event* is an element of $\mathcal{F}$. Let $(A_n : n \in \mathbb{N})$ be a sequence of events. We define

$$\bigcup_n A_n := \{\omega \in \Omega : \exists n, \omega \in A_n\}$$
$$\bigcap_n A_n := \{\omega \in \Omega : \forall n, \omega \in A_n\}$$

The sequence A_n is said *non-decreasing* if $A_n \subset A_{n+1}$ for every n, *non-increasing* if $A_{n+1} \subset A_n$ for every n.

Proposition A.2. *If (A_n) is non-decreasing, then $\mathbb{P}(\bigcup_n A_n) = \lim_n \mathbb{P}(A_n)$.*
If (A_n) is non-increasing, then $\mathbb{P}(\bigcap_n A_n) = \lim_n \mathbb{P}(A_n)$.
If $\mathbb{P}(A_n) = 0$ for all n, then $\mathbb{P}(\bigcup_n A_n) = 0$.
If $\mathbb{P}(A_n) = 1$ for all n, then $\mathbb{P}(\bigcap_n A_n) = 1$.

We define

$$\limsup_n A_n := \bigcap_n \bigcup_{k \geq n} A_k \,.$$

A point ω lies in $\limsup_n A_n$ if and only if

$$\forall n, \ \exists k \geq n, \ \omega \in A_k \,.$$

In common language, this means that A_n is realized infinitely often.

Lemma A.3 (Borel-Cantelli)**.** *If $\sum_n \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup_n A_n) = 0$.*

Proof. The sequence $B_n := \bigcup_{k \geq n} A_k$ is non-increasing. Thus $\lim \mathbb{P}(B_n) = \mathbb{P}(\bigcap_n B_n) = \mathbb{P}(\limsup_n A_n)$. As $\mathbb{P}(B_n) \leq \sum_{k \geq n} \mathbb{P}(A_k)$, the convergence of the series implies that $\mathbb{P}(B_n) \to 0$, hence the result. □

A.3 Characteristic functions

Let $\mathcal{X} = \mathbb{R}^d$ be finite dimensional.
The *characteristic function* of a probability measure $P \in \mathcal{P}(\mathcal{X})$ is the mapping $\phi_P : \mathcal{X} \to \mathbb{C}$ defined for every $t \in \mathcal{X}$ by

$$\phi_P(t) := \int e^{\imath \langle t, x \rangle} dP(x) \,,$$

where $\langle\,,\,\rangle$ stands for the inner production on $\mathcal{X}$ and where $\imath := \sqrt{-1}$. The characteristic function is well defined, and is continuous by simple application of Lebesgue's dominated convergence theorem.

Theorem A.4. *Distinct measures cannot have the same characteristic function.*

Proof. As a preliminary, consider the function $S : x \mapsto \int_0^x \frac{\sin u}{u} du$ on $\mathbb{R}_+ \to \mathbb{R}$, It is not difficult to prove that $S(x)$ admits a (positive) limit as $x \to +\infty$, because $\int_{(n-1)\pi}^{n\pi} u^{-1} \sin u \, du$ alternates in sign and its absolute value decreases to zero. You might remember that the limit is equal to $\frac{\pi}{2}$, but since we won't need this, just call the limit $\frac{\pi_0}{2}$ (in fact $\pi_0 = \pi$).
Consider the case $\mathcal{X} = \mathbb{R}$. Let $P \in \mathcal{P}(\mathbb{R})$. The proof consists in proving the following inversion formula:

$$P((a,b]) = \lim_{T\to\infty} \frac{1}{2\pi_0} \int_{-T}^{T} \frac{e^{-\imath ta} - e^{-\imath tb}}{\imath t} \phi_P(t) dt\,, \tag{A.1}$$

for every $a < b$ satisfying $P(\{a\}) = P(\{b\}) = 0$. Indeed, if Eq. (A.1) holds, then by letting $a \to -\infty$, we obtain that the distribution F_P is uniquely determined by ϕ_P, hence the conclusion.
We now prove Eq. (A.1). Denote by I_T the quantity inside the limit in (A.1). By Fubini's theorem,

$$I_T = \frac{1}{2\pi_0} \int \left(\int_{-T}^{T} \frac{e^{\imath t(x-a)} - e^{\imath t(x-b)}}{\imath t} dt \right) dP(x)\,.$$

Note that $\int_{-T}^{T} \frac{e^{\imath t(x-a)}}{\imath t} dt = 2\int_0^T \frac{\sin((x-a)t)}{t} dt = 2\text{sign}(x-a)S(T|x-a|)$ where $\text{sign}(u)$ is equal to 1 if $u > 0$, -1 if $u < 0$ and zero otherwise. Therefore,

$$I_T = \frac{1}{\pi_0} \int \left(\text{sign}(x-a)S(T|x-a|) - \text{sign}(x-b)S(T|x-b|)\right) dP(x)\,.$$

Using that $\lim_{u\to\infty} S(u) = \frac{\pi_0}{2}$, the integrand is bounded, and converges as $T \to \infty$ to the function

$$\psi_{a,b}(x) = \begin{cases} 0 & \text{if } x < a, \\ 0.5 & \text{if } x = a, \\ 1 & \text{if } a < x < b \\ 0.5 & \text{if } x = b, \\ 0 & \text{if } x > b. \end{cases}$$

By the dominated convergence theorem, Eq. (A.1) holds, as $P(\{a\}) = P(\{b\}) = 0$.
When $\mathcal{X} = \mathbb{R}^d$ with $d \geq 2$, an inversion formula very similar to Eq. (A.1) can be established, and the proof uses the same arguments, see [2, Eq. (29.3)] □

Proposition A.5. *Consider $P \in \mathcal{P}(\mathbb{R})$ and $p \in \mathbb{N}$ s.t. $\int |x|^p dP(x) < \infty$. Then, ϕ_P is p-times continuously differentiable and for every $t \in \mathbb{R}$, its p-th order derivative satisfies $\phi_P^{(p)}(t) = \int \imath^p x^p e^{\imath tx} dP(x)$.*

Proof. The statement was proven in the first year probability class. The proof is by induction and uses Lebesgue's dominated convergence theorem. □

We recall that the standard Gaussian distribution $\mathcal{N}(0,1)$ is the probability measure on $\mathbb{R}$ having the probability density function $x \mapsto (2\pi)^{-1/2}e^{-x^2/2}$.

Proposition A.6. *For every $t \in \mathbb{R}$, $\phi_{\mathcal{N}(0,1)}(t) = e^{-\frac{t^2}{2}}$.*

Proof. Denote by f the $\mathcal{N}(0,1)$-density function and note $\phi := \phi_{\mathcal{N}(0,1)}$. By Prop. A.5, ϕ is continuously differentiable and $\phi'(t) = \int \imath x e^{itx} f(x)dx$. Noting that $f'(x) = -xf(x)$ and integrating by parts, we obtain $\phi'(t) = -t\phi(t)$. The result follows by solving the differential equation and using $\phi(0) = 1$. □

The *characteristic function* of X, noted ϕ_X, is defined as the characteristic function of its law, that is $\phi_X := \phi_{X\#\mathbb{P}}$. With this notation, we can write

$$\phi_X(t) = \mathbb{E}(e^{\imath\langle t,X\rangle}) .$$

Let A be a linear operator on $\mathcal{X}$ to $\mathcal{Y}$, an other finite dimensional space. Let $b \in \mathcal{Y}$. Let X be a r.v. on $\mathcal{X}$

Lemma A.7. *For every $t \in \mathcal{Y}$, $\phi_{AX+b}(t) = e^{\imath\langle t,b\rangle}\phi_X(A^*t)$ where A^* is the adjoint of A.*

Proof. Immediate from the definition. □

A.4 A simpler proof of the Prokhorov theorem in $\mathbb{R}$

Denote by $\mathscr{F}$ the set of non-decreasing right-continuous functions on $\mathbb{R} \to [0,1]$.

Lemma A.8 (Helly)**.** *For every sequence (F_n) on $\mathscr{F}$, there exists $F \in \mathscr{F}$ and a subsequence (F_{φ_n}) such that*

$$F_{\varphi_n}(x) \to F(x)$$

for every x that is a point of continuity of F.

Proof. The set $\mathbb{Q}$ of rational numbers being denumerable, we write it as $\mathbb{Q} = \{x_0, x_1, \dots\}$. The sequence $(F_n(x_0))$ being bounded, there exists a value in $[0,1]$, which we call $G(x_0)$, and a strictly increasing map $\psi^0 : \mathbb{N} \to \mathbb{N}$ s.t.

$$F_{\psi^0_n}(x_0) \to G(x_0) .$$

The sequence $(F_{\psi^0_n}(x_1))$ being bounded, there exists a value in $[0,1]$, which we call $G(x_1)$, and a strictly increasing map $\psi^1 : \mathbb{N} \to \mathbb{N}$ s.t.

$$F_{(\psi^0\circ\psi^1)_n}(x_1) \to G(x_1) .$$

Continuing the process, be can recursively construct a sequence (ψ^k) of mappings on $\mathbb{N} \to \mathbb{N}$, and a sequence $(G(x_k))$ on $[0,1]$, such that for every k,

$$F_{(\psi^0 \circ \cdots \circ \psi^k)_n}(x_k) \to G(x_k)\,. \tag{A.2}$$

For every n, define $\varphi_n := (\psi^0 \circ \cdots \circ \psi^n)_n$. The mapping $\varphi : \mathbb{N} \to \mathbb{N}$ is stricly increasing. Moreover, for every $k \in \mathbb{N}$ and every $n \geq k$, note that

$$\varphi_n = (\psi^0 \circ \cdots \circ \psi^k)_{k_n}$$

where $k_n := (\psi^{k+1} \circ \cdots \circ \psi^n)_n$ tends to ∞ as $n \to \infty$ (to prove this, show that $k_n \geq n$). Thus for every $n \geq k$, $F_{\varphi_n}(x_k) = F_{(\psi^0 \circ \cdots \circ \psi^k)_{k_n}}(x_k)$ which proves that

$$\forall k \in \mathbb{N},\; F_{\varphi_n}(x_k) \to G(x_k)\,.$$

For every $x \in \mathbb{R}$, define

$$F(x) := \inf\{G(x_k)\, : k \in \mathbb{N},\, x_k > x\}\,.$$

Obviously, F is non-decreasing. We prove that it is right-continuous. Set $x \in \mathbb{R}$ and $\varepsilon > 0$. By definition of the infimum, there exists a rational number $x_k > x$ s.t. $G(x_k) < F(x) + \varepsilon$. Consider any point $y \in [x, x_k)$. It holds that $F(y) \leq G(x_k) < F(x) + \varepsilon$. Moreover, $F(x) \leq F(y)$ because F is non-decreasing. We have shown that for every $\varepsilon > 0$, there exists $\delta > 0$ (namely, $\delta = x - x_k$) s.t.

$$\forall y \in [x, x+\delta), \quad |F(y) - F(x)| \leq \varepsilon\,.$$

This proves that F is right-continuous.
Now, we must prove that $F_n(x) \to F(x)$ as $n \to \infty$. As above, chose $\varepsilon > 0$ and a rational number $x_k > x$ s.t. $G(x_k) < F(x) + \varepsilon$. By (A.2), this also reads $\lim_n F_{\varphi_n}(x_k) < F(x) + \varepsilon$. Since F_{φ_n} is non-decreasing, $F_{\varphi_n}(x) \leq F_{\varphi_n}(x_k)$, thus,

$$\limsup_{n\to\infty} F_{\varphi_n}(x) < F(x) + \varepsilon\,. \tag{A.3}$$

On the otherhand, choose any rational number x_ℓ s.t. $x - \varepsilon < x_\ell < x$. By the definition of F as an infimum, $F(x - \varepsilon) \leq G(x_\ell)$. By (A.2), this also reads $F(x-\varepsilon) \leq \lim_n F_{\varphi_n}(x_\ell)$. Since F_{φ_n} is non-decreasing, $F_{\varphi_n}(x_\ell) \leq F_{\varphi_n}(x)$, thus,

$$F(x - \varepsilon) \leq \liminf_{n\to\infty} F_{\varphi_n}(x)\,. \tag{A.4}$$

Putting together (A.3) and (A.4) and letting ε tend to zero, we obtain that $\lim_n F_{\varphi_n}(x) = F(x)$ whenever x is a point of continuity of F. □

Proposition A.9. *Any sequence on $\mathcal{P}(\mathcal{X})$ which converges weakly⋆ is tight.*

Proof. Consider a sequence on $\mathcal{P}(\mathcal{X})$ s.t. $P_n \xrightarrow{\mathcal{L}} P$. Define $N(x) := \|x\|$ for every $x \in \mathcal{X}$. Set $\varepsilon > 0$. By Prop. 2.6, any single probability measure is tight, thus, there exists $K > 0$ s.t. $PN^{-1}([-K,K]^c) > 1 - \varepsilon$. By the mapping theorem, $N_\# P_n \xrightarrow{\mathcal{L}} N_\# P$. Choose $K' > K$ s.t. $PN^{-1}(\{K'\}) = 0$. By Th. 2.3, $N_\# P_n([-K', K']) \to N_\# P([-K', K'])$. Thus $\liminf_n P_n(N^{-1}([-K', K'])) \geq 1 - \varepsilon$, or equivalently, $\limsup_n P_n(N^{-1}([-K', K'])^c) \leq \varepsilon$. Hence, (P_n) is tight. □

We establish the following version of the Prokhorov theorem in $\mathbb{R}$.

Theorem A.10. *A sequence of probability measures on $\mathbb{R}$ is tight if and only if every subsequence admits a further subsequence which converges weak$\star$.*

Proof. We prove the direct implication. Denote by F_n the distribution function of P_n. By Helly's lemma, there exists $F \in \mathscr{F}$ and a subsequence (F_{ψ_n}) s.t. $F_{\psi_n}(x) \to F(x)$ at every point of continuity x of F. If moreover F can be shown to be a distribution function *i.e.* if it satisfies $\lim_{x\to+\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$, then, (P_{ψ_n}) converges weakly$\star$ to the measure $P := \lambda F^{-1}$, where λ is the lebesgue measure on $[0, 1]$ and where F^{-1} is the generalized inverse defined by $F^{-1}(t) = \inf\{x : F(x) \geq t\}$.
Set $\varepsilon > 0$. As (P_n) is tight, there exists a compact set, which can be chosen of the form $[-K, K]$ s.t. $P_n([-K, K]) > 1 - \varepsilon$. This implies that for every $t > K$, $F_n(t) - F_n(-t) > 1 - \varepsilon$. In particular $F_n(t) > 1 - \varepsilon$. Choosing t as a point of continuity of F, and letting $n \to \infty$ along the subsequence ψ_n, it follows that $F(t) > 1 - \varepsilon$. As F is non decreasing, $\lim_{x\to+\infty} F(x) > 1 - \varepsilon$. By letting $\varepsilon \downarrow 0$, we conclude that $\lim_{x\to+\infty} F(x) = 1$. Finally, the inegality $F_n(-t) < \varepsilon$ leads to $\lim_{x\to-\infty} F(x) = 0$ by the same type of arguments.
Thus, we have shown that if (P_n) is tight, it admits a subsequence which converges weakly$\star$. To prove the conclusion, consider an arbitrary subsequence (P_{φ_n}). As (P_n) is tight, (P_{φ_n}) is tight as well, and thus admits a further subsequence which converges weakly$\star$.

We prove the converse of the theorem. Assume that every subsequence has a further subsequence which converges weakly$\star$. Assume, by contradiction, that (P_n) is not tight. Then, there exists $\varepsilon > 0$ s.t. for every compact set $K \subset \mathcal{X}$, $\sup_n P_n(K^c) > \varepsilon$. Consider the sequence of compact sets $(B_k : k \in \mathbb{N})$ given by the closed balls $B_k = \{x \in \mathcal{X} : \|x\| \leq K\}$. For every k, there exists $n \in \mathbb{N}$ s.t. $P_n(B_k^c) > \varepsilon$. Thus, one can construct a strictly increasing $\varphi : \mathbb{N} \to \mathbb{N}$ s.t. for all k, $P_{\varphi_k}(B_k^c) > \varepsilon$. By the standing hypothesis, one can extract from (P_{φ_k}) a further subsequence which converges weakly$\star$. Let us denote this subsequence by (P_{ψ_k}). By Prop. A.9, the latter sequence is tight. There exists a compact set K s.t. $P_{\psi_k}(K^c) < \varepsilon$ for every k. Choose a particular value of k in such a way that $K \subset B_k$. One has $P_{\psi_k}(B_k^c) < \varepsilon$, hence a contradiction. □

A.5 The Lévy-Prokhorov distance

To do.

Bibliography

[1] P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[2] P. Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1995.

[3] R. M. Dudley. *Real Analysis and Probability.* Wadsworth and Brooks/Cole, 1989.